Giardia trophozoite-secreted proteins and their effects on intestinal epithelia

By

Audrey Dubourg, M.Sc., B.Sc. (Hons)

A thesis submitted in fulfilment of the requirements of the University of East Anglia for the degree of Doctor of Philosophy

Norwich Medical School

University of East Anglia

June 2014

© This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that no quotation from the thesis, nor any information derived therefrom, may be published without the author's prior, written consent.

Declaration

I declare that the content of this thesis entitled "*Giardia* trophozoite-secreted proteins and their effects on intestinal epithelia" was undertaken and completed by myself, unless otherwise stated and has not been submitted in support of an application for another degree or qualification in this or any other University or institution.

Audrey Dubourg, MSc., B.Sc. (Hons)
University of East Anglia
June 2014

Acknowledgements

Firstly, I would like to thank my supervisors Dr Kevin Tyler and Prof Paul Hunter who have supported me throughout my PhD. Their help, guidance and trust in me have made this journey possible. I would also like to thank them for the many opportunities they gave me to present my work at various conferences. I am also grateful to them for loving all the food I baked and introducing me as one of the best baker in the lab every time they had visitors. I would like to acknowledge and thank all the members, past and present, of the "Tyler" Lab, but in particular Dr Claire Butler who guided me through the different protocols and was my well of "Lab knowledge", and Dr Maha Bouzid for her faith in me prior to the beginning of this PhD adventure and her guidance throughout it.

I would like to thank Dr Andrew Goldston and Dr Darren Sexton for their help with the flow cytometry assay; I would also like to thank Darren for all the French-Irish debates we had over the "Six Nations Tournament" throughout the years. I am grateful to Dr Gabriel Muntungi for his help and patience with all the western blot issues I faced during my thesis. I am also thankful to Dr Yongping Bao for supplying the intestinal cell lines any time I needed some; and Dr Paul Thomas for his help with the different microscopic assays and imaging issues. I am extremely grateful to Prof Jonathan Wastling and Dr Dong Xia for their help and implication in the *Giardia*-secreted proteins project, this thesis could have not happened without them.

I am also very grateful to the "BMRC floor 01 crew", your support and help in times of need, your jokes and silly discussion topics over the many hours of experiments have made this PhD greatly enjoyable or bearable when I hated it with all my heart. A big thank you to my Stats guru too, Mr Adrian Byrne! I am also very grateful to Dr Sumanto Haldar and Dr Noemi Tejera (my unicorn twin) for offering me a job, support and fun when I needed it the most.

I would like to thank with all my heart the 31st Norwich Rainbow unit and their leaders for welcoming me as part of their team/family. Volunteering with you has reminded me of what is important in life, your love and attention to these girls

and us leaders has given me the strength and motivation to keep going and “do my best”.

I would like to thank all my friends for their constant love, support, and faith in me even when I doubted myself. To quote Sorina: “I thank you all (Jaz, Sorina and Isa) for your relentless support, without which this work would have not been possible”. Thank you Isa, Sorina, Jaz, Helen, Gayatri for our girly evening and talks cheering me up every time I needed to be. I would also like to thank Stevie, David V., and Matt M for their fun, silliness and guidance; but also Davey for his famous “chippy dinners and game nights”. I would also like to thank my “Frenchy” friends, Julie and Dada, Baptiste, Celine (mon bouquet boulant), ma Vivi, Narjes (ma Holly Donkey) and ma Delfolle for their constant support too. Even thousand miles away, you were still in this with me no matter what happened. There are more people I would like to thank, but that would fill up a whole new thesis. No need to name you guys, you already know how awesome you are! Whether I met you during this journey or before, you were all a part of this, thank you all for making this adventure so interesting and fun.

And last but not the least; I would like to thank my family. My parents for their inconditional love and faith in me. My sister, Adeline who has given me the best present and most precious gift in my niece Lya. May I be an example to her of what women can do in life. My brothers, Adrien and Lisandro, for torturing me and reminding me that life is not just about work. Thank you all for the encouragements and motivation. You all made this possible.

Abstract

Giardia is a major cause of diarrheal disease worldwide. It is a flagellated cyst-forming enteric pathogen that inhabits the lumen of the small intestine. Two genetically distinct lineages (assemblages A and B) are of public health relevance and are often associated with water-borne outbreaks. Yet, the mechanism of pathogenesis and virulence in *Giardia* is poorly understood. A soluble component derived from healthy, viable and human infective *Giardia* trophozoites was shown to be able to mediate profound changes in the physiology of human derived enteric cells, consistent with the production of secreted virulence factors by the parasite. Quantitative proteomic analysis was successfully applied to the whole parasite and supernatants derived from the parasite in order to ascertain which parasite proteins are secreted. The genome of *Giardia* is believed to contain open reading frames which could encode as many as 6,000 proteins although hitherto there was only direct evidence for expression of a few hundred of these. Approximately 1,600 proteins were identified from each assemblage, the vast majority of which being common to both lineages. To look for actual enrichment in the supernatant, the ratio of proteins in the supernatant was compared with the pellet. This defined a far smaller group of putatively secreted proteins enriched comprising a high proportion encoded by genes annotated to have signal peptides, known virulence factors such as the Cathepsin B cysteine proteases and Variable Surface Proteins, scavenging proteins such as an extracellular nuclease and a high proportion of hitherto hypothetical proteins and proteins of unknown function. Further analysis of the genes encoding these proteins indicated that they were highly variable and likely to be under positive selection pressure, confirming their probable role in host-pathogen interactions and their potential as markers for discriminating virulent strains. Based on the proteomic analysis, a new model of pathogenic mechanism for *Giardia*-induced damage to enteric epithelium in which extracellular nuclease, Cathepsin B and Tenascin may have a concerted action was proposed and may have important implications in the understanding of *Giardia* pathogenesis.

CONTENTS

Contents

Contents

List of Figures

Chapter I

Chapter II

Chapter III

List of Figures

Chapter IV

Chapter V

Chapter VI

List of Tables

Chapter I

Chapter II

Chapter IV

Chapter V

List of Tables

Chapter VI

Abbreviations

2D-PAGE: Two-dimensional PAGE

AA: Amino acid

ABC: ATP-binding cassette

ABS-HI: Albumin bovine serum-heat inactivated

ADH: Arginine dehydrolase

ADI: Arginine deiminase

ADP: Adenosine diphosphate

AEE: Apical early endosome

AF: Anterior flagella

AIF: Apoptosis inducing factor

Ambic: Ammonium bicarbonate

APS: Ammonium persulfate

AQP: Aquaporin

ARE: Apical recycling endosome

ARF: ADP-ribosylation factor

Ass: Assemblage

ATP: Adenosine triphosphate

BEE: Basolateral early endosome

bg: β-giardin

BCA: Bradford acid assay

b.i.d.: Bis in die (twice a day)

BiP: Immunoglobulin heavy chain-binding protein

BSA: Bovine serum albumin

CaCC: Calcium-activated chloride channel

Caco-2: Human colonic adenocarcinoma derived epithelial cell line-2

cAMP: Cyclic adenosine monophosphate

CDC: Centre for Disease Control and Prevention

CDK: Cyclin-dependent kinase

CE: Common endosome

CF: Caudal flagella

CFTR: Cystic fibrosis conductance regulator

CID: Collision induced dissociation

CP: Cysteine protease

CoASH: Coenzyme A

COPII: Coatomer protein II

CVID: Common variable immunodeficiency

CWP: Cyst wall protein

Da: Dalton

DAPI: 4'-6-diamino-2-phenylindole

DFA: Direct fluorescence assay

DIDS: 4,4'-disothiocyanatostibene-2,2'-sulfonic acid

DIC: Differential interference contrast

DMEM: Dulbecco's Modified Eagle Medium

DNA: Desoxyribonucleic acid

dsDNA: Double-stranded DNA

ECC-RT-PCR: Post-trophozoite attachment electrophysiology analysis of cell culture combined with RT-PCR

ECDC: European Centre for Disease prevention and Control

EF-1 α: Elongation factor-1 alpha

EGF: epidermal growth factor

ELISA: Enzyme-linked immunosorbent assay

ER: Endoplasmic reticulum

ERK: Extracellular-regulated kinase

ESCP: Encystation-specific cysteine protease

ESI: Electrospray ionization

ESV: Encystation-specific vesicle

FCS: Foetal calf serum

FDR: False discovery rate

FISH: Fluorescence *in situ* hybridization

FITR: Fourier transform infrared spectroscopy

FNR: Ferredoxin: NAD(P)H oxidoreductase

G6DPH: Glucose-6-phosphate dehydrogenase

GAG: Glycosaminoglycan

***G*CATB:** *Giardia* Cathepsin B

gdh: Glutamate dehydrogenase

GlNR-1: *Giardia lamblia* nitroreductase-1

GLORF-C4: *Giardia lamblia* open reading frame-C4

Gly 101: Glybenclamide

Abbreviations

GO: Gene ontology
HCMP: High cysteine membrane protein
HIV: Human immunodeficiency virus
HPLC: High performance liquid chromatography
HRM: High resolution melt
Hsp90: Heat shock protein 90
iBAQ: Intensity based absolute quantification
IBS: Irritable bowel syndrome
IC: Immuno-chromatography
ICC: Immunocytochemistry
IFA: Immunofluorescence assay
IFN: Interferon
Ig: Immunoglobulin
IHC: Immunnohistochemistry
IHC-FR: IHC-frozen section
IHC-p: IHC on paraffin
IL-8: Interleukin 8
Isc: Short-circuit current
ITS: Internal transcribed spacer
InVS: Institut de Veille Sanitaire
JAM-1: Junctional adhesion molecule-1
kAE1: Kidney anion exchanger-1
KCC: Potassium-chloride co-transporter
kDa: Kilodalton
LBA: Long-branch attraction
LC: Liquid chromatography
LDL: Low-density lipoprotein
LE: Late endosome
LPS: Lipopolysaccharide
LTQ: Linear ion trap quadrupole mass filter
MAO: Monoamine oxidase
MAPK: Mitogen-activated protein kinase
Mb: Mega base pair
mRNA: Messenger RNA
miRNA: Micro RNA
ML: Maximum likelihood
MLCK: Myosin light chain kinase
MLG: Multi-locus genotyping
MPHB: Manitoba Public Health Branch
MS: Mass spectrometry
MsrB: Methionine-R-sulfoxide reductase
MT-PCR: Multiplexed tandem-PCR
MTZ: Metronidazole
MudPIT: Multi-dimensional protein identification technology
MW: Molecular weight
MWCO: Molecular weight cut-off
NADH: Nicotine adenine dinucleotide
NaTHNaC: National Travel Health Network and Centre
NEAA: Non-essential amino acid
NHS: National Health Service
NKCC: Na^{+}- K^{+}- $2Cl^{-}$ cotransporter
NMR: Nuclear magnetic resonance
NO: Nitric oxide
NSF: N-ethylmaleimide-sensitive factor
NZMH: New Zealand Ministry of Health
OCT: Ornithine carbamoyltransferase
ORF: Open reading frame
P: Pellet
P2Y2: Purinoreceptor-2
PBS: Phosphate buffered saline
PCD: Programmed cell death
PCR: Polymerase chain reaction
PCR-RFLP: PCR-based restriction fragment length polymorphism
PDI: Protein disulfide isomerase
PFA: Paraformaldehyde
PFOR: Pyruvate: ferredoxin oxidoreductase
PHAC: Public Health Agency Canada
PHE: Public Health England
PI: Propidium iodide
PKA: Protein kinase A
PKB: Protein kinase B
PKC: Protein kinase C
PLF: Posterior/lateral flagella
PP2A: Protein phosphatase 2A
PPIB: Peptidyl-prolyl cis-trans isomerase B
ppm: Parts-per-million
P/S: Penicillin-Streptomycin
PV: Peripheral vacuole

q.d.: Quaque die (once a day)

q.i.d.: Quater in die (four times a day)

q-PCR: Real time quantitative-PCR

RAPD: Random amplification of polymorphic DNA

rcf: Relative centrifugal force

rDNA: Ribosomal DNA

RER: Rough endoplasmic reticulum

RNA: Ribonucleic acid

RNAi: RNA interference

rRNA: Ribosomal RNA

rpm: Rotations per minute

ROCK: Rho kinase

ROS: Reductive oxygen species

SCIM: Scanning conductance ion microscopy

SDS: Sodium dodecyl sulfate

SDS-PAGE: Sodium dodecyl sulfate-polyacrylamide gel electrophoresis

SEM: Standard error mean

SFHP: Secreted formerly hypothetical protein

SGLT-1: Sodium-glucose co-transporter 1

SignalP: Signal peptide

SNAP: Soluble NSF attachment protein

SNARE: SNAP receptor

snoRNA: Small nucleolar RNA

SNP: Single-nucleotide polymorphism

SP: Supernatant

spp: Species

SR: SRP receptor

SRP: Signal recognition particle

***SSU*:** Small subunit

TAC: *Taq*Man array card

Taglin: Trypsin activated *Giardia* lectin

TBST: Tris-buffered saline and Tween20

TCA: Trichloroacetic acid

TEER: *Trans*-epithelial electrical resistance

TEM: Transmission electron microscopy

TEMED: N, N, N', N'-Tetramethylethylenediamine

TGN: *Trans*-Golgi Network

t.i.d.: Ter in die (three times a day)

TJ: Tight junction

TM: Transmembrane domain

TNF: Tumour necrosis factor

tpi: Triosephosphate isomerase

TRITC: Tetremethylrhodamine B isothiocyanate conjugate

tRNA: Transfer RNA

UDP-MurNAc: Uiridine diphosphate-peptidyl N-acetylmuramic acid

UTP: Uridine 5'-triphosphate

VD: Ventral disc

VF: Ventral flagella

VSP: Variant specific protein

WHO: World Health Organisation

XLA: X-linked agammaglobulemia

ZO-1: Zonula occludens-1

Chapter I

General introduction

1.1/ *Giardia* Historical Background

In 1681, Antony van Leeuwenhoek described *Giardia* for the first time while he was observing his own diarrhoeal stools under a microscope(Steuart et al., 2008) (Ford, 2005, Dobell, 1920). The next description was not until 1859 when the organism was described in greater detail by Vilém Dušan Lambl who thought this parasite belonged to the genus *Cercomonas*. Thus, he named it *Cercomonas intestinalis*. Thereafter, some investigators named the genus after Lambl while others named the species infecting humans after him (*Giardia lamblia*) (Adam, 2001). In 1879, Battista Grassi discovered a rodent parasite, now known to be a *Giardia* species, and named it *Dimorphus muris* apparently unaware of Lambl's earlier description. In 1882 and 1883, Johann Künstler described an organism in tadpoles (possibly *G.agilis*) that he named *Giardia*. It was the first time that *Giardia* was used as a genus name. This genus name was chosen to honor Professor Alfred Giard, an eminent french zoologist. Raphaël Blanchard, in 1888, suggested *Lamblia intestinalis* after Lambl. Charles W. Stiles changed it to *Giardia duodenalis* in 1902 and to *Giardia lamblia* in 1915 to commemorate Professor Giard and Dr. Lambl (Ford, 2005, Adam, 2001). The same year (1915), Charles A. Kofoid and Elizabeth B. Christiansen proposed *Giardia lamblia* instead of Blanchard's suggestion of *Lamblia intestinalis* (Kofoid and Christiansen, 1915) and, in 1920, Kofoid and Christiansen used *Giardia enterica* (Adam, 2001).

For many years, the naming of species caused controversy. Some researchers suggested naming species on the basis of the host of origin, which could have led to name over 40 species. On the other hand, others were focusing on morphology like Charles E. Simon, in 1922, who used morphologic criteria to distinguish between *G. lamblia* and *G. muris* and accepted the name *G. lamblia* for the forms infecting humans (Simon, 1922).

In 1952, Francis Filice published a detailed morphologic description of *Giardia* and proposed that three species names be used on the basis of the median body's morphology: *G. duodenalis*, *G. muris* and *G. agilis* (Adam, 2001). The names for the human parasite *Giardia duodenalis*, *Giardia lamblia* and *Giardia intestinalis* are all currently used despite the potential for confusion that this has created.

Indeed, through the 1970s, the species name *Giardia lamblia* became widely accepted; and, since the 1980s, some scientists have encouraged the use of the name *Giardia duodenalis.* Moreover, in the 1990s, the use of *Giardia intestinalis* has been encouraged. Since the beginning of the XXI century, the use of one of the three names rather than another seems to have lessened. In fact, *G. duodenalis*, *G. intestinalis* and *G. lamblia* appear to be used equally in publications. According to PubMed, for a total of 12,256 publications on *Giardia*, 31 % (3,777) used the term *G. lamblia*, 34 % (4,209) used *G. intestinalis* and 35 % (4,270) used *G. duodenalis*. In 2009, Monis and collaborator proposed a revised taxonomy based on genetics and suggesting that each assemblage is a lineage evolving independently. *G. duodenalis* was suggested for assemblage A, *G. enterica* for assemblage B, *G. canis* for assembalges C and D, *G. bovis* for assemblage E, *G. cati* for assemblage F and *G. simondi* for assemblage G (Table 1.1). However, these names have been hardly used in publications. In 2013, however, 259 studies on *Giardia* were published, 145 (56%) referred to *Giardia* as *G. lamblia*, the 114 publications left used equally *G. intestinalis* (22 %) and *G. duodenalis* (22 %). For the purposes of this dissertation, I will use *Giardia* to identify the species complex and I will identify the different assemblages as it is currently common practice.

1.2/ Classification of *Giardia*

1.2.1/ Phylogeny

As described above, for many years, the classification of *Giardia* was only based on morphology, which had led to the description of a large number of species and a history of taxonomic confusion and controversy for the *Giardia* genus (Caccio et al., 2005, Thompson, 2002). Today, with the development of molecular techniques, genotypic classifications are based on sequence analysis of genetic markers integrated with structural, biochemical correlations and with pathology; which allows a more robust classification of *Giardia* and other protists (Plutzer et al., 2010, Cavalier-Smith, 2003).

Traditionally, *Giardia* has been considered as belonging to Phylum Sarcomastigophora, Subphylum Mastigophora (=Flagellata), Class

Zoomastigophorea, Order Diplomonadida and Family Hexamitidae (Morrison et al., 2007).

A decade ago, a new systematic classification was developed and based on genetic, structural and biochemical data (Keeling et al., 2005). According to this, *Giardia* would belong to Domain Eukaryota, Kingdom Excavata (unicellular Eukaryotes which contain free-living and symbiotic forms and also some important parasites of humans), Phylum Metamonada (large group of flagellate protozoa), Subphylum Trichozoa, Superclass Eopharyngia, Class Trepomonadea, Subclass Diplozoa, Order Giardiida, Family Giardiidae and Genus *Giardia* (Keeling et al., 2005) The members of the Excavata kingdom, which is a divergent and controversial "super-group", are loosely united by a combination of morphological and molecular evidence. The phylogenetic robustness of the Excavata kingdom as a whole has yet to be determined; however, three subgroups are considered as clades: (i) Preaxostyla (including *Trimastix* and *Oxymonads*), (ii) Fornicata (Diplomonads, Retortamonads and *Carpediemonas*), (iii) an unnamed clade consisting of euglenozoa, Heterolobosea and Jakobida (Hampl et al., 2009). In 2005, however, Keeling et al showed that there were morphological similarities in cytoskeletal ultrastructure uniting this kingdom (Keeling et al., 2005). In 2006, multigene phylogenetic analyses of six nuclear-encoded protein-coding genes showed that excavates were further associated (Simpson et al., 2006). In 2009, Hampl et al also showed the monophyly of Excavata by progressively removing long-branch gene sequences and taxa to decrease the long-branch attraction (LBA) artefact. However, the robust monophyletic group obtained was still not fully recovered as globally optimal (Hampl et al., 2009) (Fig 1.1). Although many studies suggest that *Giardia* belongs to the eukaryotic super-group Excavata, the small nucleolar ribonucleic acids (snoRNAs) latterly identified from *Giardia* might be considered to imply *G. lamblia* emerged somewhat later or more adjacent to the higher eukaryotes (Luo et al., 2009).

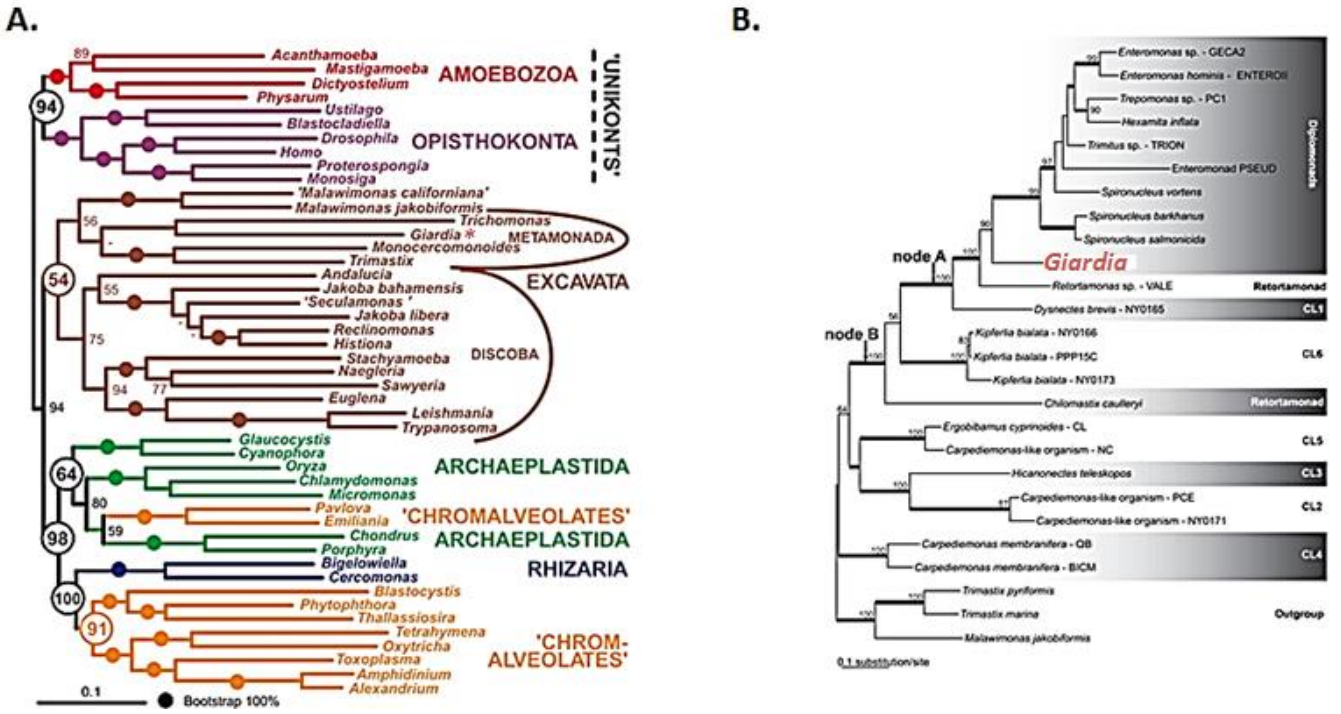


Fig 1. 1: *Giardia*'s taxonomy in the evolutionary tree of life (based Hampl et al, 2009, and Takashita et al, 2012, with permission). **A.** Maximum-likelihood (ML) tree obtained after the removal of 1,750 of the longest-branch genes sequences. The numbers at the nodes indicate bootstrap support; branches with a maximum support are shown by full circles. Dashes indicate bootstrap value < 50 % (Hampl et al., 2009). **B.** Maximum-likelihood phylogenetic tree of Fornicata based on six concatenated proteins (α-tubulin, β-tubulin, HSP70, HSP90, EF-1α and EF2). Trimastix pyriformis, Trimastis marina and Malawimonas jakobiformis (three excavate species) were used to root the tree. Boostrap values are indicated for nodes with support > 50 %. Relationships with over 0.90 Bayesian posterior probabilities are shown by thick branches (Takishita et al., 2012).

1.2.2/ Taxonomy

The first original species descriptions were based on presumed host specificity because all species possess cysts with the same morphology except for *G. microti* which can be identified by the fact that its cysts contain two differentiated trophozoites whereas cysts from other species contain a single trophozoite. However, this was abandoned in favour of trophozoite morphology based on comparisons initiated by Filice in 1952 (Monis and Thompson, 2003, Filice, 1952). In the Genus *Giardia*, six species have been identified on the basis of light microscopic (shape of trophozoite and median body) and electro microscopic (form of ventrolateral flange, ventral disc and flagellum) characteristics (Fig1.2) (Plutzer et al., 2010, Monis and Thompson, 2003, Adam, 2001). *G. agilis* which infects Amphibians; *G. ardeae* and *G. psittaci* are found in Birds; *G. muris* and *G. microti* are found in rodents (Caccio et al., 2005, Adam, 2001). The sixth species, *G. lamblia* (*G. duodenalis* or *G. intestinalis*), includes *Giardia* strains isolated from a large range of mammalian hosts (Fig1.3).

Considering its wide host range and the small amount of morphological variation displayed, *Giardia lamblia* can be considered as a species complex. Seven distinct groups, named assemblages, are assigned to *Giardia* (assemblage A to G) based on the genetic analysis of different *Giardia* genes such as triosephosphate isomerase (*tpi*), glutamate dehydrogenase (*gdh*), and β-giardin (*bg*) genes. The closely related genotypes are grouped into assemblages and sub-assemblages (Feng and Xiao, 2011, Plutzer et al., 2010, Sprong et al., 2009, Adam, 2001).

Assemblage A and B are responsible for human infections and are also found in a wide range of mammals. Assemblage A can also be found in livestock; and assemblage B has been found in wildlife and dogs but to a much lesser extent than in humans. The remaining assemblages show more restricted host ranges, assemblages C and D are found in Canids, E in livestock, F in cats, and G in rodents (Sprong et al., 2009, Caccio et al., 2005) (Table 1.1). Isoenzyme analysis of laboratory-adapted strains and deoxyribonucleic acid (DNA) sequence analyses at different loci showed the presence of subgroups, called sub-assemblages, in

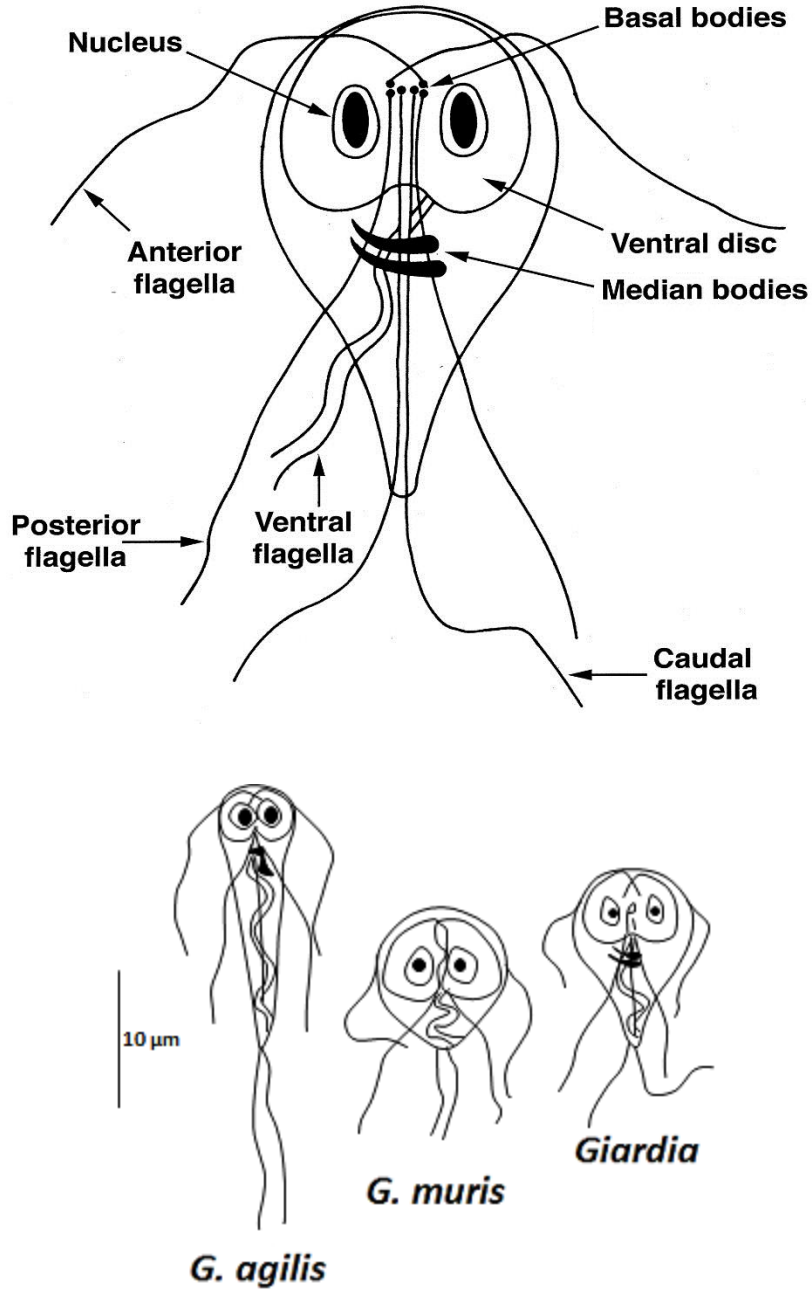


Fig 1. 2: Morphological features of a *Giardia* trophozoite and examples of trophozoite morphology of three species of *Giardia* (modified from Monis and Thompson, 2003, with permission).

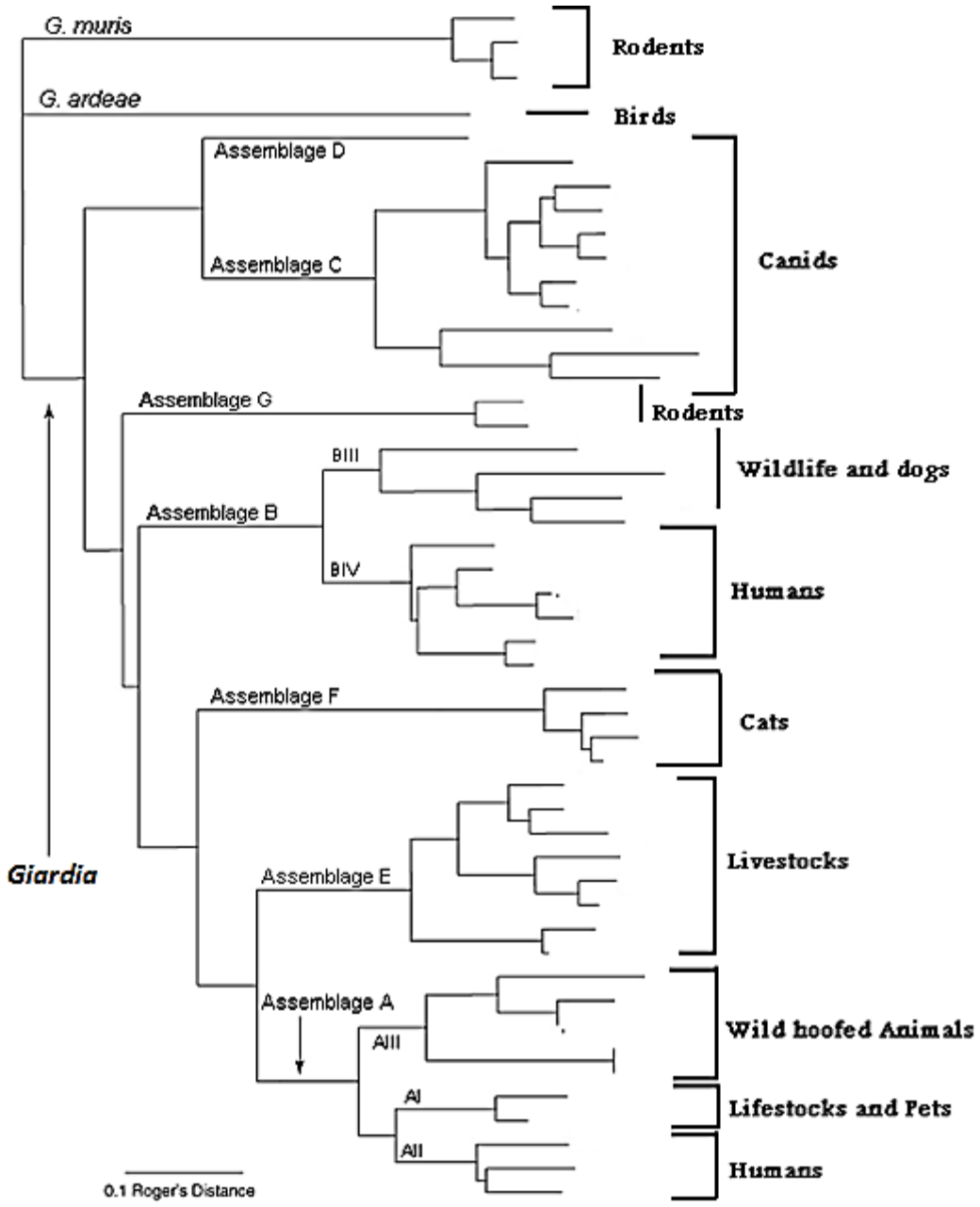


Fig 1. 3: Dendogram depicting the genetic relationships of isolates of G*iardia.* (modified from Monis et al, 2003, with permission). Genetic relationships of *Giardia*, *G. muris* and *G. ardeae* were inferred by Neighbour-Joining analysis of allozymes of 21 enzymes detected in different isolates (and, rarely, within single isolate), via isoenzyme profiles, using Roger's Distance. Roger's distance was obtained from the mean Euclidean distance between allele frequencies at each locus identified by allozyme analysis. The mean Euclidean distance is considered as the shortest straight-line distance between the focal point (here *Giardia*) and its nearest neighbour (*G. muris*, *G. ardeae* and *Giardia* different assemblages), the Euclidian distance is based on the Pythagorean formula.

assemblages A and B (Sprong et al., 2009, Monis et al., 1999). Some studies have shown that assemblages A and B encompassed genetic groups I to IV, with sub-assemblages AI, AII, BIII, and BIV (Monis et al., 2009, Robertson et al., 2007, van der Giessen et al., 2006, Monis et al., 2003, Mayrhofer et al., 1995). The *Giardia* sub-assemblages AI and AII were shown to cycle predominantly within defined hosts, AI in livestock, AII in humans; and these cycles don't interact significantly (Monis et al., 2009, Sprong et al., 2009). Some genotyping studies have, however, reported AII in dogs and cats in some Indian and Mexican communities (Ballweber et al., 2010, Eligio-Garcia et al., 2008, Traub et al., 2004). A new sub-assemblage, AIII, was identified in wild hoofed animals; (Sprong et al., 2009) (Table 1.1). In assemblage B, two sub-assemblages were identified: BIII and BIV. In human isolates, these two sub-assemblages were found with a very similar frequency (Sprong et al., 2009). They were also found in other mammals (beavers, monkeys, marine animals) with a similar frequency too but in a lesser prevalence than in humans (Sprong et al., 2009) (Table 1.1). Several genes, such as *tpi*, *bg* and the small subunit rRNA gene (SSU-rRNA) are used to discriminate *Giardia* DNA at an assemblage level; however, only *bg*, *tpi* and *gdh* are sufficiently discriminatory to characterise amongst sub-assemblages (Scorza et al., 2012, Sprong et al., 2009). Most sub-assemblage studies only used single locus approaches to characterise and subtype strains from both assemblage A and B which can lead to some discrepancy in the taxonomy. In order to increase accuracy of *Giardia* assemblage and sub-assemblage investigations, a multi-locus genotyping strategy is now used to study *Giardia* genetic diversity (Huey et al., 2013, Scorza et al., 2012, Covacin et al., 2011, Lebbad et al., 2010, Caccio et al., 2008). Beside assemblages A-G, a new *Giardia* assemblage was identified and characterised in marine vertebrates (seals and gulls): Assemblage H (Lasek-Nesselquist et al., 2010). This new assemblage was only isolated in a single geographic area suggesting a local source; and it was also only genetically characterised by *gdh* amplification. The amplification of *tpi* was also used to characterise *Giardia* species but did not provide evidence of this new assemblage. Moreover, *tpi* amplified sequences showed a bias towards assemblage B. Plus, none of the other gene usually used to characterise *Giardia* species were used during this study. Therefore, this assemblage H, isolated in marine vertebrates, cannot be

completely accepted as a new *Giardia* assemblage yet and requires further investigation.

In 2009, Monis et al proposed a revised taxonomy of *Giardia* species including the host specificity (Monis et al., 2009) (Table 1.2). They showed that the genetic distance separating assemblages A and B is at the same level as the genetic distance separating the other suggested species. They stated that a revised taxonomy of *Giardia* was necessary for two reasons. First, taxonomy should recognise and reflect the biological and evolutionary differences within *G. lamblia*, specifically the host specificity. Second, recognising the different assemblages as distinct species could affect ways of thinking in terms of zoonotic potential and human health threats. To Monis et al, the fact that the genetic characteristics of the assemblages are preserved in sympatric, local areas where the cycles of transmission possibly interact reinforces the argument that *Giardia* assemblages represent independently evolving metapopulations. In their study, Monis et al showed some limitations in the current taxonomy of *Giardia*. However, too many studies have only used a single locus genotyping sequence to identify *Giardia* assemblages and their subtypes. It has been shown that a gene history may be different from a species phylogeny (Heled and Drummond, 2010). Indeed, discrepancies such as horizontal transfer, lineage sorting and gene duplication/extinction, between a single locus tree and a species tree may be problematic for closely related species and sub-species. Therefore, more genetic information on the different assemblages and sub-assemblages are needed to determine whether or not assemblages are distinct *Giardia* species. Using the multi-locus genotyping approach instead of a single marker may determine if a new taxonomy is required for *Giardia*.

1.2.3/ Sex: an important factor to consider when classifying all forms of life such as *Giardia*

The origin of sex has always been a central evolutionary question in phylogeny and taxonomy. Indeed, one of the major innovations in eukaryotes was

Table 1.1: *Giardia* species and their relationship with their hosts.

***Giardia* species**	**Major hosts**
G. agilis	Amphibians
G. ardeae	Birds
G. psittaci	Birds
G. muris	Rodents
G. microti	Muskrats, Voles
G. lamblia :	
Assemblage A	
Sub-assemblage AI	Livestock, pets
Sub-assemblage AII	Humans
Sub-assemblage AIII	Wild hoofed animals
Assemblage B	Humans, wildlife and dogs
Sub-assemblage BIII	
Sub-assemblage BIV	
Assemblage C/D	Canids
Assemblage E	Livestock
Assemblage F	Cats
Assemblage G	Rodents
Assemblage H	Marine Vertebrates

Table 1.2: Revised taxonomy of *Giardia* and other species (Monis et al., 2009, with permission).

Species (= assemblage)	**Host**
G. duodenalis (=assemblage A)	Humans and other primates, dogs, cats, livestock, rodents,and other wild mammals
G. enterica (= assemblage B)	Humans and other primates, dogs, some species of wild mammals
G. agilis	Amphibians
G. muris	Rodents
G. psittaci	Birds
G. ardeae	Birds
G. microti	Rodents
G. canis (= assemblage C/D)	Dogs, other canids
G. cati (= assemblage F)	Cats
G. bovis (=assemblage E)	Cattle and other hoofed livestock
G. simondi (= assemblage G)	Rats

[a]Designation based on original taxonomic descriptions

the origin of sexual reproduction by meiosis with its consequences for effecting genome-wide recombination (Cavalier-Smith, 2002). Genetic exchanges have been extensively studied in protists such as kinetoplastids and apicomplexans (Heitman, 2006, Gaunt et al., 2003, Bingle et al., 2001). And yet, little is known about sex in other protists such as *Giardia* and other metanomads. Several studies have though suggested that *Giardia* might undergo sexual reproduction (Lasek-Nesselquist et al., 2009, Logdson, 2008, Cooper et al., 2007, Heitman, 2006, Ramesh et al., 2005, Tibayrenc et al., 1991).

In eukaryotes, sexual reproduction is accomplished by meiosis. It is a specialised process of cell division that results in haploid cells (Villeneuve and Hillers, 2001, Heywood and Magee, 1976). In 2001, Villeneuve and Hillers identified stereotypical reductive division in meiosis as a major evolutionary innovation in eukaryotic cells; stipulating that the delineation of its history might be key to understanding the evolution of sex. In 2005, Ramesh et al identified a set of genes encoding homologues of known components of the "core meiotic recombination machinery". By conducting a phylogenic study of the genes found in *Giardia* and their putative homologues in other eukaryotes, Ramesh et al showed that the *Giardia* genes are orthologs of their counterparts in organisms known to undergo sexual reproduction. From the 17 meiotic genes identified, all but four were found in *Giardia.* According to Ramesh et al, when compared to the same genes in the databases of animals, plants and fungi known to have meiotic sex, each to the four genes not found in *Giardia* is also missing from one or more sexual organisms (Birky, 2005, Ramesh et al., 2005). None of these genes contain mutations that would interrupt their open reading frames (ORFs) (Birky, 2005). Such mutations would accumulate in non-functional genes and not in functional genes; which shows that *Giardia* might be capable of sexual reproduction.

Population genetics studies have analysed the diversity of genes at molecular levels, focusing on allelic sequences. Alleles are alternate forms of a gene that vary from each other in sequence, yet code for the same polypeptide or RNA (Birky, 1996). Alleles usually occupy the same locus on a particular chromosome. In diploid or polyploid organisms, the copies (two or more) of a gene can be the same or different alleles; this is referred as allelic homozygosity and allelic heterozygosity

respectively (Birky, 1996). A population genetics theory suggests that high levels of divergence in a lineage are evidence that the lineage has been asexual for a very long time. Indeed, organisms that asexually reproduce do not share alleles among lineages; and the allelic copies begin to acquire different mutations from the moment that sexual reproduction is lost (Birky, 1996). In 2008, Poxleitner et al suggested that a fusion of the two *Giardia* nuclei might occur during the encystation of *Giardia* trophozoites; this phenomenon of nuclear fusion is called karyogamy and is found in many polyploid and haploid organisms (Poxleitner et al., 2008, Birky, 1996, Tibayrenc et al., 1991). According to Poxleitner et al, this potential karyogamy of *Giardia* nuclei could result to *Giardia* genome's low levels of allelic heterozygosity. In their study, Poxleitner et al also showed that karyogamy and material genetic exchanges take place during encystation using fluorescence *in situ* hybridization (FISH) and transmission electron microscopy (TEM). Moreover, two other studies showed that recombination and genetic exchanges within and between assemblages were possible in *Giardia*; using population genetics study, polymerase chain reaction (PCR) and sequencing, genealogies and tests for recombination (Lasek-Nesselquist et al., 2009, Cooper et al., 2007). However, despite the fact that genetic exchange was shown to be possible in *Giardia* via tests for recombination, this phenomenon hasn't been directly observed in *Giardia* yet. The only conclusions we can draw from Lasek-Nesselquist et al., Poxleitner et al., Cooper et al., and Ramesh et al. studies are: (i) *Giardia* might be capable of meiosis and, thus, sexual reproduction; (ii) the evolution of meiosis might have happened early in the eukaryotic evolution; and (iii) the "conserved meiotic machinery" is composed of a set of genes that encode a variety of component proteins (Lasek-Nesselquist et al., 2009, Poxleitner et al., 2008, Cooper et al., 2007, Birky, 2005, Ramesh et al., 2005).

The studies by Cooper et al (2007) and Lasek-Nesselquist et al (2009) provide convincing evidence that *Giardia* exhibits genetic exchange. However, direct evidence of meiotic recombination is still lacking. Furthermore, if meiotic reproduction does take place in *Giardia*, its frequency should be determined as well as the mechanism of meiotic reproduction involved. Ascertaining whether meiotic

reproduction occurs in *Giardia* will yield a better understanding of *Giardia*'s phylogeny and taxonomy (Morrison et al., 2007, Birky, 1996).

1.3/ Giardiasis, transmission and health importance

1.3.1/ Transmission

Transmission of *Giardia* occurs via direct contact with infected humans or animals, or via ingestion of contaminated water or contaminated food. Transmission of protozoan pathogens by contaminated food can occur via 3 main routes: (i) through contamination of food ingredients on the farm by polluted irrigation water, (ii) through contaminated materials involved in the final step for product processing or washing, or (iii) through transfer or spread via infected food handlers or food preparers (Dawson, 2005). Food transmission of *Giardia* mainly happens by cross-contamination via people with children in nappies and/or pets. The first food transmission by food handler was reported in 1981 by Osterholm et al. In this outbreak, 29 people developed giardiasis; investigations showed that this outbreak was caused by eating contaminated home-canned salmon. A child in nappies was present in the home where contaminated food was prepared; it was therefore suggested that the child was infected with *Giardia* and that the food preparer failed washing her hands after changing the child nappies before handling food (Smith et al., 2007, Porter et al., 1990, Osterholm et al., 1981). In 1990, Porter et al reported another giardiasis outbreak caused by eating contaminated fruit salad. A child in nappies was also present in the food handler's home in this case. The first hypothesis of contamination for this case was the same as for the first outbreak due to home-canned salmon, the food preparer failed to wash her hands after changing the nappy and before handling the food. However, a pet rabbit was also present in the case reported in 1990 by Porter and collaborator. This rabbit may have been positive for *Giardia* when it entered the home and therefore was excreting cysts. Cysts may have been transferred to the food preparer's hands who failed to wash her hands prior to handle the fruit salad. In both cases, contamination came from food handlers failing to wash their hands before cooking which shows the importance of a good hygiene prior to handle any raw food.

During the outbreak reported by Porter and collaborator, the question of whether domestic animals could have also played a part in this transmission had been raised. In 2005, Hunter and Thompson discussed the possibility of zoonotic transmission of *Giardia* by comparing different case control studies. They came to the conclusion that zoonotic transmission is not strongly supported for giardiasis. Indeed, most types of *Giardia* are adapted to single host species which reduces the probability of zoonotic transmission as a major contamination route for *Giardia* (Hunter and Thompson, 2005) (Table 1.1).

The cyst, which is the infective form responsible for the infection dissemination, is excreted in the environment by infected hosts. It can remain viable and infectious for long periods in a moist, cool environment (WHO, 2008). In 2006, Roberston and Gjerde studied the resistance of *Giardia* cysts and their viability in a Norwegian river environment over winter time. After 39 days, viable cysts were still detected in the control samples (constant temperature: +4°C in cold chambers) but not after 54 days. For the cysts in the river environment, viable cysts were detected up to 29 days (water temperature varied between 7.3 and 1.1°C over the 162 days of winter) (Robertson and Gjerde, 2006). In 1954, Rendtorff determined that *Giardia* had a low infective dose, between 10 and 100 cysts had to be ingested by the host to develop the symptoms of giardiasis. According to the World Health Organisation (WHO), fewer than 10 cysts represent a meaningful risk of infection for one host (WHO, 2008, Rendtorff, 1954). Although *Giardia* cysts have been shown to be resistant to cold water for long period of time, it has been suggested that their infectivity may not be as high as for freshly excreted cysts (Alum et al., 2012). A new method combining RT-PCR with electrophysiology analysis of cell culture post-trophozoite attachment (ECC-RT-PCR) was developed to test the viability and infectivity of *Giardia* cysts (Alum et al., 2012). However, this technique was only tested on *G. muris* and only for its development. For how long *Giardia* cysts remain infective has yet to be investigated.

One of the most efficient ways to remove microorganisms from drinking water is through a filtration process. Filtration consists in the physical removal of turbidity and microorganisms (Betancourt and Rose, 2004). During filtration, water passes through a pore structure composed of diverse bed materials such as a bed of

sand (sand filtration), a layer of diatomaceous earth which are siliceous sedimentary rocks (diatomaceous earth filtration), and a combination of coarse anthracite coal overlying finer sand (dual- and tri-media filtration) (Betancourt and Rose, 2004). However, a study showed that conventional water treatment systems did not always ensure the complete removal of waterborne parasites as some *Giardia* cysts were found (LeChevallier et al., 1991). This study also suggested that high disinfection levels or more efficient disinfection procedures were required to protect against passage of *Giardia* cysts. Chlorination is the disinfection step the most used in most countries to obtain water for human consumption; but *Giardia* cysts are highly resistant to chlorination (Betancourt and Rose, 2004). Most of the giardiasis outbreaks are due to contaminated recreational/drinking water (Nichols et al., 2009, Karanis et al., 2007, Hunter and Thompson, 2005, Wallis et al., 1996). Swimming in contaminated waters and swimming pools has also been recognised as an important route of transmission. Contamination of recreational waters and swimming pools are due to human/animal wastes, accidental faecal contamination and rainfall/storm waters (Nichols et al., 2009, Shields et al., 2008, Karanis et al., 2007). Besides the transmission via contaminated treated water, several studies have highlighted the major role of untreated water, such as lakes and rivers (Hunter and Thompson, 2005, Welch, 2000). Untreated water is mainly used by populations from developing countries in which main urban areas have water treatment systems; but these systems are not extended to agricultural areas. Populations use water from rivers and lakes on a daily basis for purposes such as cooking, drinking, bathing, swimming and washing food. The lack of latrines can also contribute to a high transmission rate; inhabitants are defecating directly into the water (Noor Azian et al., 2007, Balcioglu et al., 2007, Prado et al., 2003, Cifuentes et al., 2000)

Infected hosts act as a reservoir potentially contaminating new environments when releasing cysts in their faeces. In 1954, Rendtorff showed that the duration of cysts shedding may vary from 5 to 41 days after the end of the symptoms for most patients after feeding prisoners with different doses of *Giardia* cysts from asymptomatic human donors. However, for two patients, the release of cysts persisted for 129 and 132 days. According to Rendtorff, symptoms usually last

for 6 to 15 days, with an average of 9.1 days, in immunocompetant people (Rendtorff, 1954).

1.3.2/ Pathology and epidemiology

1.3.2.1/ Giardiasis

Giardiasis clinical manifestations are highly variable between infected hosts and can range from acute to chronic infections; while some individuals may remain asymptomatic. The incubation period usually varies from 9 to 15 days with an acute stage generally beginning with a feeling of intestinal disturbance, nausea and anorexia (Wolfe, 1992). The main symptom of giardiasis is diarrhoea which generally leads to the disease diagnosis. Characteristically, the diarrhoea is watery and can be explosive and foul-smelling; it is often associated with fatigue and loss of weight. The other typical symptoms developed by infected hosts are bloating, abdominal pain, vomiting, and mild-fever (Cotton et al., 2011, Wolfe, 1992, Flanagan, 1992). Diarrhoea can occur with or without malabsorption syndrome. Malabsorption may be responsible for the loss of weight; even in the absence of diarrhoea, nutrients may be malabsorbed (Beaumont and James, 1986). Occasionally, some non-specific symptoms, such as lassitude, headache and myalgia, occur during a chronic phase (Wolfe, 1992). Moreover, some extra-intestinal symptoms may occur including polyarthritis, apthous ulceration, or urticaria (Flanagan, 1992). These symptoms can also be associated with failure to thrive. Children, especially infants, are more infected than adults (Botero-Garces et al., 2009, Bartlett et al., 1991); those malnourished and from developing countries in particular (Botero-Garces et al., 2009). Several studies have shown the relation between *Giardia* infections and stunting in infected children. Children with chronic giardiasis often fail to thrive and may develop growth defects (Kotloff et al., 2013, Botero-Garces et al., 2009, Berkman et al., 2002, Sullivan et al., 1991). Early childhood diarrhoeal illness has been associated with stunted growth which may lead to low cognitive function, metabolic syndromes and pre-diabetes, altered intestinal mucosal architecture associated with changes in small bowel functions, reduced enterocyte mass, and evidence of immune activation and inflammation of

the mucosa (Scharf et al., 2014, Kotloff et al., 2013, Johnston et al., 2013, Barakat and Halawa, 2013, Richard et al., 2013).

The acute phase of giardiasis may last for days to weeks. The host factors, such as immune status, diet, age or concomitant intestinal microbiota, play a part in the severity of symptoms and the duration (Cotton et al., 2011, Robertson et al., 2010, Grazioli et al., 2006); *Giardia* parasite factors, such as genotype, resistance to treatment and ability to evade the host immune system have key roles in the severity and duration of the disease (Robertson et al., 2010). The acute stage can often be followed by a chronic phase; chronic symptoms may be sporadic, continuous, intermittent or recurrent with episodes of diarrhoea. This chronic stage may last for years. Immunocompromised individuals seem more at risk to undergo chronic giardiasis. Indeed, patients with hypogammaglobulinemia have been associated with a predisposition to develop chronic giardiasis; although HIV patients don't seem to show increased sensitivity to *Giardia* infection (Nkenfou et al., 2013, Oksenhendler et al., 2008, Escobedo and Nunez, 1999). This suggests that $CD4^+$ T lymphocytes may not be the T cells involved in the clearance of *Giardia* during infection. Chronic giardiasis seems to have an increased frequency in patients with immunodeficiency diseases such as Immunoglobulin (Ig) deficiency, X-linked agammaglobulemia (XLA), and common variable immunodeficiency (CVID) (Agarwal and Mayer, 2009, Herbst et al., 1994, Hermaszewski and Webster, 1993). XLA, Ig deficiency and CVID are either responsible for defective mature B cells or for failure to normally develop and differentiate B cells into plasma cells (Agarwal and Mayer, 2009, Herbst et al., 1994). The most common immunodeficiency is an Ig A deficiency. Luminal Ig A might play a role in the clearance of parasite; therefore, the lack of secretory Ig A in patients may allow attachment and proliferation of intestinal parasites such as *Giardia* (Agarwal and Mayer, 2009). CVID has also been suggested to cause a T cell dysfunction due to the fact that gastrointestinal diseases, and giardiasis in particular, are more common in patients with CVID (Herbst et al., 1994). However, why some patients with CVID develop gastrointestinal diseases while others do not still remains unclear (Agarwal and Mayer, 2009)

A few studies also showed similarities of symptoms between *Giardia* infections, especially chronic infections, and irritable bowel syndromes (IBS). D'Anchino et al, in 2002, demonstrated the association of persistent low grade inflammation with faecal persistence of *Giardia*; while Grazioli et al, in 2006, investigated the prevalence of *Giardia* in 137 patients with IBS (Grazioli et al., 2006, D'Anchino et al., 2002). To prevent information and observational bias, Grazioli et al excluded from their study patients with previous gastrointestinal consultations, patients with symptoms lasting for more than one year, patients who had recently been taking anti-inflammatory drugs and antibiotics. Patients with alarming symptoms, such as weight loss, vomiting and bleeding were also excluded of the study. They evaluated the presence of *Giardia* infection in patients with IBS via (i) direct search of parasites in duodenal biopsies, (ii) histological examination of duodenal biopsies and (iii) parasitological evaluation of stool samples. They showed that 6.5 % of these patients with IBS also had giardiasis. A significant association between giardiasis and *H. pylori* infection was also found in the 137 patients selected for this study. Indeed, 14 % of patients diagnosed with *H. pylori* had giardiasis while 1% of patients without *H. pylori* had giardiasis. This suggests that the two infections might share some risk factors such as a common route of transmission (Grazioli et al., 2006, Moreira et al., 2005). These results showed that *Giardia* may be a common cause of IBS-like symptoms.

1.3.2.2/ Epidemiology

Giardia is one of the most widespread protozoa causing diarrhoea in the world with approximately 200 million symptomatic individuals, and 500,000 new cases reported each year (WHO, 1996). *Giardia* is especially common in areas where poor sanitary conditions and defective water treatment prevail. In the United Kingdom (UK), the United States (US) and Mexico, giardiasis seasonality has been shown to have a peak incidence during late summer. However, no seasonal pattern has been observed during day care outbreaks (Ortega and Adam, 1997). This waterborne pathogen has a prevalence of 2 to 5 % and 20 to 30 % in industrialised and developing countries respectively. Several studies have shown that children were more frequently infected than adults, particularly those

malnourished. The prevalence of *Giardia* infection in infants, and children under 10 years old, vary between 15 to 20 % in developing countries (Noor Azian et al., 2007, Ortega and Adam, 1997).

Recent studies using PCR suggest that the proportion of reported giardiasis represents 5% of common waterborne diseases; and that campylobacteriosis and salmonellosis are the most common waterborne diseases in Europe with proportions of 55% and 35% respectively (Fig 1.4 A) (ECDC, 2013, ECDC, 2012). These studies also show that reported cases of *Giardia* infection accounted for 8% of all reported cases of gastro-intestinal infections in the UK (Fig 1.4 B.) (PHE, 2013, ECDC, 2013, ECDC, 2012). In the US, similar studies show that approximately 20 % of all the reported cases of common gastro-intestinal diseases between 2004 and 2013 were giardiasis (Fig 1.4 C) (CDC, 2013). Giardiasis appears to have a very low incidence rate in European countries where *Giardia* infection were reported and monitored (Fig 1.5). The highest and lowest incidence rates were observed in Bulgaria, with an incidence rate of 234.1 reported cases per million per year, and in Lithuania, with an incidence rate of 0.843 reported cases per million per year, respectively (Fig 1.5) (ECDC, 2013, ECDC, 2012, ECDC, 2010a). When compared to the other European countries with an incidence rate for giardiasis, the UK has one of the highest incidence rates observed, with 58.9 reported cases per million per year (Fig 1.5) (ECDC, 2013, ECDC, 2012, ECDC, 2010a).

Assemblage A and B are associated with human infections; their distribution varies considerably among studies and within the same country. Feng and Xiao reviewed the analysis of over 4,000 human isolates, from different geographical areas, by PCR amplification of DNA extracted from faeces. According to them, assemblage B appears to be slightly more common in both industrialised and developing countries than assemblage A, in the case of endemic giardiasis. Indeed, 708 cases in developed countries and 1,589 cases in developing countries have been reported as assemblage B infections; for assemblage A, 482 and 1,096 cases have been reported in developed and developing countries respectively (Feng and Xiao, 2011). Even with the data summarised in this chapter, there is only a small amount of data available to epidemiologically link *Giardia* genotypes/subtypes – in source and/or drinking water- and human endemic giardiasis. Data on human

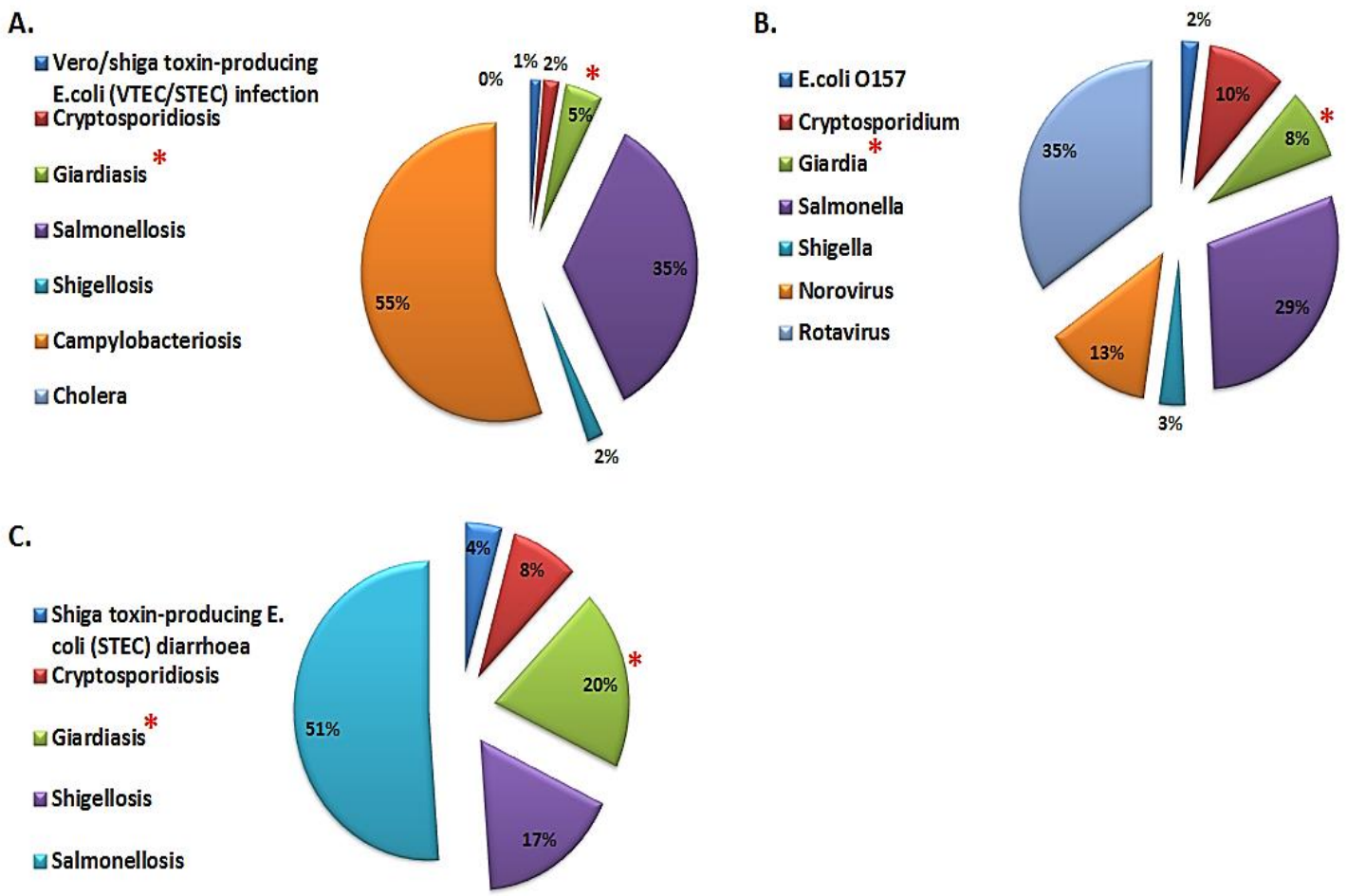


Fig 1. 4: Proportion of giardiasis and other common gastrointestinal diseases in Europe and the US and proportion of *Giardia* infection in the UK ***Giardia*** (ECDC, 2013, ECDC, 2012, CDC, 2013, PHE, 2013). **A.** Proportion of waterborne diseases in Europe (2006-2011); **B.** Proportion of common gastrointestinal infection in the UK (2000-2012); **C.** Proportion of common gastro-intestinal diseases in the USA (2004-2013).

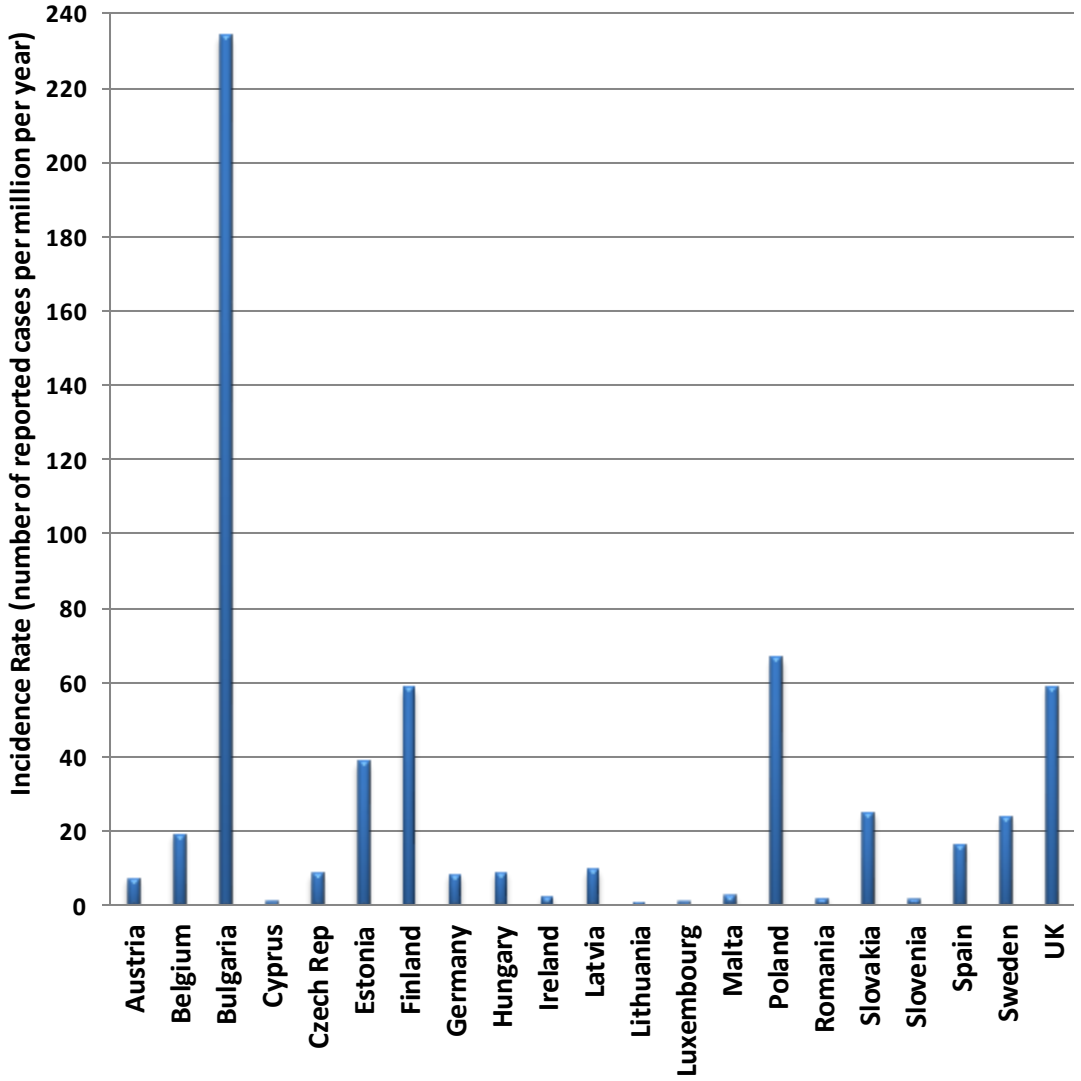


Fig 1. 5: Giardiasis incidence rate per country in Europe between 2006 and 2011 (ECDC, 2013, ECDC, 2012, ECDC, 2010a). Incidence rates for giardiasis were calculated via the following formula: $\frac{Number\ of\ cases\ arising\ in\ the\ country\ over\ 6\ years}{Average\ population\ number\ of\ the\ country\ between\ 2006-2011}$, for each European country with available data on number of *Giardia* infection between 2006 and 2011. Incidence rates were expressed in Number of reported cases per million per year.

giardiasis epidemiology in outbreaks are very limited. However, a systematic review analysed 14 publications on *Giardia* sporadic infections, outbreaks and case-control study from 22 countries and suggested that assemblage B was more present, during outbreaks, than assemblage A or assemblages A/B co-infection, with 60, 35 and 5% respectively (Caccio and Ryan, 2008). Even if assemblage B seems overall more present that assemblage A, no strong conclusions can be yet drawn from current data available. The fact that genetic exchange was shown to be possible *in vitro* suggests that *Giardia* may be capable of genetic recombination among member of the species complex (Caccio and Ryan, 2008, Poxleitner et al., 2008, Cooper et al., 2007, Ramesh et al., 2005); as a consequence, its virulence may vary according to the genetic exchange occurring during infection. Many questions still remain unanswered and more investigations are needed to understand *Giardia* infection epidemiology.

Giardiasis is associated with wide symptomatology, ranging from the absence of symptoms to acute or chronic diarrhoea, malabsorptive and allergic manifestations, weight loss, vomiting and childhood failure to thrive; but whether the severity of the disease is influenced by host, parasite, or host-parasite factors remains still largely unknown (Feng and Xiao, 2011, Caccio et al., 2005). The study of the association between *Giardia* assemblages and virulence has produced some inconsistent and contradictory results so far. Some studies have linked assemblage A to mild intermittent diarrhoea and assemblage B to severe, acute or persistent diarrhoea (Feng and Xiao, 2011, Caccio and Ryan, 2008, Caccio et al., 2005, Homan and Mank, 2001). Other studies have linked assemblage A to severe giardiasis and assemblage B to asymptomatic to mild giardiasis (Feng and Xiao, 2011, Read et al., 2002). Another study, realised in south-west London between 1999 and 2005, has shown similar epidemiological and clinical features. But differences between both assemblages, in term of fever intensity, have been observed; assemblage A being more frequently associated with fever than assemblage B, and yet, assemblage B is predominant in the urban UK setting (Breathnach et al., 2010). At present, data available for assemblage virulence are inconclusive (Nash, 2013, Feng and Xiao, 2011, Breathnach et al., 2010, Caccio and Ryan, 2008, Caccio et al., 2005, Homan and Mank, 2001). The exact contribution of *Giardia* genetic variability to

symptomatology and virulence is still under debate and requires further investigation.

Mixed infections have also been reported in several studies and countries, the percentage of mixed infections ranges from 2 % to 21 % (Gelanew et al., 2007). Most of these infections are assemblage A/B; some studies have identified assemblages A or B with C or D (canine-specific) infections (Traub et al., 2009). Mixed infections of assemblages A and F (cat-specific) have been identified in humans in Ethiopia (Gelanew et al., 2007). However, none of these results could be confirmed by analysis of either *tpi* or *ssu rRNA* genes (Feng and Xiao, 2011, Traub et al., 2009, Caccio and Ryan, 2008, Gelanew et al., 2007). These mixed infections are endemic and have only been observed in small local communities.

1.3.3/ Control and prevention

Nine national and international institutions: the Centre for Disease Control and Prevention (CDC), the World Health Organization (WHO), the New Zealand Ministry of Health (NZMH), the National Health Service (NHS), the National Travel Health Network and Centre (NaTHNaC), the "Institut de Veille Sanitaire" (the French Centre for disease control, InVS), the Canadian Public Health Agency (PHAC), the Manitoba Public Health Branch (MPHB), and the European Centre for Disease control (ECDC), were consulted to establish the main recommendations for the control and prevention of giardiasis. In this survey, 9 out of 9 showed the same recommendations for the prevention and control of giardiasis (WHO, 2013, NaTHNaC, 2013, MPHB, 2013, NZMH, 2013, InVS, 2012, ECDC, 2010b, CDC, NHS, PHAC). The most important recommendation is to maintain general good hygiene by washing hands before and after handling food, after changing nappies and using the toilet. Raw food such as fruits and vegetables should be washed with uncontaminated water; it is also suggested to avoid faecal exposure during sexual activity. Potentially contaminated water such as recreational bodies of water (swimming pool) and untreated drinking water (shallow wells, lakes, rivers, ponds, streams) should not be swallowed (Lebwohl et al., 2003).

Potentially contaminated water can be treated by boiling it for longer than one minute or by filtering it through a pore size of one µm (or smaller). Public water

supplies are regulated and treated according to standard protocols; which vary from one country to another according to their own laws and regulations.

1.3.4/ Clinical diagnosis and typing

The identification of *Giardia* can be achieved either by using classical diagnostic methods such as serological and microscopic techniques or by using molecular techniques for diagnosis and genetic analysis.

1.3.4.1/ Classical diagnostic methods

- **Microscopy**

The conventional detection method of *Giardia* cysts is microscopy; and the microscopic examination of stool is the most established diagnostic technique (Koehler et al., 2013, Lebwohl et al., 2003). Other biological samples such as duodenal, tissue, environmental and or/water samples are also examined via microscopy in order to detect and identify *Giardia* (Schuurman et al., 2007, Behr et al., 1997, Sauch, 1985). Some morphological features of the trophozoite such as median body shape and location, shape, ventro-lateral disc length, length of caudal flagella, and number and shape of nuclei, can be used for the identification of *Giardia*. It is however impossible to identify trophozoite or cysts to species level via light microscopy; staining techniques can however be used to help detect these *Giardia* stages and differentiate them from other micro-organisms and environmental debris present within samples (Koehler et al., 2013, Adam, 2001).

Simple staining protocols can be performed, such as Lubol's iodine staining, iron-haematoxylin, Giemsa and trichrome (Zhang et al., 2012, Arbo et al., 2006, Kellogg and Elder, 1999, Shetty and Prabhu, 1988); or cysts can be concentrated using, for example, formalin-ether (Koehler et al., 2013, Wegayehu et al., 2013). Motile trophozoites can also be detected by direct microscopy of fresh samples, smears being immediately prepared with warm (37°C, saline solution. Dead trophozoites can be detected by air-dried faecal smears stained with Giemsa. Due to the fact that *Giardia* infection is localised in the upper part of the gastrointestinal tract, cyst excretion is intermittent and multiple successive faecal samples analysis is required for a period of 1 to 2 weeks (Koehler et al., 2013).

❖ **Immunological tools**

In the detection of *Giardia* cysts and trophozoites, the use of immunological techniques can be advantageous over microscopy. Fluorescence microscopy can also be used to detect *Giardia* cyst and trophozoites. Indeed, fluorescence microscopy and direct fluorescence antibody (DFA) such as a fluorescein isothiocyanate-conjugated anti-*Giardia* antibody recognising surface epitopes, such as 65 kDa *Giardia* antigen, on cysts have been shown to be very efficient, with a sensitivity and specificity average ranges between 85% to 100% and 99 to 100 % respectively for the detection of cysts in faecal smears and environmental samples (Table 1.3) (Koehler et al., 2013, Rimhanen-Finne et al., 2007, Aldeen et al., 1998).

The detection of *Giardia* antigens in faecal samples, coproantigens, is another way of detecting *Giardia* infections. Enzyme-linked immunoassays such as enzyme-linked immunosorbent assays (ELISAs) have been used, and have shown high specificity and sensitivity with a range between 63 and 100 % (Aldeen et al., 1998, Behr et al., 1997). Immuno-chromatographic (IC) tests have also been achieved and have given variable specificity and sensitivity depending on study, between 79 to 100 % and 26 to 100 % respectively (Elsafi et al., 2013, Goni et al., 2012, Costache et al., 2009, Garcia and Garcia, 2006, Pillai and Kain, 1999). However, a study determined that the sensitivity of coproantigen assays could be lower than microscopic approaches (Johnston et al., 2003). Most IC tests are easy to implement and are usually a combined set of several antibodies to test for different parasites. This could affect the specificity of the test as cross-reactions may occur when testing for several pathogens (Koehler et al., 2013). Variation in the different assays diagnostic performance may also be due to intermittent shedding of cysts and *Giardia* antigens. A decreased sensitivity might be correlated to a low numbers of cysts in faeces or trophozoites in the host intestine (Strand et al., 2008). In their study, Strand et al suggested that coproantigen methods are useful in an outbreak setting but not for diagnosis of patients with refractory or persistent giardiasis following treatment. Yet, coproantigen detection methods have the capacity to detect infections prior to the cysts excretion in the host faeces; which makes them one the best choice for a cost-effective and rapid screening of large numbers of

Table 1. 3: List of antibodies available for the identification of *Giardia* via immunological techniques from the main antibodies supplying companies in the UK

Primary Antibody	Target	Techniques	Company
Mouse antibodies			
anti-65kDa *Giardia* antigen	*Giardia* cysts	ELISA, ICC[a]/IF[b], WB[c]	AbCam, Thermo Scientific, Santa-Cruz Biotechnology, Stratech
anti-*Giardia* external surface	*Giardia* Portland 1 strain, including its flagellae	IHC-P[d]	AbCam
anti-*Giardia* cyst	*Giardia* cysts	ELISA, ICC/IF	AbCam
anti-*Giardia*	*Giardia* cysts	IF, WB	Santa-Cruz Biotechnology
anti-*Giardia*	*Giardia* cysts	ELISA, ICC/IF, WB	Thermo Scientific
anti-*G. muris* trophozoite	*G. muris* trophozoites	IF	Merck Millipore
Goat antibodies			
anti-*Giardia* antibody (FITC)	Purified intact *Giardia* cysts	ICC/IF, IHC-Fr[e]	AbCam
anti-*Giardia* antibody	Purified intact *Giardia* cysts	ELISA, ICC/IF	AbCam, Thermo Scientific

[a]Immunocytochemistry
[b]Immunofluorescence
[c]Western blot
[d]Immunohistochemitry on paraffin
[e]Immunohistochemistry frozen sections

faecal samples (Koehler et al., 2013, Johnston et al., 2003); but, like widely used diagnostic techniques, they do not currently allow for the identification and characterisation of *Giardia* species or genotypes.

1.3.4.2/ Molecular techniques

Since classical diagnostic methods cannot be used to distinguish between different species and assemblages of *Giardia*, various nucleic acid-covalent techniques have been developed to identify and characterise *Giardia* and the genetic variations within and amongst *Giardia* species and assemblages. Some molecular methods rely on specific *in situ* hybridisation of probes to particular genetic loci within *Giardia* genome, but most rely on the specific amplification of one or more loci from small quantities of genomic DNA by PCR. Such studies have contributed to a better understanding of *Giardia* biology, epidemiology and population genetics (Feng and Xiao, 2011, Ankarklev et al., 2010, Adam, 2001). Comparative studies between microscopic, enzyme immunoassays and PCR have shown that molecular techniques were highly sensitive and specific with a range between 80 % and 100 % (Elsafi et al., 2013, Goni et al., 2012).

- **DNA fingerprinting**

PCR-fingerprinting coupled methods depend on the screening of a genome in search for neutrally evolving repetitive DNA, and also for variation in organisation and sequence length. These techniques do not necessarily require prior knowledge of the genome or genes to characterise a parasite, which can be advantageous. However, the genetic fingerprint of *Giardia* cyst or trophozoite isolates represents a population of organisms rather than an individual. The random amplification of polymorphic DNA analysis (RAPD) has been applied to *Giardia* because of its ability to detect variation within and among populations of a particular species (Pelayo et al., 2003, Welsh and McClelland, 1990, Williams et al., 1990). In RAPD analysis, only a single arbitrary primer is used instead of two sequence-specific primers (Deng and Cliver, 1999). At first, RAPD was considered useful due to its ability to amplify from small amounts of genomic DNA and its capacity to rapidly screen genomes for variation (Hadrys et al., 1992). But, some significant issues with the specificity and

reproducibility of RAPD were found because of the low stringency often used in PCR (Perez et al., 1998). Also the fact that a single arbitrary primer is used may give false positive results due to the techniques susceptibility to contaminant-organism or foreign DNA; and is therefore recommended for *Giardia* only with axenic isolates grown in vitro (Deng and Cliver, 1999).

Minisatellites display allelic hypervariability due to differences in repeat copy number within alleles, hence variation in size (Jeffreys et al., 1985). They have been shown to be a very useful tool in the discrimination between *Giardia* individual strains and in epidemiological analysis (Carnaby et al., 1995, Carnaby et al., 1994, Upcroft et al., 1990). The use of the human polycore minisatellite probes 33.6 and 33.15 and the M13 phage DNA probe were shown to allow the identification of a range of putative hypervariable minisatellites in *Giardia* (Carnaby et al., 1995, Upcroft et al., 1990). Even with such promising results, the study of *Giardia* minisatellites to identify between strains has surprisingly been neglected over the last 20 years. The development and use of minisatellites probes for *Giardia* may give more insights on *Giardia* taxonomy as its taxonomy is still under debate. Indeed, fingerprinting minisatellites may allow identifying whether *Giardia* assemblages are different species or not; and therefore, it may also allow a better discrimination on a sub-assemblage/species level.

❖ **Restriction fragment length polymorphism-, specific-, real time-, multiplex PCR and sequencing**

Discrimination of *Giardia* species and assemblages are based on single-nucleotide polymorphisms (SNPs). PCR techniques using specific primer pairs for a selective amplification of different genetic loci, followed by enzymatic cleavage or sequencing, are employed. The *Giardia* specific key markers used for this are:

- Genes coding for β-giardin (*bg*), elongation factor 1-α (*ef1-α*), triose-phosphate isomerase (*tpi*), glutamate dehydrogenase (*gdh*), and variable surface proteins (*vsp*)
- The *Giardia lamblia* open reading frame-C4 (GLORF-C4)
- Nuclear ribosomal RNA genes such as 18S rRNA and internal transcribed spacer 1 and 2 (ITS-1, -2)

SSU sequence can also provide useful genetic markers for specific identification of *Giardia* as it has relatively low intraspecific and relatively high interspecific sequence variations (Koehler et al., 2013, Feng and Xiao, 2011).

Additional markers, used to achieve assemblage or sub-assemblage identification based on specific PCR, RFLP or sequencing are present in the *tpi*, *bg*, and *gdh* genes (Monis et al., 1999). These genes have been used in systematic investigation of *Giardia* to provide the basis of the current classification of members within the *Giardia* genus (Feng and Xiao, 2011, Caccio and Sprong, 2010, Sprong et al., 2009). In addition, ITS of ribosomal DNA (rDNA) are also useful for assessing genetic variability within *Giardia* species due to their higher intraspecific variation compared to the rRNA gene regions examined thus far (Caccio and Sprong, 2010).

PCR-based restriction fragment length polymorphism (PCR-RFLP) approach is sometimes also used for the genetic analysis and classification of *Giardia* (Robertson et al., 2007). Although this approach has been proven useful, it does not detect all sequence and length variations within and between amplicons. This is due to the fact that the restriction enzymes used only recognise a small number of potentially variable sites. The problems of this approach can also include an incomplete enzymatic digestion of amplicons and/or sequence variations within individual amplicons leading to a misinterpretation of the RFLP profiles obtained (Koehler et al., 2013). The “gold-standard” for detecting genetic variation and polymorphism remains DNA sequencing and comparative analysis; which can be applied to single-copy as well as multi-copy genes; with the condition that there is sequence homogeneity between copies of target genes (Sprong et al., 2009). Moreover, DNA sequence data analysis is suitable for direct, comparative genetic and phylogenetic studies at any time. However, some data have shown some incongruity in genotyping results when different genetic loci are targeted. Indeed, the difficulty in assigning some isolates, predominantly for specimen from humans and dogs, has been raised (Feng and Xiao, 2011, Sprong et al., 2009). The high prevalence of mixed infections may be one factor contributing to the incongruity observed. To overcome this issue, the multi-locus genotyping (MLG) approach has been increasingly used to characterise *Giardia*. Yet, the grouping of the parasites

into individual MLGs has been further complicated by the fact that many isolates may exhibit double peaks at specific positions in electropherograms. In 2010, Lebbad et al observed double peaks in electropherograms in assemblages B, C, D and E isolates but never in assemblages A, F and G isolates. This suggests that MLG approach may be more useful for the typing of assemblage A isolates (Lebbad et al., 2010, Caccio et al., 2008).

Real time quantitative PCR (q-PCR) enables the monitoring of DNA amplification as it occurs *in vitro* (Koehler et al., 2013). This PCR can be combined with the high resolution melt (HRM) approach which allows the detection of mutations, polymorphisms and epigenetic differences in double-stranded DNA (dsDNA) (Koehler et al., 2013). This approach has been used for the specific detection of *Giardia* species or assemblages (Zhang et al., 2012, Almeida et al., 2010, Haque et al., 2007, Bertrand et al., 2004). However, HRM analyses may not always be able to delineate multiple, distinct types of sequence, such as paralogues, within samples; particularly if the melting temperature or melting profiles of such sequences are similar, which can induce merging of peaks resulting in an uninterpretable profile (Koehler et al., 2013). Yet, the high-throughput potential and relatively low cost of HRM analysis as well as its ability to differentiate between *Giardia* isolates, make the HRM combined q-PCR approach a suitable tool for the identification and differentiation of *Giardia* species and population variants if suitable genetic markers are used.

Multiplexed tandem-PCR (MT-PCR), which combines a primary phase of target loci enrichment with an analytical phase using primary amplification as template, and Luminex, which used fluorescently dyed micro-beads for large scale genotyping, approaches have also been shown to have a great potential for medium to large scale epidemiological studies of *Giardia* and other waterborne pathogens (Jex et al., 2012, Taniuchi et al., 2011, Stark et al., 2011). Multiplex PCR consists in the amplification of multiple genetic loci, representing either the same organisms or multiple organisms, in a single reaction. Although useful, multiplex PCRs can raise significant problems due to potential cross-hybridisation or competition between multiple primer sets which could result in poor PCR amplification and/or the generation of non-specific products (Koehler et al., 2013,

Elnifro et al., 2000). The development of MT-PCR addresses some of the limitations of multiplex PCR. MT-PCR consists of two amplification phases: (i) a primary phase to enrich target loci using multiplexed primers in a few PCR cycles (10-15 cycles), each primer pair targets a distinct genetic locus which is coupled to (ii) an analytical phase where a highly diluted product from the primary amplification is used as a template (Jex et al., 2012). During this analytical phase, the targeted amplification is performed in tandem for each genetic locus using a specific nested-PCR rather than by multiplex (Jex et al., 2012). A comparative study showed that MT-PCR exhibited 100 % sensitivity and specificity compared to the "gold-standard" that is microscopy (Stark et al., 2011). The micro-beads used in the Luminex approcah are internally dyed with fluorescent dyes to produce specific spectral address (Koehler et al., 2013, Taniuchi et al., 2011). The beads high-density digital coding allows the tracking of the target organisms but also of other important identifiers such as sample code and type, geographical origin, collection date, and reagent kits, which could provide a robust detection approach for multiplex assays (Koehler et al., 2013). This should enable the development and implementation of effective water and environment surveillance programmes. Moreover, a recent study combining Luminex assays with *Taq*Man array card (TAC) system allowed the detection of 19 enteropathogens, including *Giardia*, with a specificity and sensibility of 96 % and 98 % respectively suggesting that combining molecular and bioinformatics techniques may be the answer to a better and faster diagnostic of *Giardia* infection and giardiasis (Liu et al., 2013).

- **Genomic comparison**

Most of the *Giardia* genotyping studies have relied on the sequence characterisation of human and animal-derived cysts and trophozoites using only SNPs (Caccio and Ryan, 2008, Caccio et al., 2005). However, some data showed inconsistent genotyping results when different loci were targeted (Caccio et al., 2008). The development of genomics tools for *Giardia* has provided new insights on assemblages A and B whole genomes and genes function (Adam et al., 2013, Perry et al., 2011, Franzen et al., 2009, Morrison et al., 2007). These studies have also highlighted assemblage-specific genes which could potentially be good targets for a

faster characterisation of *Giardia* during an infection. In addition, genomics approach is a fast and low cost technique as it doesn't require the amplification of DNA sequences to compare different samples and data are processed through database via bioinformatics tools.

- **Fluorescent *in situ* hybridisation (FISH)**

Fluorescently labelled oligonucleotide probes are used to specifically detect target sequences within RNA or DNA. Most FISH assays used for cyst detection have relied on the hybridisation of the RNA rather than the DNA of a cell; targeting a variable region of the small subunit (*SSU*) of rRNA. This is due to the fact that rRNAs are transcribed to high abundance within the cell. (Bednarska et al., 2007, Lemos et al., 2005). It has been suggested that only viable cysts would be detected due to RNA degradation after cell death; thus, rRNA would not be detectable in dead cysts. This makes FISH probes of use to study the cyst viability in faecal or environmental samples (Dorsch and Veal, 2001). However, the use of FISH probes for cyst viability assays is dependent on the rate of target decay. It has been assumed that this decay is fast upon cell death, but this remains to be quantitatively assessed for rRNA. It is likely that the rate of degradation will vary depending on different environmental conditions such as temperature, pH or potential RNase contamination due to cellular damage.

Several studies have demonstrated the utility of FISH probes as a dependable method to detect cysts, but most of the probes used to date have not yet allowed for reliable discrimination between species or assemblages (Koehler et al., 2013). Nonetheless, in one study probes to the rRNA *SSU* were able to discriminate between *Giardia muris*, *Giardia ardeae* and *Giardia lamblia* (=*duodenalis*/*intestinalis*) (Erlandsen et al., 2005).

1.3.5/ Treatment

Giardia infections can remain asymptomatic or induce severe and/or chronic giardiasis, in which case treatment is generally indicated (Gardner and Hill, 2001). Recent studies have shown strain-specific pathogenicity of different genotypes and possibly some sub-genotypes in rodent models (Solaymani-

Mohammadi and Singer, 2011, Benere et al., 2011). Yet, in humans, association between strain-specific genotypes and clinical symptoms have been inconclusive so far (Hahn et al., 2013, Robertson et al., 2010). Six main classes of agents are currently used to treat giardiasis (i) Nitroimidazoles, (ii) Quinacrine, (iii) Furazolidone, (iv) Benzimidazoles, (v) Paromomycin, and (vi) Bacitracin zinc (Gardner and Hill, 2001).

1.3.5.1/ Nitroimidazoles

The nitroimidazoles, in particular metronidazole (MTZ), were discovered in 1955 and were found to be highly effective against several protozoan such as *Trichomonas vaginalis* or *Entamoeba histolytica* (Durel et al., 1960) (Fig 1.6 A.). MTZ [1-(β-hydroxyethyl)-2-methyl-5-nitroimidazole; Flagyl] was reported to be effective against giardiasis in 1962 (Darbon et al., 1962, Durel et al., 1960). MTZ current model of action suggests its intracellular reduction to a toxic radical form via *Giardia* anaerobic metabolic pathway and by several parasites reductases, including pyruvate: ferredoxin oxidoreductase (PFOR), nitroreductase and thioredoxin reductase (Nillius et al., 2011, Leitsch et al., 2011). Indeed, pyruvate: ferredoxin proteins from the parasite donate electrons to the drug's nitro group; after reduction of this nitro group, MTZ becomes "activated", and a gradient allowing MTZ intracellular transport is established (Hahn et al., 2013, Gardner and Hill, 2001). Once activated, the reduced MTZ binds to the parasite DNA macromolecules which results in DNA damage in the form of loss of helical structure, impaired template function and strand breakage. This DNA damage results in the death of the trophozoite (Gardner and Hill, 2001). In addition, MTZ also inhibits trophozoite respiration (Upcroft and Upcroft, 1998, Paget et al., 1989). Since its discovery, MTZ and other nitroimidazoles (such as tinidazole and ornidazole) have been the first choice for the treatment of giardiasis.

MTZ is readily and completely absorbed after oral administration, it penetrates body tissues and secretions such as saliva, breast milk or semen and vaginal secretion. MTZ is mainly metabolised in liver and excreted in the urine (Gardner and Hill, 2001, Lau et al., 1992). Since their discovery, MTZ and other nitroimidazoles (tinidazole, secnidazole, benzoyl MTZ, nimorazole and ornidazole)

have been the first choice for the treatment of giardiasis. However, patients treated with nitroimidazoles can show a wide range of mild to severe side effects (Table 1.4). The most common side effects of MTZ treatment include headache, metallic taste, vertigo, nausea. In addition, pancreatitis, central nervous system toxicity, at high doses, and transient reversible neutropenia (abnormally low number of neutrophils) have all been attributed to MTZ (Gardner and Hill, 2001, Harris et al., 2001, Thompson et al., 1993). The other nitroimidazoles used in the treatment of giardiasis have been shown to provoke mild side effects such as nausea, anorexia, and abdominal pain (Harris et al., 2001).

1.3.5.2/ Quinacrine

Quinacrine (Mepacrine/Atabrine) was first introduced as antimalarial drug in 1930; it became an important agent against *Giardia* after World War II (Gardner and Hill, 2001) (Fig 1.6 D.). Quinacrine has been proven to be very effective by different studies, and to have one advantage over nitroimidazole drugs as it also kills cysts as a decrease excystation was observed in both *in vitro*- and patient-derived cysts (Harris et al., 2001, Gillin and Diamond, 1981). The antiprotozoal mechanism of action of quinacrine has yet to be fully elucidated; however several alternative hypotheses have already been suggested. Firstly, this drug is thought to interfere with flavin compounds of enzymes such as NADH oxidase, which could decrease the oxygen consumption by the parasite (Paget et al., 1989). Secondly, quinacrine is a known cholinesterase inhibitor and therefore could be involved in the inhibition of *Giardia* acetylcholine pathway. Finally, quinacrine is thought to intercalate with the parasite DNA, thereby inhibiting the nucleic acid synthesis (Gardner and Hill, 2001, Thompson et al., 1993).

Quinacrine is rapidly absorbed from the intestinal tract and widely distributed in body tissues. Quinacrine is clinically very effective against *Giardia*; some studies have indeed shown an average 95 % efficacy over 5 to 10 days (Wolfe, 1992, Kavousi, 1979). Yet, this drug has been reported to have a very wide range of adverse effects from a bitter taste, vomiting and nausea to more severe symptoms such as blood dyscrasias (pathologic conditions or disorders in which the blood's constituents are either abnormal or present in abnormal quantity), toxic psychosis,

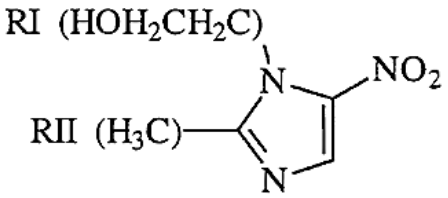
insomnia, myalgia, or haemolysis on glucose-6-phosphate dehydrogenase deficient patients (Harris et al., 2001, Gardner and Hill, 2001) (Table 1.4).

1.3.5.3/ Furazolidone

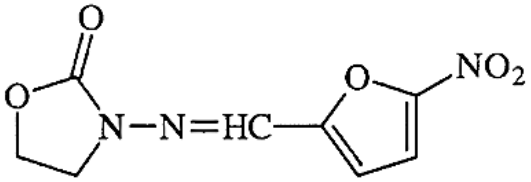
Furazolidone is a member of the nitrofuran family which was discovered in the 1940s (Fig 1.6 B.). Furazolidone is effective against many bacterial infections, including *Klebsellia spp*., *Clostridium spp*., *Campylobacter spp*., *Staphylococcus aureus* and *E. coli* (Harris et al., 2001, Gardner and Hill, 2001); and in 1960, furazolidone was reported to be an effective antigiardial agent (Webster, 1960). As for Quinacrine, forazolidone mechanism of action is not completely explained and requires further investigation. It is thought to undergo reductive activation in *Giardia* trophozoite, like metronidazole (MTZ). However, instead of being reduced by ferredoxin I, furazolidone may be reduced by NAD(P)H oxidases as well as PFOR. This is mainly due to the fact that nitrofurans have a more positive redox potential than nitroimidazoles (Brown et al., 1998, Upcroft and Upcroft, 1998). Once reduced, forazolidone is thought to damage important cellular components including DNA. In one study, furazolidone was shown to act on two specific stages of *Giardia* cell cycle, S and G2+M; which indicates an inhibition of DNA synthesis and of the cell cycle completion (Hoyne et al., 1989).

Furazolidone is promptly absorbed via the gastrointestinal tract and metabolised by body tissues; which leads to low concentrations in serum and urine. According to a few studies, about 10 % of patients treated with furazolidone reported gastrointestinal symptoms such as nausea, vomiting, diarrhoea, malaise. Other side effects can also include a brown discoloration of urine, haemolysis in glucose-6-phosphate dehydrogenase-deficient patients, urticaria, hypersensitivity, pruritus (itching sensation) (Gardner and Hill, 2001, Wolfe, 1992) (Table 1.4). The drug has also a monoamine oxidase (MAO) inhibitory effect and should not be given to individuals taking MAO inhibitors. Plus, some rare cases of disulfiram-like reaction were reported when furazolidone was taken with alcohol (Gardner and Hill, 2001).

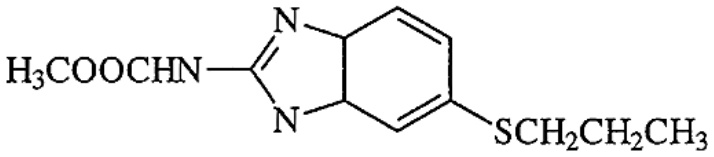
A. Metronidazole



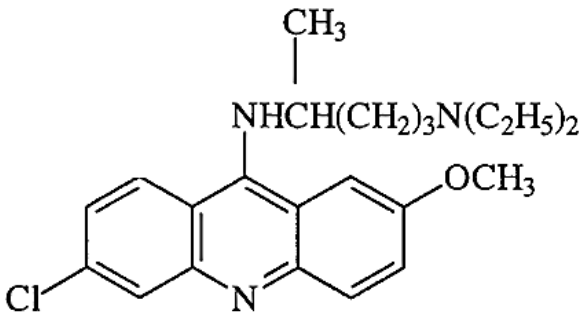
B. Furazolidone



C. Albendazole



D. Quinacrine



E. Paromomycin

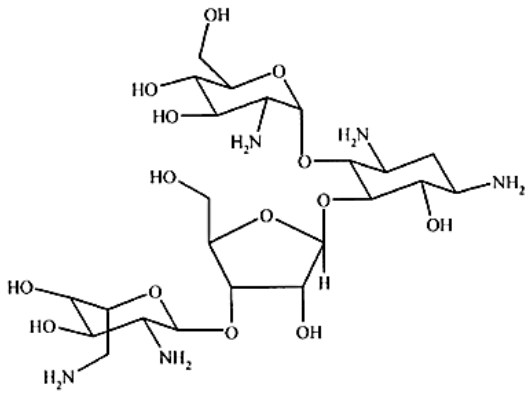


Fig 1. 6: The four most used antigiardial drugs (modified from Harris et al, 2001 and Upcroft and Upcroft, 1998 with permission). **A.** Metroinidazole, member of the Nitroimidazole family drugs and main way of treatment of giardiasis. **B.** Furazolidone, a nitrofuran compound used in treatment of many bacterial infections and giardiasis. **C.** Albendazole,member of the Benzimidazole family drug, used in treatment of many parasitic worm infections and giardiasis. **D.** Quinacrine, a substituted acridine and antimalarial drug also effective against *Giardia*. **E.** Paromomycin (or humatin) member of the aminoglycoside family, indicated for the treatment of *E. hystolitica*, *T. vaginalis* and *Giardia*.

1.3.5.4/ Benzimidazoles

Benzimidazoles are bicyclic aromatic compound formed by the fusion of benzene and imidazole. Their chemotherapeutic property against helminths and parasites was established in the 1950s. Thousands of benzimidazole molecules were synthetized but only a few had an effect on parasites. Amongst these agents, two were shown to have an effect on *Giardia* infections and used to treat giardiasis: albendazole (Fig 1.6 C) and mebendazole (Aguayo-Ortiz et al., 2013, Gardner and Hill, 2001). Benzimidazole agents exert their toxic effect on parasites by binding to the β-tubulin cytoskeleton, resulting in the disruption/inhibition of cytoskeleton polymerisation and impaired glucose intake (Gardner and Hill, 2001, Venkatesan, 1998). It has been postulated that the colchicine site in the microtubules may play a role in the benzimidazole binding to the giardial β-tubulin (Clark and Holberton, 1988, Lacey, 1988). In 1994, Katiyar et al suggested that the glutamic acid 198 (Glu 198) on the β-tubulin amino acid chain may be involved without being able to prove it. However, a recent study has demonstrated that this amino acid (AA) was involved in the benzimidazole-β-tubulin binding (Aguayo-Ortiz et al., 2013, Katiyar et al., 1994).

Benzimidazoles are poorly absorbed by the gastrointestinal tract, although their absorption can be improved by co-ingestion of a fatty meal. Compare to the other types of drugs used in the treatment of giardiasis, benzimidazoles have less side effects on patients (Table 1.4). During a short-term use, it may cause anorexia and constipation; and during a long-term use, a high dose of albendazole can cause reversible neutropenia and elevated hepatic enzyme levels (Gardner and Hill, 2001, Liu and Weller, 1996). Due to a possible teratogenicity, albendazole is contra-indicated in pregnancy.

1.3.5.5/ Paromomycin and bacitracin zinc

❖ **Paromomycin**

Paromomycin, also called humatin, is a member of the aminoglycoside family and is a broad-spectrum antibiotic (Fig 1.6 E.). Paromomycin was first isolated in 1956 and shown to be effective against *T. vaginalis* and *E. histolytica*

infections. It was also proposed as a treatment for *Giardia* in resistant infections and during pregnancy (Gardner and Hill, 2001, Kreutner et al., 1981). Paromomycin inhibits *Giardia* protein synthesis by interfering with the parasite's 50S and 30S ribosomal subunits; which causes the misreading of messenger RNA (mRNA) codons (Edlind, 1989).

Paromomycin is poorly absorbed from the intestinal lumen; indeed, it has been shown that even large-dose of oral administration achieve only minimal concentrations in the patients' blood and urine (Kreutner et al., 1981). However, due to this minimal absorption, Paromomycin is selectively toxic as its action depends on the unique giardial ribosomal RNA (rRNA) structure (Edlind, 1989). Via *in vitro* susceptibility testing, its activity against *Giardia* has been demonstrated to be lower than the activity of nitroimidazoles, quinacrines or forazolidones (Boreham et al., 1985, Gordts et al., 1985). Paromomycin can cause side effects such as ototoxicity and nephrotoxicity if absorbed systemically (Table 1.4). However, with limited systemic absorption, toxicity should not be a problem for patients with normal kidney function (Gardner and Hill, 2001, Harris et al., 2001).

- **Bacitracin zinc**

Bacitracin was first isolated in 1945 from a strain of *Bacillus*; it used systemically against severe staphylococcal infections until 1960 when its toxicity and the development of other antibiotics restricted its use. To promote stability into the bacitracin complex, zinc was added, giving it the name of bacitracin zinc (Gardner and Hill, 2001, Kucers et al., 1997). Following the *in vitro* demonstration of its action against *E. histolytica* and *Trichomonas*, bacitracin zinc was found to be active against *Giardia* (Andrews et al., 1995, Andrews and Bjorvatn, 1994, Andrews et al., 1994). Bacitracin acts on pathogens by interfering with their dephosphorylation step in cell membrane synthesis. Undecaprenyl pyrophosphate is an important molecule involved in the peptidoglycan synthesis. The tight binding of bacitracin to undecaprenyl pyrophosphate was shown to prevent undecaprenyl pyrophosphate hydrolysis. As a result, undecaprenyl monophosphate, the product of undecraprenyl pyrophosphate hydrolysis, is less available for binding with the UDP-peptidyl N-acetylmuramic acid (UDP-MurNAc) – pentapeptide complex to

initiate the second stage of the peptidoglycan synthesis suggesting that cell wall biosynthesis was inhibited (Ming and Epperson, 2002).

Some adverse effects such as diarrhoea, abdominal discomfort and nausea were observed in some patients. The use of bacitracin zinc can also lead to nephrotoxicity and gastrointestinal disturbance with prolonged oral administration (Gardner and Hill, 2001) (Table 1.4).

1.3.5.6/ Drug resistance and new chemotherapeutic agents

- **Drug resistance and relapse**

Besides the fact that all the antigiardial drugs can provoke adverse effects in patients (Table 1.4), treatment failures have also been reported for all of them. However, it is important to differentiate between drug resistance, cure followed by reinfection and post-*Giardia* lactose intolerance. If it is a reinfection (positive stool sample examination for *Giardia* antigen), re-infected patients should respond to original therapeutic agent used; reinfections are common in *Giardia* endemic regions around the world and in situation of poor faecal-oral hygiene. Therefore, risk factors should be assessed and guidance on hygiene and prevention should be given to re-infected patients. Disaccharidases are enzymes present in the intestinal lumen which cleave complex sugars such as lactose into simple sugars such as glucose (Bayless and Christopher, 1969). Deficiency in disaccharidase activity leads to an increase of complex sugars concentration within the intestinal lumen; this excess of complex sugar acts as osmotic loads resulting in an outpouring of fluid into the small intestine which increases the gastrointestinal motility (Bayless and Christopher, 1969). Post-*Giardia* lactose intolerance is the most common disaccharidase deficiency associated with giardiasis; it can occur in 20 to 40 % of patients with *Giardia* infection (Gardner and Hill, 2001). Thus, if the stool sample examination is negative for *Giardia*, patient will be advised to avoid lactose-containing food, and this syndrome may last for several weeks but will normally resolve eventually.

True treatment failure could mean that patients are infected with a drug-resistant *Giardia* isolate. Resistance to most antigiardial drugs has been

Table 1. 4: Recommended dosage and adverse effects of anti-*Giardia* drugs (modified from Gardner and Hill, 2001, with permission.)

Drug	Adult dose[a]	Paediatric dose	Adverse effects
Nitroimidazoles			
Metronidazole	250 mg t.i.d. x 5 - 7 days	5 mg/kg t.i.d. x 5 - 7 days	Headache, vertigo, nausea, metallic taste, urticaria Disulfiram-like reaction with alcohol ingestion Rare: pancreatitis, central nervous system toxicity, reversible neutropenia, peripheral neuropathy, T-wave flattening with prolonged use Mutagenic/carcinogenic?
Tinidazole	2 g, single dose	50 mg/kg , single dose (max, 2 g)	As for Metronidazole
Ornidazole	2 g, single dose	40 -50 mg/kg , single dose (max, 2 g)	As for Metronidazole
Acridine derivatives			
Quinacrine (Mepacrine/Atabrine)	100 mg t.d.i. x 5 - 7 days	2 mg/kg t.i.d. x 5- 7 days	Nausea and vomoting, dizziness, headache Yellow/orange discoloration of skin and mucous membranes Rare: toxic psychosis
Nitrofuran derivatives			
Furazolidone	100 mg q.d.i. x 5 - 7 days	2 mg/kg q.i.d. x 5- 7 days	Nausea, vomiting, diarrhoea Brown discoloration of urine; disulfiram-like reaction with alcohol ingestion Reacts unfavorably with MAO[b] inhibitors Mild hemolysis in G6DPH[c] deficency Carcinogenic?
Aminoglycoside			
Paromomycin	500 mg t.d.i. x 5 - 10 days	30 mg/kg/day in 3 doses x 5 - 10 days	Ototoxicity and nepthrotoxicity with systemic administration
Benzimidoles			
Albendazole	400 mg q.d. x 5 days	15 mg/kg/day x 5 - 7 days (max, 400 mg)	Anorexia, constipation Rare: reversible neutropenia and elevated liver function tests Teratogenic?
Bacitracin zinc	120,000 U b.i.d. x 10 days	Not tested in children under 10 years	Nausea, vomiting, abdominal discomfort Nephrotoxicity with systemic absorption

[a] q.d., once a day; b.i.d., twice a day; t.i.d., three times a day; q.i.d., four times a day

[b] Monoamine Oxydase

[c] Glucose-6-phosphate dehydrogenase

documented or studied *in vitro*; and different drug class susceptibilities for several genotypically different clones of *Giardia* found in the human duodenum were observed (Upcroft and Upcroft, 2001, Upcroft et al., 1999, Carnaby et al., 1994, Farthing, 1992). A consistent correlation between *in vitro* resistance and clinical failure or success has yet to be determined; which still leaves unanswered the question of true parasite resistance for *Giardia*. When clinically resistant strains have been identified, treatments with longer repeat courses or higher doses of the original drug are given to the patients (Gardner and Hill, 2001). Yet, the most successful way of eradicating *Giardia* seems to be the use of a drug from a different class to avoid potential cross-resistance. However, this may not be always effective e.g. Brasseur and Favennec reported that two of their patients failed albendazole treatment following two courses of MTZ but did respond positively to quinacrine (Brasseur and Favennec, 1995). In case of failure, the last option is to combine drugs together to increase the response from patients. For example, MTZ-albendzole, MTZ-quinacrine, or a nitroimidazole plus quinacrine; for courses of at least two weeks (Gardner and Hill, 2001).

- **New chemotherapeutic agents**

Most of the drugs used in the treatment of giardiasis have considerable side effects and are, most of the time, contraindicated (Gardner and Hill, 2001, Harris et al., 2001) (Table 1.4). Therefore, the study of new chemotherapeutic agents plays a fundamental role in the rationale for the treatment if *Giardia* infections.

In 2009, Muller et al reported *Giardia* protein disulfide isomerases (PDI), more specifically PDI2 and PDI4 and *Giardia lamblia* Nitroreductase-1 (GlNR-1), as promising targets for new drugs (Rossignol, 2010, Muller et al., 2007). A study has shown that GlNR-1 increases *Giardia* susceptibility to nitro drugs such as MTZ and albendazole; this study also suggests that GlNR-1 may play a role in the reduction of the nitro drugs used in the treatment of giardiasis (Nillius et al., 2011). In another recent study, *Giardia* has been shown to be highly sensitive *in vitro* to tetrahydrolipsatin, an anti-obesity drug also called Orlistat (Hahn et al., 2013). Orlistat is a derivative of the naturally occurring lipase inhibitor lipstatin from *Streptomyces toxytricini* (Hahn et al., 2013, Weibel et al., 1987). Orlistat was shown

to inhibit pancreatic lipases making it a good anti-obesity drug (Hahn et al., 2013, Borgstrom, 1988). Although lipid metabolism is crucial for rapid proliferation and survival of organisms, *Giardia* was shown to have restricted resources to synthesis and metabolise lipid suggesting that *Giardia* depends strongly on the exploitation of lipids supplied by the host environment (Das et al., 2002). According to Hahn et al, Orlistat may inhibit *Giardia*'s growth; it may also alter the parasite trophozoite's morphology and detachment, *in vitro.* Even with promising targets and drugs, *in vivo* model experimentations are required as *Giardia*'s susceptibility may vary according to the infective strain. There is also an important need of standardisation for the evaluation of antigiardial drugs.

1.4/ Biology of *Giardia*

During its life cycle, *Giardia* alternates between two different stages (Fig 1.7) (i) the infectious cyst stage, found in external environment (Fig.1.8 A.) and (ii) the trophozoite stage, found in host small intestine (Fig. 1.8 B.). *Giardia* infections are initiated by ingestions of cysts from contaminated water or food, or through direct faecal-oral contact (Carranza and Lujan, 2010, Adam, 2001, Adam, 1991).

1.4.1/ Cyst structure

The cyst is the infective form of *Giardia* and is resistant to the external environment conditions as well as to the gastric acid in the stomach of the infected host. The *Giardia* cyst contains four nuclei and is approximately 5 by 7 to 10 µm in diameter, and is covered by a wall of 0.3 to 0.5 µm thickness composed of an outer filamentous layer and an inner membranous layer with two membranes (Fig.1.8A). The outer portion of the wall is surrounded by a web of 7- to 20-nm filaments whereas the inner portion contains an outer and inner cyst wall membrane separated by a thin layer of cytoplasm (Adam, 2001, Adam, 1991, Erlandsen et al., 1990). Four major proteins (29, 75, 88 and 122 kilodaltons (kDa) in size) have been identified in the outer wall by the hybridization of cyst wall-specific antibodies to specific bands on Western immunoblots (Adam, 1991, Reiner et al., 1989). Immunogold electron microscopy has been used to localise these proteins to the filamentous structures of the outer portion of the cyst wall (Adam, 1991, Erlandsen

et al., 1990). Flagellar axonemes, vacuoles, ribosomes and fragments of the ventral disk are found inside the cyst wall

1.4.2/ Trophozoite structure

Trophozoites are non-invasive, and proliferate while attached to the epithelial cells of the gut. They are the primary cause of the clinical manifestation of the disease (Adam, 1991). Trophozoites replicate by vegetative growth in the intestine (Elmendorf et al., 2003). The trophozoite is between 12 to 15 µm long, 5 to 9 µm wide, and 1 to 2 µm thick (Carranza and Lujan, 2010, Adam, 2001) (Fig. 1.8 B.). *Giardia* trophozoite possesses two nuclei surrounded by nuclear envelopes that remain intact throughout the cell cycle; and has a complex cytoskeleton that maintains the shape of the parasite and anchors the four pairs of flagella (Fig. 1.8B), the median body, and the ventral disc (VD) (Carranza and Lujan, 2010, Elmendorf et al., 2003). The four pairs of flagella emerge ventrally, anteriolaterally, posteriolaterally, and caudally from the cell body. The flagella, median body, and VD have been proposed to be important for the attachment of trophozoites to enterocytes in the intestine (Carranza and Lujan, 2010, Palm et al., 2005, Elmendorf et al., 2003, Lujan et al., 1997).

From a structural perspective, *Giardia* trophozoites are simple. They possess some basic characteristics of eukaryotic cells such as the presence of nuclei with nuclear envelopes linked to the endoplasmic reticulum (ER), and lysosome-like peripheral vacuoles that are located underneath the plasma membrane (Carranza and Lujan, 2010, Elmendorf et al., 2003, Adam, 2001). However, they lack some typical eukaryotic organelles like peroxisomes, a morphologically evident Golgi apparatus, and classical mitochondria; although a mitochondrial remnant has been described and characterised (Tovar et al., 2003). *Giardia* has a well-developed ER and can form secretory vesicles (Gottig et al., 2006, Lujan and Touz, 2003, Lujan et al., 1997). In the cell cytoplasm and adjacent to the plasma membrane (except in the region of the ventral disc), there are numerous vesicles and tubules (peripheral vacuoles or PVs) which have been proposed to have an endosome and lysosome function (Carranza and Lujan, 2010, Gottig et al., 2006).

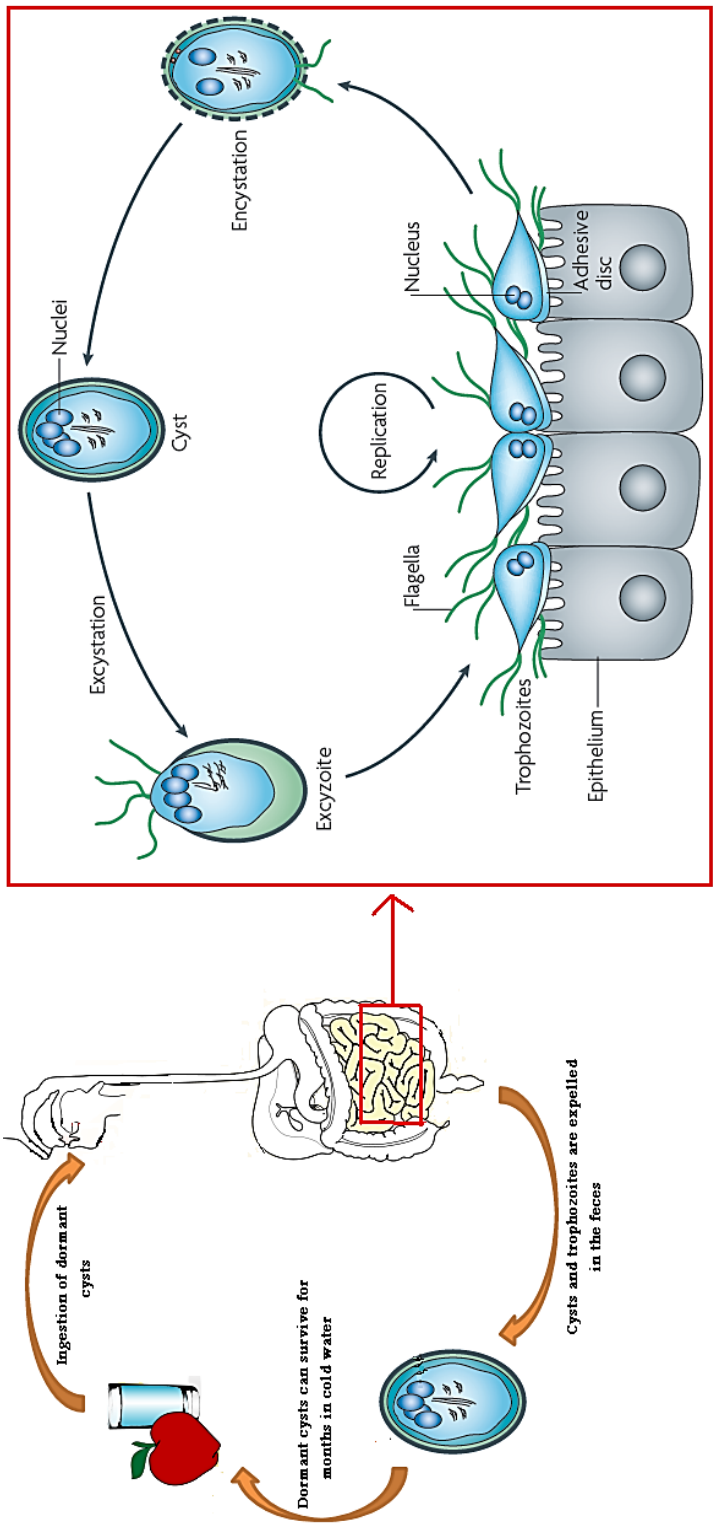
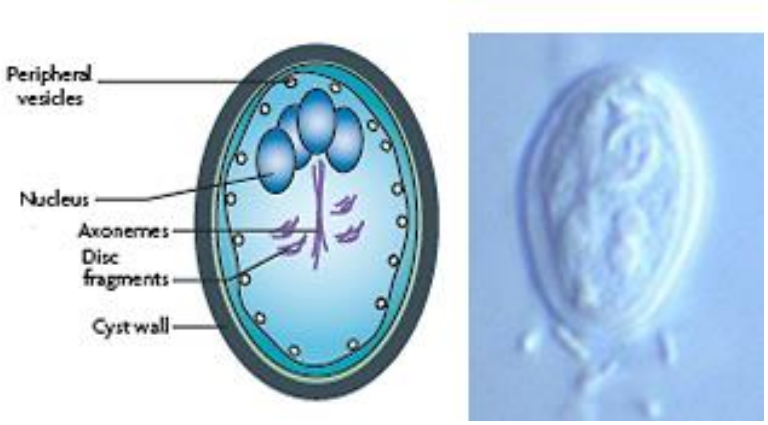


Fig 1. 7: The life cycle of *Giardia* (modified from Ankarklev et al, 2010, with permission and http://en.wikipedia.org/wiki/File:Giardia_life_cycle_en.svg)

A.



B.

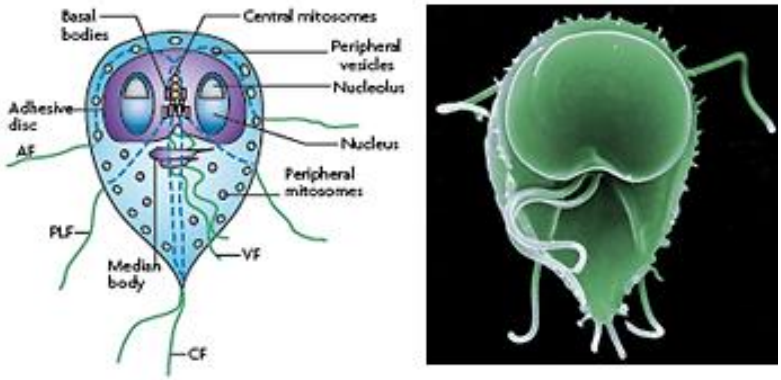


Fig 1. 8: *Giardia*'s different stages (modified from Ankarklev et al, 2010, with permission)**. A.** Thick and resistant cyst found in external environment like water, the cyst wall and inner layer consist of two membranes and protect the parasite. Cysts are oval-shaped and non-motile; they measure 8-12 µm long by 7-10 µm wide. A cyst has four tetraploid nuclei; its outer wall is 0.3-0.5 µm thick and is composed of a network of filaments ranging from 7 to 20 nm diameter; the main proteins composing this wall are N-acetylgalactosamine and three cyst wall proteins (CWP1, CWP2 and CWP3). The adhesive disk and flagella are disassembled and stored in the parasite.
Photo: **http://techalive.mtu.edu/meec/module02/images/giardia_cysts.jpg.**
B. Flagellated trophozoites are found in host small intestine, and measure 12-15 µm long by 5-9 µm wide. There are 8 flagella organised in four pairs: the anterior flagella (AF), ventral flagella (VF), posterior/lateral flagella (PLF) and caudal flagella (CF); dashed lines indicate internal structures. The median body is microtubular structure of unknown function; the adhesive disk is a large, rigid attachment structure composed of microtubules. There are several central and peripheral mitosomes in the trophozoite. Peripheral vesicles are lysosome-like vesicles that lie beneath the plasma membrane throughout the cell.
Photo: **http://hermes.mbl.edu/news/press_releases/2007/2007_09_27.html**, credit: Joel Mancuso, University of California, Berkeley, USA.

1.4.3/ Excystation

Once in the host digestive system, cysts release trophozoites in the upper small intestine; this mechanism is called excystation and is initiated by the acids present in the host stomach. (Alvarado and Wasserman, 2010, Bernal et al., 1998). Excystation occurs rapidly (between 10 and 15 min) and involves a reorganisation of the parasite's cytoskeleton enabling an excyzoite to emerge through the cyst wall (Ankarklev et al., 2010). Several studies suggested that signalling proteins such as protein kinase A (PKA), protein phosphatase 2A (PP2A) and calmodulin may have an important role during excystation (Alvarado and Wasserman, 2010, Lauwaet et al., 2007, Reiner et al., 2003, Abel et al., 2001, Bernal et al., 1998). It was suggested that PKA and PP2A may be key modulators during the first stage of acid induction; and calmodulin has been shown to be involved in the liberation of the excyzoite (Lauwaet et al., 2007, Abel et al., 2001, Bernal et al., 1998). A cysteine protease (CP2) was also shown to be involved in the excystation mechanism; CP2, released from lysosome-like peripheral vesicles, may degrade the cyst wall from the inside, facilitating the liberation of the excyzoite (Adam, 2001). Once liberated, excyzoites undergo cytokinesis twice, without intervening S phases to produce four trophozoites which are able to attach to the intestinal epithelia (Ankarklev et al., 2010, Bernander et al., 2001).

1.4.4/ Encystation

To permit the parasite to survive in hostile external environmental conditions and to guarantee the transmission of the infection to susceptible hosts, *Giardia* undergoes a mechanism, called encystation, to form a resistant cyst (Adam, 2001, Lujan et al., 1997). New formed cysts are released in the environment by evacuation in faeces. Encystation is an adaptive process to cope with the decrease of nutrients, cholesterol in particular, in the presence of high bile concentration in the lower small intestine (Lujan et al., 1996). Encystation is accomplished in three crucial steps: (i) the induction of encystation-specific gene expression, specifically those necessary for the synthesis and processing of cyst wall proteins (CWPs) (Lujan et al., 1998, Lujan et al., 1995b, Mowatt et al., 1995); (ii) the synthesis and

intracellular transport of CWPs through newly developed organelles such as the Golgi apparatus and encystation-specific secretory vesicles (ESVs) (Lujan et al., 1995a); and (iii) the assembly of the extracellular wall (Hehl et al., 2000).

During the early phase of encystation, CWPs are synthesized and transported to cell surface for assembly into the extracellular cyst wall (Gottig et al., 2006, Lujan and Touz, 2003, Lujan et al., 1997). This major protein transport is a regulated secretion process depending on intracellular signals, and consists in the export of at least three CWPs and glycans. CWPs are synthesised as luminal proteins without membrane anchors and are likely to be dependent on transmembrane cargo receptor or escorter proteins, at least for their export from the ER (Hehl and Marti, 2004). Before secretion, the cyst wall material is accumulated and concentrated in specific intermediate compartments: ESVs. A study suggested that the soluble CWPs were not exported from the ER by a default pathway but by a signal-dependent mechanism (Hehl et al., 2000). Another study confirmed that most regions of CWP3 were important for correct transport of the cyst wall component to ESVs and to the extracellular matrix while investigating targeting of a chimeric CWP3 reporter (Sun et al., 2003). Marti et al (2003) showed that CWP1 export signals were recognised by cargo receptors or escorter molecules which have yet to be determined. This study also showed that endogenous CWP1 or reporters containing CWP targeting signals were only able to exit the ER and accumulate in ESVs after trophozoites had received an encystation stimulus (Marti et al., 2003). This suggests that the transport of CWPs is a tightly regulated secretory process and not just an automatic consequence of increased synthesis of soluble cargo into the ER lumen. Later during encystation, sites for initiation of the cyst wall filament assembly appear on the trophozoite plasmalemma; followed by the assembly of the filamentous portion of the cyst wall (Adam, 2001, Reiner et al., 1989). Upon completion of encystation, motility is lost, and the outer portion becomes rounded and filamentous. Moreover, *Giardia* cysts are no longer attached to a surface. ESVs disappear and the internal cyst part included two trophozoites with four nuclei that have not yet completed cytokinesis (Adam, 2001).

It is crucial for *Giardia* survival that trophozoites encyst not only when faced with harsh environmental conditions, such as low cholesterol or alkaline pH and

high bile concentration, but also without the presence of any external inducer in order to guarantee its dissemination. A recent study showed that *Giardia* trophozoites can undergo encystation under both vegetative growth and encysting conditions (Sulemana et al., 2014). Indeed, after three to six hours in inducing conditions, encysting trophozoites continued to encyst regardless of whether the initial stimulus remained. This suggests that there is a point of commitment during encystation, after which encysting trophozoites cannot revert to growing or dividing trophozoites; therefore maximising *Giardia*'s chance of survival. However, the mechanisms regulating these decision-making processes have yet to be determined.

The full sequencing of *Giardia* genome has identified an extensive number of putative protein kinases, which suggests a complex requirement of signal transduction to coordinate cell proliferation but also the encystation-excystation processes (Morrison et al., 2007). Even though the proteins and pathways involved in signal transduction have only begun to be understood, few studies have described a possible mechanism involving several enzymes (Carranza and Lujan, 2010, Li et al., 2009, Touz et al., 2008, Bazan-Tejeda et al., 2007, Lauwaet et al., 2007, Kim et al., 2005):

- Extracellular regulated kinases 1 and 2 (ERK1 and ERK2) possess homology to the Mitogen-activated protein kinase (MAPK) family which members are involved in signalling in higher eukaryotes (Ellis et al., 2003). In *Giardia*, ERK1 and ERK2 were shown to be activated by phosphorylation, to translocate to the nuclei and to phosphorylate other proteins, including putative transcription factors (Ellis et al., 2003). In their study, Ellis et al demonstrated that these two enzymes localised to different structures; ERK1 to the median body, VD and basal bodies whereas ERK2 localised to the nuclei and caudal flagella. ERK2 was also shown to change its localisation pattern during encystation, being mostly cytoplasmic.
- PP2A is a highly conserved serine/threonine protein phosphatase which has been associated with encystation-excystation process in *Giardia* (Lauwaet et al., 2007). Lauwaet and collaborator have shown PP2A to localise to cytoskeleton structures and occasionally to cyst wall. Western

blot analyses showed that PP2A protein level in *Giardia* resembled the mRNA expression; and was up-regulated in cysts and also in the beginning of excystation (Carranza and Lujan, 2010, Lauwaet et al., 2007).

- Protein kinase B (PKB) is part of a serine/threonine kinases subgroup; *Giardia* PKB has a transmembrane domain in the cytoplasmic region and a potential nuclear localisation signal in the regulatory domain. PKB was shown to be constitutively expressed and up-regulated during encystation (Kim et al., 2005).
- Protein kinase C (PKC) is also a member of the serine/threonine kinases family. PKC plays key regulatory roles in signal transduction pathways controlling cellular responses including differentiation (Bazan-Tejeda et al., 2007). Several isoforms of PKC have been identified in *Giardia* which all show changes in expression during encystation. Bazan-Tejeda et al demonstrated that selective PKC inhibitors blocked the encystation in a dose-dependent manner.
- Arginine deiminase (ADI) is a cytoplasmic enzyme that belongs to the guanidine modifying super family; it catalyses the irreversible catabolism of arginine to citrulline in the arginine dehydrolase (ADH) pathway. ADI serves also as an important source of energy (Li et al., 2009). In 2008, Touz et al reported that ADI played multiple roles in signal transduction pathways leading to the control of encystation and antigenic variation. It was suggested that this was due to the potential of ADI to citrullinate arginines present in *Giardia* proteins (Touz et al., 2008). However, this additional function of *Giardia* ADI is still discussed since another study reported that recombinant dimeric enzymes were unable to perform this process (Li et al., 2009).
- Heat shock protein 90 (Hsp90) is a versatile chaperone protein involved in the modulation of diverse cellular processes such as cell cycle, signal transduction, differentiation. The majority of proteins chaperoned by Hsp90 have been shown to be transcription factors and protein kinases (Roy et al., 2012, Taipale et al., 2010). A recent study has shown Hsp90 implication in *Giardia* encystation by inhibiting *Giardia* Hsp90 in

trophozoites in pre-encystation stage and during encystation (Nageshan et al., 2014). This inhibition resulted in a robust induction of encystation during the pre-encystation stage; but not during encystation. This suggests that Hsp90 may be involved in the modulation of encystation in *Giardia*. However, the downstream events upon Hsp90 inhibition leading to induction of encystation remain unclear and have yet to be investigated.

These different studies have shown the implications of these enzymes in *Giardia* signal transduction pathway; however a link between these elements is still necessary and requires further investigation.

1.4.5/ *Giardia* secretory pathway

Two pathways have so far been characterised for the transport of proteins in *Giardia*: (i) the endomembrane protein transport system, (ii) lysosome-like peripheral vesicles, (Lujan and Touz, 2003, Adam, 2001).

1.4.5.1/ The endomembrane protein transport system

ER and Golgi complex are key components of the eukaryote endomembrane system which is involved in the protein folding and translocation. In higher eukaryotes, proteins destined for secretion possess a signal sequence directing them to the ER as they are translated in the ribosomes. The protein signal sequence binds to the signal recognition particle (SRP); this complex then binds to the SRP receptor (SR) present on the cytoplasmic side of the ER. The ER is composed of chaperonins such as Ig heavy chain-binding protein (BiP) and protein disulfide isomerases (PDIs).

Giardia ER was first observed after the cloning and characterization of SRα and the identification of an extensive membrane system labeled with anti-BiP antibody (Soltys et al., 1996). BiP functions as a chaperone protein aiding in the folding and transport of newly synthesized proteins (Craig et al., 1993). BiP chaperones were essentially found in reticular membranous net in vegetative and encysting trophozoites, but BiP proteins were also observed inside encystation-specific secretory vesicles (ESVs) during encystation; which was probably due to them still being associated with CWPs during transit of CWPs to the plasma

membrane (Lujan and Touz, 2003). PDIs were shown to be involved in the disulfide bond formation of newly synthetized secretory proteins; this event occurs in the ER lumen (Gilbert, 1997). Three PDIs have been identified in *Giardia*; two of them – PDI-1 and PDI-2 - were described as atypical due to the presence of KRKK ER-retention signal on their short cytoplasmic tails (Knodler et al., 1999). PDI-3, on the contrary, was shown to possess a C-terminal KQRL ER-retention signal. All three PDIs localise in the ER in vegetative trophozoites; however, PDIs were also observed within ESVs and on the surface of encysting parasites (Reiner et al., 2001). BiP and PDI-2 expression was shown to increase during trophozoite differentiation into cyst, which suggests that the rapid synthesis of cyst wall proteins during encystation requires more ER chaperones (Lujan et al., 1997). The presence of ER chaperones outside the ER during encystation also suggests that, due to the presence of an elementary endomembranous system in *Giardia*, the proper quality control for ER-protein retention may either be absent, fail, or not be sufficient to deal with an increase in expression of cyst wall components (Lujan and Touz, 2003, Knodler et al., 1999).

In eukaryotes, Golgi complex serves as a major sorting point in the secretory pathways. It selectively targets proteins and lipids to different organelles. The Golgi apparatus consists of a series of flattened cisternal membranes forming a stack with a defined *cis* to *trans* orientation (Marti and Hehl, 2003, Lujan and Touz, 2003). Cargo modifications during transport through this cisternae stack include: (i) processing N- and O-linked oligosaccharides; (ii) addition of mannose-6-phosphate residues to lysosomal proteins; (iii) sulfation; (iv) phosphorylation; and (vi) acylation (Marti and Hehl, 2003). At the most *trans* side of the Golgi apparatus, the *trans*-Golgi network (TGN), proteins, glycans and lipids are packed into different classes of vesicles bound for the plasma membrane or other organelles in the cell (Marti and Hehl, 2003). The Golgi apparatus has a unique architecture which is remarkably conserved throughout eukaryotic evolution (Mellman and Simons, 1992). However, the typical eukaryotic Golgi complex appears to be missing in vegetative *Giardia* trophozoites. It has been suggested that *Giardia* trophozoites may possess one or more organelles where typical Golgi functions take place, even though they do not have a Golgi-like appearance (Lujan and Touz, 2003). Interestingly, the rise of an

elaborate Golgi complex, consisting in 3 – 20 parallel cisternae and dramatic changes in the nuclear envelope, including the appearance of numerous nuclear pores, were observed close to the nuclear membrane only in encysting trophozoites via the use of electron microscopic techniques (Lujan and Touz, 2003). Whether this perinuclear region contains Golgi markers has yet to be determined. Interestingly, a set of membrane-bound structures other than peripheral vesicles were identified and localised near the ER and/or nuclear envelope compartment; but their implication as part of a putative trophozoite secretory system remains to be elucidated (Benchimol, 2002). All this suggests that *Giardia* trophozoite nuclear envelope may play an important role in protein sorting and transport. The sequences of two *Giardia* syntaxins were reported (Dacks and Doolittle, 2002). Syntaxins are associated with transport through the Golgi apparatus; they belong to the SNAREs family. SNAREs are soluble NSF (N-ethylmaleimide-sensitive factor) attachment proteins receptors (SNAP receptors). They play crucial roles in vesicular transport and membrane fusion in higher eukaryotes (Parlati et al., 2002). The identification of Golgi-related syntaxins in *Giardia* trophozoites provides further evidence for the presence of a cryptic Golgi complex in the parasite.

The formation of transport vesicles and their delivery to the target membranes were shown to be controlled by small GTPases of the ADP-ribosylation factor (ARF) and Rab families respectively (Pfeffer, 2001, Sogaard et al., 1994). ARFs are required for maintaining intracellular transport and the organelle structure integrity. In *Giardia*, an ARF homologue was identified and immunolocalised to a region surrounding both nuclei, in an area described as likely to be the rough ER (RER) (Lujan and Touz, 2003). A *Giardia*-Rab 1 GTPase was also described and shown to localise to the ER and PVs (Langford et al., 2002). Rab 1 was also associated with ESVs during encystation. Rab proteins were described as organelle-specific GTPases that facilitate the delivery of transport vesicles by regulating the protein-protein interaction (Parlati et al., 2002, Pfeffer, 2001). The investigation of the transport and protein-protein interaction of Golgi resident markers like-Rab 1 may help localising and identifying the minimal Golgi-like structure still missing in *Giardia* trophozoite.

1.4.5.2/ Protein transport to the lysosome-like peripheral vesicles

Most eukaryotes possess a system of endosomes and lysosomes that degrade and recycle endogenous proteins or proteins acquired by endocytosis and phagocytosis from the extracellular space. Early endosomes internalise endocytosed proteins to allow for their return to the cell membrane or transport to late endosomes. Once in the late endosomes, proteins are transported to and degraded by the lysosomes (Scott et al., 2014). Both endosomes and lysosomes are acidic with a pH < of 6 and of 5 respectively. *Giardia* lacks compartments resembling early/late endosomes and lysosomes. Instead, it possesses peripheral vacuoles (PVs) that are located underneath the plasma membrane and seem to function as endosomes and lysosomes (Lanfredi-Rangel et al., 1998). The lysosomal targeting pathway has been well studied and defined in yeast and mammalian cells. In this pathway, soluble hydrolases were shown to be sorted away from the secretory pathway by reaching the lysosomes through a receptor-mediated process, via oligosaccharide signals attached to the proteins in the Golgi apparatus (Lujan and Touz, 2003). In *Giardia* trophozoites, soluble enzymes such as acid phosphatase and cathepsin B were shown to localise to the PVs by their immune-localisation (Lanfredi-Rangel et al., 1998, Ward et al., 1997). Supposing that *Giardia* possesses a mechanism similar to the one described above, soluble proteins would need to interact with a receptor to be transported to the PVs; so far, no receptor were described to be involved in protein delivery to PVs. Moreover, protein glycosylation was shown to be controversial in *Giardia* due to the lack of evidence for either N- or O-glycosylation in any trophozoite protein (Lujan et al., 1995a).

Transmembrane proteins are sorted from the Golgi to endosome/lysosomes via motifs present in their AA sequence. Tyrosine-based signals have been shown to be the motifs usually present in the cytoplasmic tail of integral proteins involved in lysosomal protein transport (Le Borgne and Hoflack, 1998). Touz et al. showed that tyrosine-based motifs were playing an important role in *Giardia* lysosomal trafficking. They also found an encystation-specific cysteine protease (ESCP) in the lysosome-like PVs (Touz et al., 2003, Touz et al., 2002). ESCP was described as a

transmembrane protein carrying an YRPI motif within its cytoplasmic tail. This motif was demonstrated to be crucial for ESCP delivery as deletion and point mutations of the motif missorted ESCP to the plasma membrane (Touz et al., 2003). All these findings indicate that *Giardia* possesses an endosomal/lysosomal system concentrated in one single organelle, the PV. This suggests that the subdivision of this organelle into endosome/lysosome specific compartments occurred later than the development of a basic machinery for lysosomal-protein delivery. Additionally, a study showed that in certain periods of *Giardia* life cycle, PVs could function as secretory organelles (Slavin et al., 2002).

1.5/ Host-parasite interactions

The human intestine acts as an anaerobic "bioreactor" programmed with a colossal population of bacteria. This population of bacteria, up to 1 x 10^{14}, far surpasses all the other microbial communities associated with the body's surfaces in size (Backhed et al., 2005, Savage, 1977). All three domains of life - Archaea, Eukarya and Bacteria- are found in the gastrointestinal tract and formed a microbiota approximately 10 times greater than the total number of human somatic and germ cells in a body (Backhed et al., 2005).

1.5.1/ The human intestinal epithelium

The intestinal epithelium is a critical interface between the environment and the organism. It is composed of the mucous coat, the closely packed microvilli and the glycocalix which prevent direct contact of most macromolecules on the apical side; and the tight junctions which restrict paracellular transport to small molecules (Fig 1.9) (Snoeck et al., 2005).

The mucosal surface of the intestine forms a major mechanical barrier that separates the host's internal milieu from the external environment. However, it is not a passive barrier. Indeed, the tight junctions are dynamic structures regulated by events taking place in the intestine; and the enterocytes have a function in the presentation of antigens to T cells and then, in their digestion (Cobrin and Abreu, 2005).

1.5.1.1/ The intestinal epithelial cells: Enterocytes

Enterocytes are the intestinal absorptive cells found in the small intestine and colon. These cells are hyperpolarized, joined together by tight junctions (TJs), and efficient protein sorting and addressing are essential to the establishment and maintenance of cell polarity, on which the integrity of the epithelial barrier depends (Snoeck et al., 2005, Jung et al., 1995) (Fig 1.9).

Enterocytes have a distinctive structural asymmetry. Indeed, sorting compartments of the secretory apparatus, subsets of membrane and cytoskeletal proteins and microtubules are asymmetrically distributed throughout the cytoplasm. These subsets of proteins localise to functionally and structurally distinct membrane domains, termed apical and basolateral (Drubin and Nelson, 1996). They are composed of a TGN, nucleus, ER, TJs which join cells together; and a lot of vesicles: apical early endosome, apical recycling endosome, late endosome, common lysosome, basolateral early endosome, lysosomes and mitochondria. These vesicles are involved in the transport of macromolecules and nutrients from the apical surface to the basal-lateral surface and inversely (Fig 1.9) (Snoeck et al., 2005).

Since the intestinal cells are hyperpolarized, a short-circuit current (Isc) allows a current to travel along a path where essentially no (or a very low) electrical impedance is encountered; and reflects net ions transport in the form of cation absorption or anion secretion on the apical side (Fig 1.10). In epithelial cell, the Isc represents the highest electric current between the apical (negatively charged) and the basolateral (positively charged) surfaces. The electric resistance of this Isc is called *trans*-epithelial electrical resistance (TEER) and seems to be at the origin of the cell permeability to molecules (Narai et al., 1997).

Some of the major functions of enterocytes are (i) ion uptake, including sodium, calcium for example; this uptake typically occurs through active transport with hydrolysis of one molecule of ATP; (ii) Water uptake which follows the osmotic gradient established by the Na^+-K^+-ATPase channel on the basolateral surface; it can be a transcellular or a paracellular transport; (iii) Sugar uptake, sugar molecules are absorbed and cross the apical membrane of the enterocytes using the Na^+-

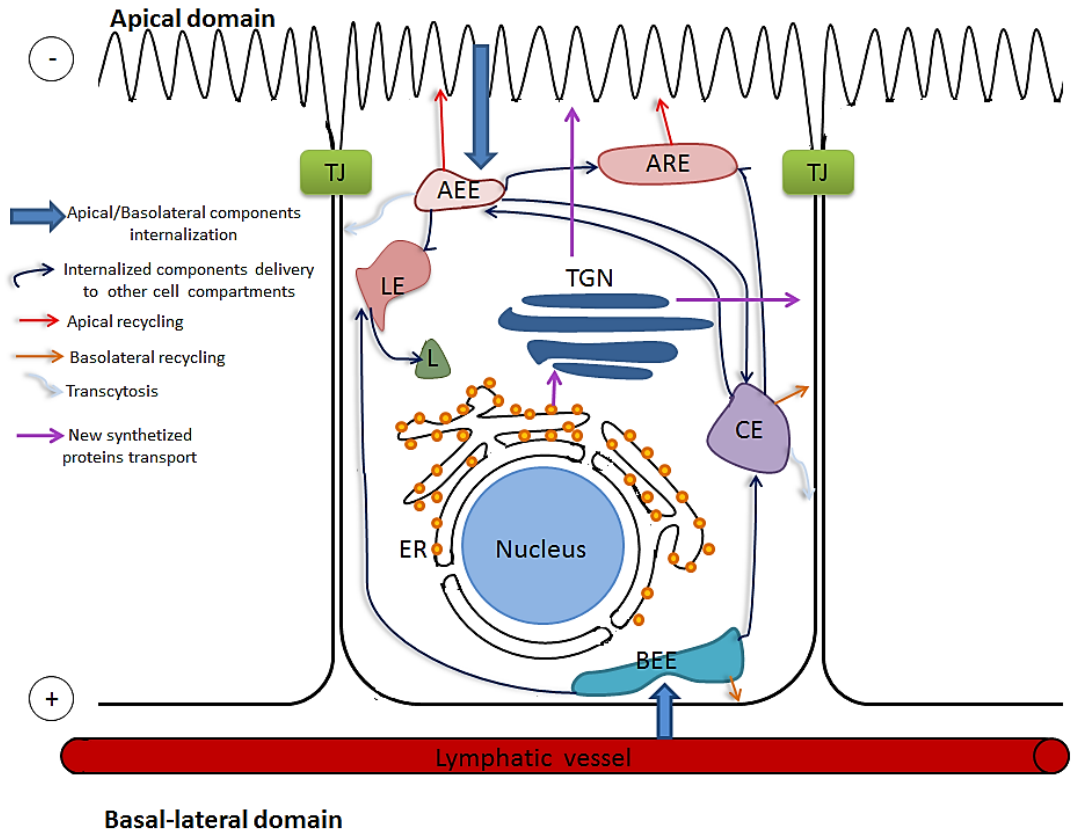


Fig 1. 9: Model for endocytic traffic in polarized enterocytes (Based on Snoeck et al, 2005, with permission). The apical and basolateral surfaces are separated by tight junctions (TJs). Upon internalization (➡), fluid and membrane components are delivered (↗) to distinct early endosomes, the apical (AEE) or the basal (BEE). Then, from the AEE, internalized components can be recycled (→) or transcytosed (↘), be delivered to the late endosome (LE) and ultimately to lysosomes; or they can be delivered to the apical recycling endosome (ARE) or to the common endosome (CE) for recycling through the ARE or to transcytose. From the BEE, internalized components can recycle, be delivered to the CE for recycling to the basolateral membrane (→) or transcytose through the AEE or the ARE, or they can be delivered to the LE and lysosomes. Membrane proteins newly synthesized (●) can be directly delivered to the apical or basolateral surface after sorting in the TGN (*trans*-Golgi Network).

dependent glucose transporter (iv) peptides and amino acid uptake, proteins are cleaved to amino acids or small peptides and are then absorbed on the apical membrane.

1.5.1.2/ The intestinal molecular pathways

There are three way possible for the transport of molecules throug enterocytes.

- **Transcellular transport**

Because of the enterocyte physical structure, it is very unlikely that macromolecules can cross the lipid bilayer into the cytosol. Indeed, its plasma membrane being composed of a lipid bilayer in which membrane-bound proteins and glycoproteins are situated, the molecules can only be transported by receptors present at the enterocyte membrane surface. The major pathway for molecules is a transcellular transport called endocytosis which is a receptor-mediated mechanism (Fig 1.10). Besides transport of molecules, endocytosis is also involved in regulation of cell-surface receptor expression, maintenance of cell polarity, and antigen presentation (Conner and Schmid, 2003, Drubin and Nelson, 1996). In enterocytes, three other pinocytosis mechanisms have been proposed: (i) clathrin-mediated endocytosis, which is the major pinocytic pathway; (ii) caveolin-mediated endocytosis, which is a slow process consequently unlikely contributing significantly to bulk fluid-phase uptake; (iii) and clathrin-caveolin-independent endocytosis, which remains poorly understood (Snoeck et al., 2005, Conner and Schmid, 2003) (Fig 1.11).

Another way to absorb nutrients, ions or water is via epithelial transporters which are integral membrane protein pumps or channels (Snoeck et al., 2005). Nutrients, such as sugars, ions, and amino acids, enter the enterocyte cytoplasm at the apical membrane by these transporters and exit the basolateral membrane. Some are adenosine triphosphate (ATP)-dependent and others rely on the osmotic gradient between the lumen of the intestine and the mucosa (Fig 1.10).

❖ **Paracellular transport**

For some molecules, another way to go through the intestinal epithelial barrier is paracellular transport. Instead of crossing through enterocytes, molecules go through the tights junctions, which are dynamic structures, between adjacent epithelial cells (Fig 1.10). However, the passage of large molecules through tight junctions is not possible; they have to be cleaved into peptides or amino acids for proteins, or saccharose, glucose, for example, for polysaccharides. Chantret et al compared 20 different human colonic carcinoma cell lines in order to determine which line spontaneously developed normal human intestinal cell characteristics, such: (i) the organization of the cells into a polarized monolayer; (ii) the presence of an apical brush border; and (iii) the presence of brush border-associated hydrolases such as sucrase-isomaltase, lactase, or alkaline phosphatase (Chantret et al., 1988). The only human colonic carcinoma cell line able to do all three was the human colonic adenocarcinoma derived epithelial cell line-2 (CaCo-2) concluding that only the CaCo-2 cell line showed the differentiation characteristics of mature enterocytes. Moreover, in another study, Grasset et al showed that the CaCo-2 cells exhibited a similar electrical properties as enterocytes, by displaying a measurable TEER and a positive Isc reflecting ion and molecules transport through cells similarly to human enterocytes (Grasset et al., 1984).

❖ **Water uptake in the intestine**

Water transport across the cell membranes takes place via several routes: (i) by osmosis, via the lipid bilayer and the water channels called aquaporins (AQP) (Fig 1.12 A.); (ii) by co-transport, via co-transporters or uniporters such as the Potassium-Chloride co-transporter (KCC) (Fig 1.12 B.); (iii) or by co-transport and osmosis at the same time, via specific co-transporter such as the Sodium-glucose co-transporter 1 (SGLT-1) (Fig 1.12 C.) (Zeuthen, 2010)

AQP are integral membrane proteins expressed in various organs and tissues. They serve as channels in the transfer of water and, sometimes, small molecules across the membrane (Takata et al., 2004, Agre et al., 2002) (Fig 1.13). The water is transferred through the cell membranes via the osmotic pressure

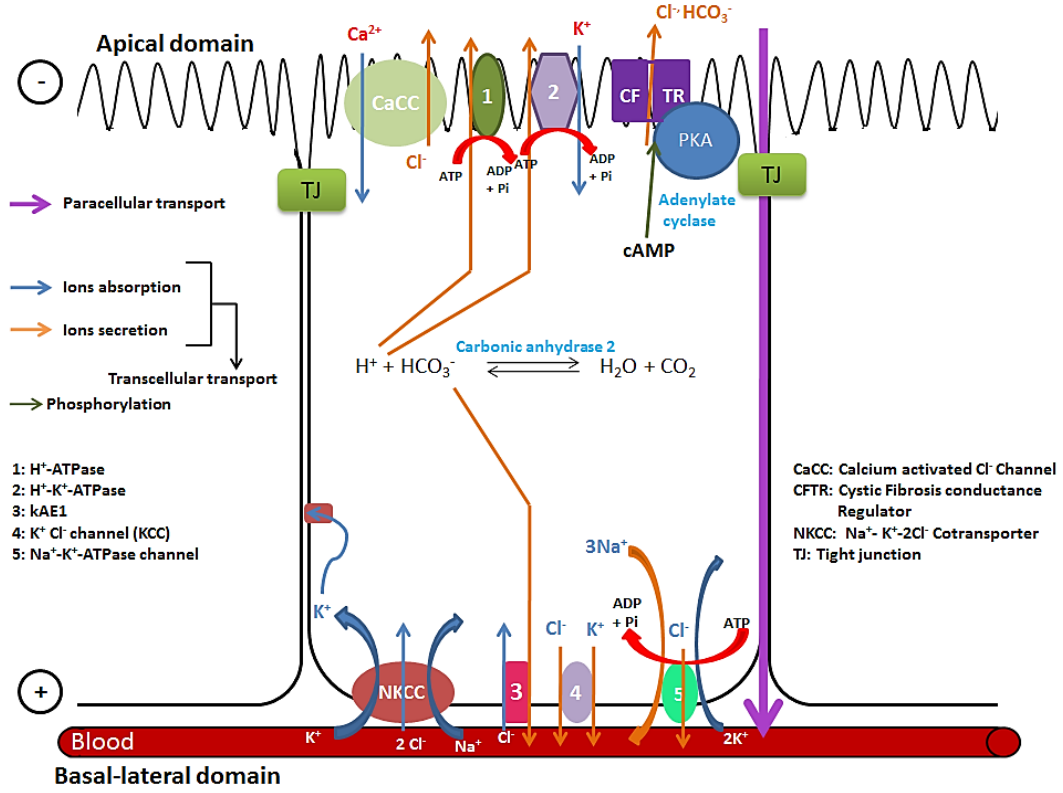


Fig 1. 10: Model of ions transport in polarized enterocytes. On the apical surface, there are four major ions transporters. The H^+-ATPase (1) and H^+-K^+-ATPase (2) are ATP-dependent, they need energy for the ions absorption and/or secretion. Via the hydrolysis of one molecule of ATP into ADP and inorganic phosphate (Pi), ions are absorbed or secreted or both. Currently, two of these channels have been identified through which Cl^- can be secreted into the intestinal lumen, thus creating the driving osmotic gradient for fluid secretion, namely: (i) CFRT which is embedded as dimers into the membrane, it is regulated by the phosphorylation of protein kinase A (PKA) and it is a cyclic adenosine monophosphate (cAMP)- dependent mechanism; (ii) CaCC which is an Ca^{2+}-dependent pathway for Cl^- secretion, when the cytosolic concentration of Ca^{2+} increases, the secretion of ions Cl^- into the lumen is increasing too.

On the basolateral surface, there are four major transporters too. Three of them have a role in the chloride secretion: (i) the Na^+-K^+-ATPase channel (5) is ATP-dependent and provides the energetic requirements for active Cl^- secretion by transporting three Na^+ ions out of the cell and two K^+ ions into the cell; (ii) the Na^+-K^+-2 Cl^- co-transporter (NKCC) which mediates Cl^- uptake at the basolateral pole and thereby provides the substrate for the apical Cl^- secretion; (iii) potassium channels (4) maintain cellular electroneutrality by compensating for Cl^- efflux and keeping the cell in a state of hyperpolarization. The last main basolateral receptor is kidney anion exchanger 1 (kAE1) (3) which has a role in the absorption of Cl^- ions from the basolateral side; it exchanges HCO_3^- anion from the cytosol with Cl^- extracellular.

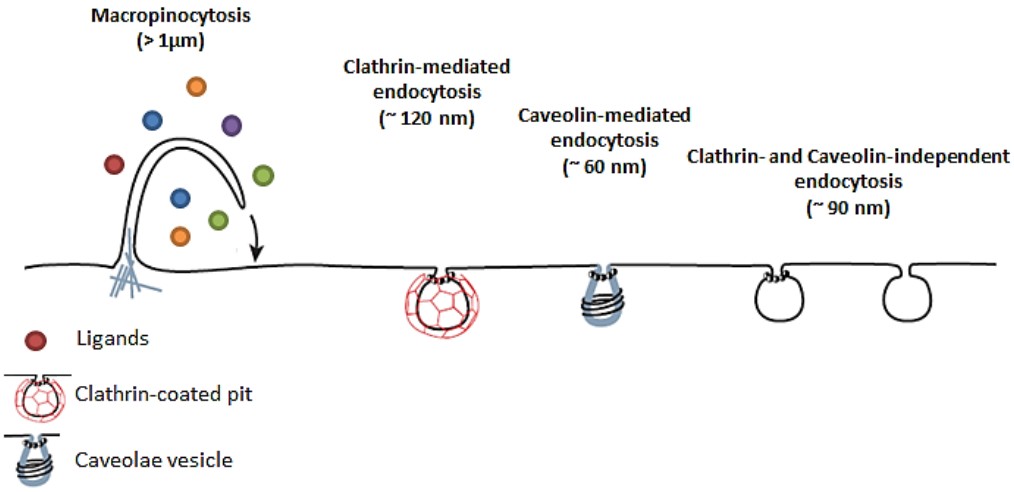


Fig 1. 11: The different types of pinocytosis in mammalian cells (reproduced and modified from Conner et al, 2003, with permission). These pathways differ with regard to the size of the endocytic vesicle, the nature of endocytosed molecules, receptors and lipids, and the mechanism of vesicle formation

The macropynocitosis is unlikely to occur by enterocytes since this mechanism involves extensive membrane ruffling and sampling of large volumes of the extracellular milieu. The clathrin-mediated endocytosis involves the concentration of high-affinity transmembrane receptors and their ligands into "clathrin-coated pits" for the formation of a coated veside bringing the molecules into the cytosol to be recycle by endosomes. Caveolae, which are flask-shaped invaginations of the plasma membrane, are involved in the caveolin-mediated endocytosis. These invaginations demarcate cholesterol and sphingolipid-rich microdomains in which many different signalling molecules and membrane transporters are concentrated.

regulation and driven by a hydrostatic gradient within the hemipore. 10 AQP channels have been identified so far, AQP0 to AQP10 (Agre et al., 2002). Some are aquaglyceroporins; they are water and glycerol permeable. AQP1 is the main AQP channel, it is found in any type of organs or tissues. In the intestine, AQP5 is mainly found in the small intestine; and AQP3, which is an aquaglyceroporin, is mainly found in the colon (Matsuzaki et al., 2004, Takata et al., 2004, Agre et al., 2002).

Co-transporters and uniporters are membrane transport proteins. They can transport one or several molecules through the cell plasma membrane. The tight coupling between water and molecules in co-transport allows co-transporters to function as molecular water pumps (Fig 1.12 B.). The free energy contained in the transported molecules gradient can be transferred to the water flux. In some co-transporters, like the sodium-glucose co-transporter 1 (SGLT-1), both modes of transport can occur (Fig 1.12 C.). They co-transport water through the plasma membrane but also possess water channel properties (osmosis). The transport of water via these transporters is bimodal; osmosis and co-transport can be done in parallel (Zeuthen, 2010).

1.5.2/ Pathophysiology

Giardia trophozoites adhere to the epithelial surface of the small intestine via the ventral disc (VD). This attachment triggers a series of events: (i) the increase of intestinal cell apoptosis; (ii): the disruption of the intestinal barrier function; (iii) a diffuse shortening of the brush border microvilli and malabsorption; (iv) hypersecretion of anion (such as chloride) and increase of the intestinal transit (diarrhoea) (Cotton et al., 2011).

1.5.2.1/ Increase of the intestinal cell apoptosis

Programmed cell death (PCD) helps to maintain homeostasis of intestinal cells either via senescence in which case cells stop dividing or via apoptosis in which cellular mechanism trigger the cell's own death. Apoptosis assists in preserving the homeostatic turnover of gastrointestinal epithelial cells; it is also involved in maintaining a correct cell number in organisms by balancing cell growth and death. The main regulatory component of apoptotic death is the

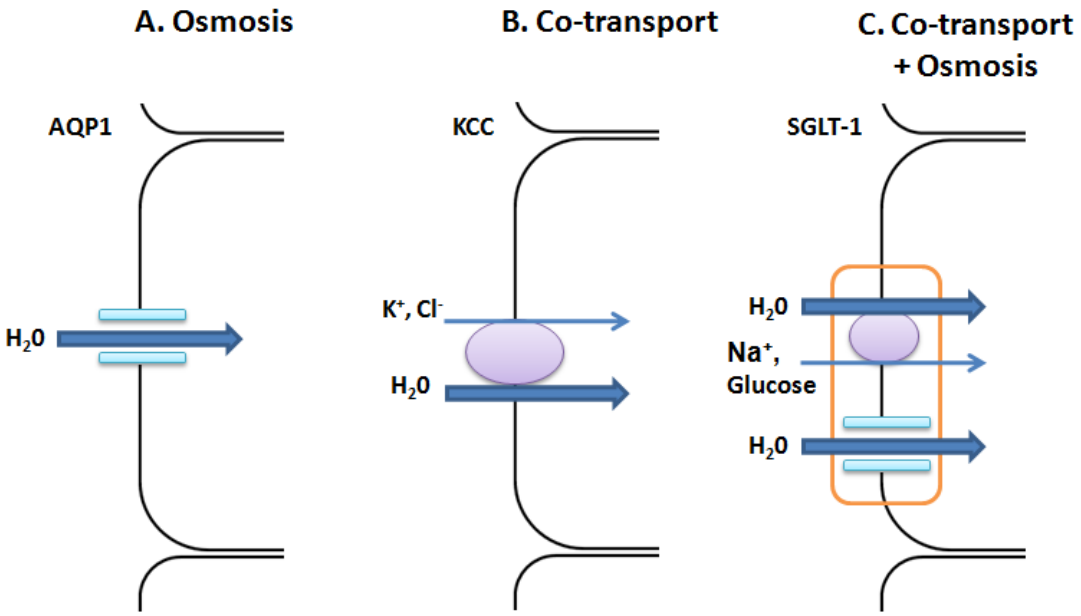


Fig 1. 12: The different water transport routes (Based on Zeuthen, 2010, with permission). **A.** Via osmosis which is driven by the water chemical potential difference. **B.** Via co-transport of substrates, such as potassium-chloride co-transporter (KCC), co-transporters function as molecular water pumps; the free energy contained in the substrate gradient is transferred to the transport of water. **C.** Via co-transport and osmosis, some co-transporters can employ both co-transport and osmosis to transport water and substrates through cell membranes.

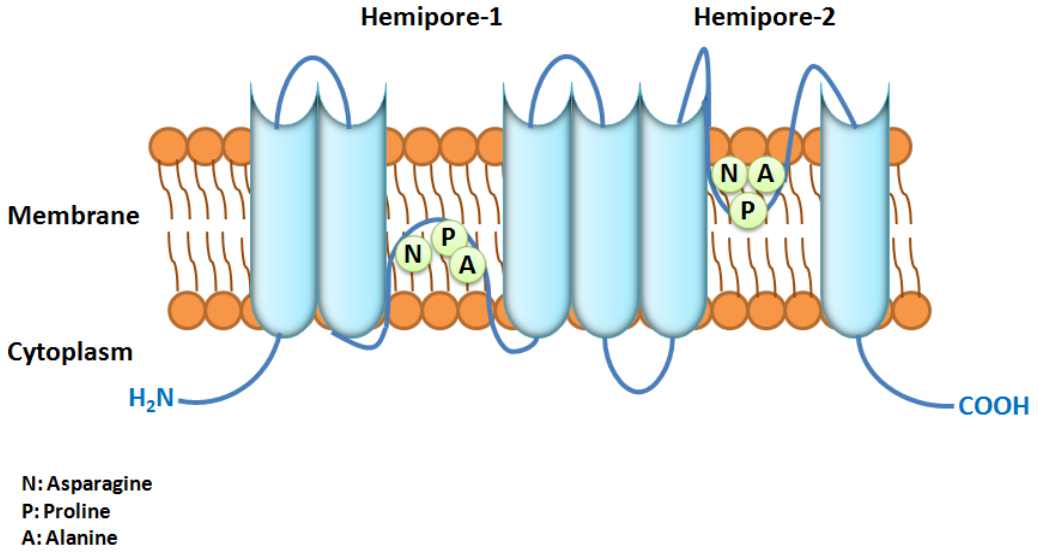


Fig 1. 13: Schematic representation of the AQP3 structure. AQP3 is mainly expressed in the epithelia of the upper digestive tract and the lower digestive tract (from the colon to the anus). Each AQP isoforms is composed of a single polypeptide chain of approximately 270 AAs and spans the membrane 6 times. Both the amino- and carboxyl-terminal ends are in the cytoplasm; and highly conserved regions (hemipores) contain the consensus motif: asparagine-proline-alanine (NPA).

sequential activation of proteolytic cysteinyl aspirate-specific proteases called caspases (Hall et al., 1994). The induction of apoptosis by *Giardia* in enterocytes was established by Roxstrom-Lindquist et al (2005). In this study, PCD-associated genes such as CC and CXC chemokine ligand and other chemo-attractant receptor genes, transcription regulators genes, and cell cycle regulator genes were up-regulated in cells exposed to *Giardia* components (Roxstrom-Lindquist et al., 2005). Furthermore, it has been established that high rates of epithelial apoptosis occur shortly after exposure to *Giardia*, both in patients with chronic giardiasis and *in vitro* (Troeger et al., 2007, Roxstrom-Lindquist et al., 2005, Chin et al., 2002). Some studies suggested that *Giardia* infections initiate apoptosis via the activation of pro-apoptotic caspase-8, known to be triggered following the activation of caspase-6, a caspase-3 downstream target (Cotton et al., 2011). Caspases are a family of proteases that cleave their substrates after aspartic acid residues (Slee et al., 1999). They are present in cytosol as inactive zymogens, called pro-caspases, which require proteolytic processing to achieve their active configuration. Several distinct routes lead to caspase activation depending upon the stimulus that initiates the death program. Their activation was shown to be stepwise with a downstream effect as caspases are able to process other caspases (Slee et al., 1999). The main extrinsic pathway was shown to be initiated at the cell membrane by ligands of tumour necrosis factor (TNF) receptors; this ligand-binding leads to assembly of the death-inducing signalling complex which recruits and activates capsase-8 and -10 (Connolly et al., 2014). The intrinsic pathway is initiated by mitochondria whose outer membrane becomes permeable to cytochrome *c* upon certain cellular stresses such as DNA damage or ER stress (Connolly et al., 2014, Tait and Green, 2010). Both intrinsic and extrinsic pathways converge on activating the executioner caspase-3 and -7 throughout the downstream induction of other caspases (Tait and Green, 2010, Slee et al., 1999). Whether or not the caspase activation by *Giardia* is direct or occurs due to the activation of caspases-6 and -3 has to be determined (Fig 1.14). The possible roles of apoptosis inducing factors (AIFs) in *Giardia*-induced enterocyte apoptosis are unknown; AIFs may activate apoptotic DNA breakdown in a caspases-independent design when released after a loss of mitochondrial integrity (Cotton et al., 2011). Although both parasite and host factors can induce intestinal

epithelial cell apoptosis, the components responsible for its activation during giardiasis have yet to be identified.

1.5.2.2/ Disruption of the intestinal barrier function

In the intestine, adjacent enterocytes are connected by TJs. These TJs establish a selective barrier that separates the external environment from underlying host tissues; they also are the first line of defence against the influx of harmful luminal contents, such as pathogens and cytotoxic products (Kotler et al., 2013, Anderson and Van Itallie, 1995). They are composed of a branching network of sealing strands. Each strand acts independently from the others and are formed from various transmembrane proteins (claudin proteins embedded in both plasma membranes), cytosolic plaque proteins (Zonula occudens, ZO-1) and cytosolic regulatory proteins (myosin light chain kinase, MLCK; and Rho kinase, ROCK) (Anderson and Van Itallie, 1995). *Giardia*-mediated increases of the intestinal permeability result from modifications of the apical TJs, including disruption of ZO-1, cytoskeletal filamentous (F)-actin, claudin-1, and α-actinin which is a component of the actomyosin ring that regulates paracellular flow (Cotton et al., 2011, Troeger et al., 2007, Scott et al., 2002). *In vivo* studies of *Giardia* show that high intestinal permeability and macromolecular uptake occur contemporaneously with the peak of trophozoite colonisation. A return to normal baseline is observed following host parasite clearance (Cotton et al., 2011, Scott et al., 2002). In 2002, Chin et al demonstrated that inhibition of caspases-3 prevents relocalisation of F-actin and ZO-1, suggesting a direct cause-effect relationship between *Giardia*-induced apoptosis and small intestinal barrier function. However, which of the tight junction regulatory protein (MLCK and/or ROCK) is activated by the caspase-3 mediated F-actin and ZO-1 disruption has to be determined (Fig 1.14). Furthermore, by which mechanisms *Giardia* is disrupting the intestinal epithelial barrier function, during giardiasis, is still unknown and needs further investigations.

1.5.2.3/ Diffuse shortening of brush border microvilli, malabsorption and anion hypersecretion

Giardiasis can also reduce the whole intestinal absorptive surface area, either in the presence or absence of villous atrophy. However, some studies have clearly demonstrated that the total duodenal surface area is decreased in patients with chronic giardiasis; this is supported by previous studies showing that *Giardia* infections induce the diffuse shortening of epithelial brush border (Troeger et al., 2007, Scott et al., 2004, Scott et al., 2000). This process is triggered by activated $CD8^+$ lymphocytes; consistent with the hypothesis that *Giardia*-mediated immunopathophysiology occurs secondary to the disruption of the intestinal barrier (Scott et al., 2004). However, $CD8^+$ lymphocytes are known to contribute to the parasite clearance as well as to contribute to the pathophysiology. Some studies have shown that maldigestion happens during giardiasis due to a deficiency in numerous brush border enzymes; they have also demonstrated that diffuse shortening of brush border microvilli causes small intestinal malabsorption due to reduced absorption of water, glucose and electrolytes (Buret, 2007, Scott et al., 2004, Scott et al., 2000, Faubert, 2000, Cevallos et al., 1995). Malabsorption of nutrients and electrolytes creates an osmotic gradient that draws water into small intestinal lumen causing a small intestine distension and rapid peristalsis (Fig 1.14). It has been reported that *Giardia*-infected individuals can also have elevated levels of faecal fatty acids (Morken et al., 2009). It is possible that high levels of undigested carbohydrates contribute to diarrhoea, in *Giardia* infections, following their transformation into short chain fatty acids by colonic microbiota (Cotton et al., 2011). An increased chloride secretion has also been shown to contribute further to diarrhoea in giardiasis (Baldi et al., 2009, Troeger et al., 2007, Cevallos et al., 1995). In summary, diarrhoea in *Giardia*-infected individuals is induced by several mechanisms including diffuse shortening of brush border microvilli, malabsorption, and anion hypersecretion. These phenomenons appear to be initiated via caspase-dependent apoptosis and tight junctional disruptions (Fig 1.14). The parasitic and/or host factors involved in these events have still to be identified.

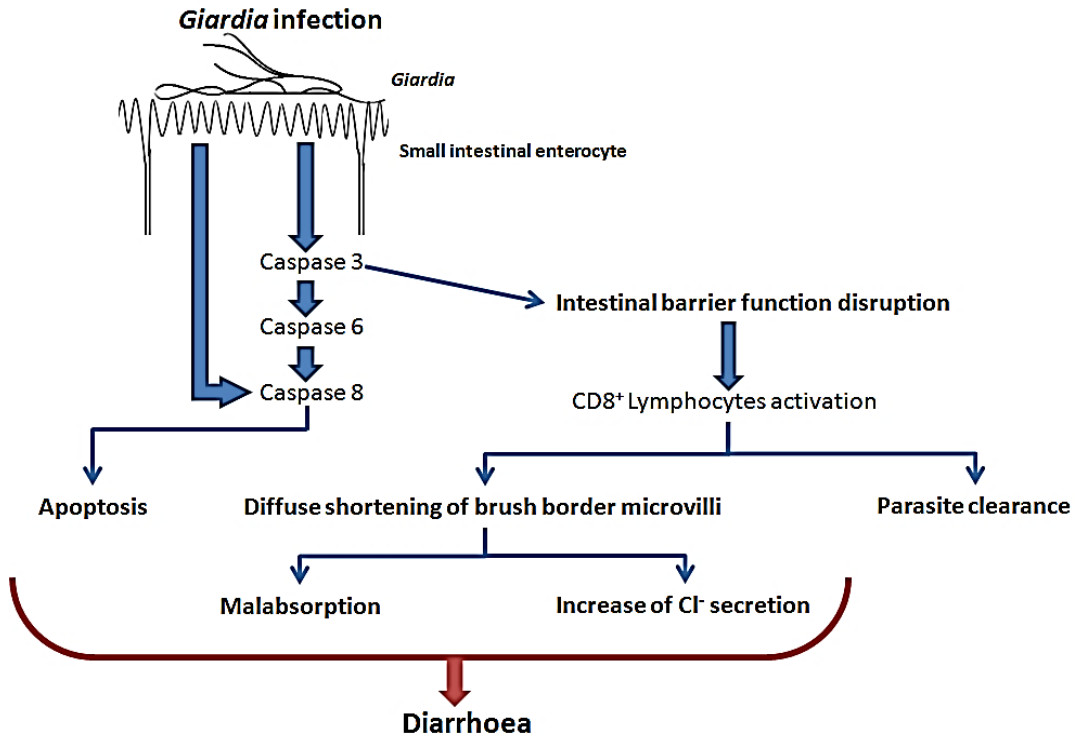


Fig 1. 14: *Giardia* infection Pathophysiology. *Giardia* infection triggers a cascade of physiological reactions within the intestinal epithelial cells: First, caspase 8 protein is activated either directly or via other caspases (caspase 3 and 6); and induces apoptosis. The intestinal barrier function is disrupted via the inhibition of caspase 3 which modifies the tight junctions' permeability; this leads to the activation of T cells inducing a parasites clearance but also a diffuse shortening of brush border microvilli. The shortening of microvilli causes malabsorption and a chloride (Cl^-) secretion increase. *Giardia*-induced diarrhoea may result from the combination of all these intestinal enterocyte physiological reactions.

1.5.3/ *Giardia* pathogenic factors

The intestinal epithelium is continuously exposed to a multitude of food and other antigens that may either be beneficial or harmful for the host. Consequently the intestinal immune system has to balance between either tolerance against commensal bacteria and food antigens or protective immune responses induced by intestinal pathogens and toxins (Cobrin and Abreu, 2005, Medzhitov and Janeway, 2000).

To survive and colonise such a hostile environment, *Giardia* has to adapt constantly to the host environment changes via modification in proteins expressed by the trophozoites. *Giardia* genome is distributed between five chromosomes, its size slightly varies between assemblage A and B (Table 1.5) (Adam et al., 2013, Morrison et al., 2007). Assemblage A (WB isolate) is approximately 11.7 mega base pair (Mb) in size, divided among 306 contigs and with a very low level of heterozygosity (< 0.01 %) (Perry et al., 2011, Morrison et al., 2007). Assemblage B genome is 12.01 Mb in size, divided among 544 contigs and with a level of heterozygosity higher than assemblage A (0.425 %) (Adam et al., 2013, Franzen et al., 2009). Sub-assemblage A2 (DH isolate) genome is 10.7 Mb in size which makes it the smallest assembled *Giardia* genome to date; its genome is divided among 239 contigs and its level of heterozygosity is also very low (0.037 %) (Adam et al., 2013). Whole genomes were compared showing that there is a substantial synteny between isolates A1 and A2 and an overall preservation of genome synteny between A1 isolate and assemblage B (Adam et al., 2013, Franzen et al., 2009). For all five chromosomes, central regions showed greater levels of similarity than the more distal regions. Only a few assemblage-specific genes are found in assemblage B, except for *vsp* gene family where the repertoires of assemblage A and B are completely different (Adam et al., 2013, Franzen et al., 2009). Interestingly, all meiosis-related genes identified in assemblage A are found in assemblage B. Several are well conserved, but some are not as conserved and some even show deletion or insertion in important regions (Franzen et al., 2009). It has been shown that exposure to epithelial cells resulted in significant changes in *Giardia* gene expression and the release of several *Giardia*-produced metabolic enzymes: (i)

surface lectins, (ii) giardin proteins, (iii) variant surface proteins (VSPs), and (iv) cysteine proteases (CPs) (Ringqvist et al., 2011, Ringqvist et al., 2008, Roxstrom-Lindquist et al., 2006). Most of these *Giardia*-expressed parasitic factors are involved in the trophozoite attachment via its ventral disc (Cotton et al., 2011).

1.5.3.1/ Lectins

Lectins are surface sugar-binding proteins that are highly specific for their sugar moieties. They play a role in many different biological functions: from regulation of cell adhesion to glycoprotein synthesis; they also have a role in the recognition phenomena involving cells and proteins. They are also found during the colonisation of a new host by pathogens. Indeed, some viruses use them to attach to host cells. In 1986, two studies have demonstrated the presence of *Giardia*-secreted lectin proteins during *Giardia* infection (Lev et al., 1986, Farthing et al., 1986). Farthing et al showed that *Giardia* possesses a surface membrane-associated lectin with specificities for glucosyl and mannosyl residues from the experimental cell line they used, rabbit erythrocyte. Lev et al confirmed this discovery by showing the presence of *Giardia*-secreted lectin during parasite attachment to enterocytes. Mannosyl residues are present on the surface of mammalian intestinal epithelial cells and appear to be receptors for bacterial mannose-binding lectins, which led to the conclusion that lectins may participate in the trophozoite attachment to the host enterocytes (Farthing et al., 1986). Lev et al also suggested that lectins may be responsible for the microvillus shortening. Even if the presence of *Giardia*-secreted lectins have been demonstrated decades ago, how these lectins and to what extend they participate in the parasite attachment remain unclear and need further investigation.

The best characterised *Giardia* lectin was observed in haemagglutination assay after trypsin treatment of sonicated trophozoites; it was therefore called trypsin activated *Giardia* lectin (taglin) (Lev et al., 1986). Taglin was shown to be a possible agent contributing to the microvillus shortening observed during *Giardia* infection (Farthing et al., 1986). A 30 kDa antigen has been identified as a common immunoreactive protein displaying the lectin activity; however this protein was not characterised until 2003. Weiland and collaborators identified the proteins

Table 1. 5: Comparison of genomic features of *Giardia* genomes sequenced
(modified from Adam et al., 2013 with permission)

Isolate Genotype	A1[a]	A2[b]	B[c]
Genome size (Mb)	11.7	10.7	12.01
Percent-coding region (%)	82.76	89.54	86.35
GC (%)	49.24	49.04	48.25
Heterozygosity (%)	< 0.01	0.037	0.425
Contigs	306	239	544
Protein-coding ORFs	5,901	6,724	7,477
ORFs with assigned function	2,905	2,900	3,946
ORFs without assigned function	2,996	3,824	3,531
ORFs in asserted pathways	710	656	942
ORFs not in asserted pathways	5,191	6,068	6,535
ORFs with assigned function, but no pathway	2,196	2,245	3,005

[a] Sub-assemblage A1, WB C6 clone isolate

[b] Sub-assemblage A2, DH isolate

[c] Assemblage B, GS isolate

recognised by taglin as alpha-1 (α1-) and alpha-2 (α2-) giardins, two proteins present in *Giardia* cytoskeleton and involved in the attachment of trophozoites to intestinal cells (Weiland et al., 2003)

1.5.3.2/ Giardins

Giardins are *Giardia*-specific proteins; they are components of *Giardia* cytoskeleton within the ventral disc. Four major classes have been identified so far: alpha (α)-giardin, beta (β)-giardin, gamma (γ)-giardin and delta (δ)-giardin (Feliziani et al., 2011, Wei et al., 2010, Palm et al., 2003, Peattie, 1990, Peattie et al., 1989).The α-giardin form a large class of annexin-like proteins, encoded by 21 different genes (named α1-α19), having significant sequence similarities. They are located at the outer edges of the ventral disc microribbons and are found in both assemblages A and B with a genome synteny (Feliziani et al., 2011, Wei et al., 2010). Although Steuart et al determined that α2-giardin was an assemblage A-specific protein, a recent study showed the presence of a α2-giardin gene-like in assemblage B, with 92 % AA identified and encoded from gene in a syntenic position (Franzen et al., 2009, Steuart et al., 2008). α1-giardin was further characterised than any other α-giardins, and was shown to be an immunodominant glycosaminoglycan (GAG)-binding protein associated with the plasma membrane (Weiland et al., 2003). GAGs are proteoglycans involved in various processes including adhesion of microbes to eukaryotic cells. The most widely expressed GAG on the intestinal epithelial cells is GAG heparan sulphate (Ianelli et al., 1998); α1-giardin was shown to bind to highly sulphated sugars such as heparin and heparan sulphate in a calcium-dependent manner (Weiland et al., 2003). Alpha-1 giardin was also shown to be partly localised on the surface of excyzoites. All of this suggests that α1-giardin may play a role in the early stage of the host-parasite interaction. Not all the functions of α-giardins have yet been identified; but all the α-giardins were shown to be expressed throughout the parasite life cycle (Palm et al., 2005).

Beta (β)-giardins are striated fibre-assemblin-like proteins and are closely associated with microtubules and δ-giardin which are component of microribbons (Feliziani et al., 2011). There are three β-giardin isoforms identified (Palm et al., 2005). They are adhesive-specific proteins which assemble in 2.5 nm filaments

which are then further assembled into the superstructure of the dorsal ribbons of the ventral disc. This suggests a primarily structural role for this protein (Crossley and Holberton, 1985, Holberton, 1981). Feliziani et al have recently showed that there was neither structural difference between assemblage A and B β-giardin, nor genetic difference. However, by Immunofluorescence assay (IFA), they have observed differences into the organisation of the protein within the ventral disc between both assemblages. Gamma (γ)-giardins have been identified as microribbon proteins and three different isoforms have been isolated. However, their localisation and functions within the ventral disc microribbon are still uncertain (Feliziani et al., 2011, Palm et al., 2005). Delta (δ)-giardins are different from α-giardin and γ-giardin; but share conserved AA motifs with β-giardin suggesting that they belong to the same protein family (Jenkins et al., 2009, Palm et al., 2005). A recent study showed that, in absence of δ-giardin, the ventral disc ability to bind/remain attached to inanimate surface was affected; this led to the conclusion that δ-giardins might be important for the trophozoite attachment to host cells; although the mechanism inducing this failure to attach is still unknown (Jenkins et al., 2009). Although most giardins so far characterised have localised to the ventral disc, how the four classes of giardins interact to provide structure to the ventral disc requires further investigations.

1.5.3.3/ Variant-specific surface proteins

In order to control giardial infections, cellular and humoral immunity is required; the production of $CD8^+$ T lymphocytes correlates in particular with the parasite clearance (Faubert, 2000). To escape the host humoral immune response, several protozoan pathogens can undergo antigenic variation. In *Giardia*, this evasion is achieved by the on-off switching of the expression genes encoding VSPs. VSPs are cysteine-rich proteins covering the entire surface of the parasite, including flagella; the VSP family comprises a repertoire of approximately 200 genes in both human-infecting assemblages, A and B (Ankarklev et al., 2010, Adam et al., 2010, Nash, 2002, Adam, 2001). No identical VSPs have been found in the two assemblages (Franzen et al., 2009). The *vsp* coding regions comprise about 3 % of *Giardia* total genome; it can be raised to 4 % with the inclusion of the upstream

intergenic areas (which may be tasked with control of the *vsp* genes expression) (Adam et al., 2010).

The VSPs size ranges from 20 to 200 kDa; VSPs have variable amino termini (CXXC) and semi-conserved carboxyl termini (containing GGCY motifs). They also have a conserved hydrophilic cytoplasmic tail (CRGKA). Immediately adjacent to this CRGKA domain is a hydrophobic domain that forms the protein membrane anchor (Fig 1.15) (Ankarklev et al., 2010, Adam et al., 2010). The entire surface of *Giardia* is covered with VSPs; the expression of VSPs at the surface is mutually exclusive, except during differentiation and switching when several VSPs are simultaneously expressed (Prucca et al., 2011). Switching has been reported to happen every 6 to 13 generations; however, it depends on the growth conditions, *Giardia* isolate and the specific VSP expressed (Ankarklev et al., 2010). According to Singer et al, there is a selection for and against VSPs in immunodeficient hosts which demonstrates that VSPs may play a key role in the host-parasite interaction. In this study, it was also suggested that clear preferences existed for specific VSPs in different hosts. Antigenic variation may allow the parasite to express VSPs that enhance survival in many different hosts and environmental conditions explaining the large repertoire of VSPs present in *Giardia* genome(Singer et al., 2001). However, the differences between hosts that may affect VSP selection have yet to be determined.

In addition to being involved in the immune evasion and host-parasite interaction, VSPs are also components of cellular signalling. Indeed, some VSPs are specifically palmitoylated on the cysteine residue in the conserved CRGKA carboxyl-terminal motif, located in the cytoplasmic tail (Fig 1.15). This palmitoylation helps to regulate the segregation of the proteins that are detergent resistant to domains on the plasma membrane (so-called lipid rafts) (Touz et al., 2005, Hiltpold et al., 2000). It was also suggested that complement-independent antibody-mediated cytotoxicity was due to specific changes or signalling mediated by the palmitoylated CRGKA cytoplasmic tail (Touz et al., 2005). Other VSPs can be citrullinated on the CRGKA arginine residue by arginine deiminase (Fig 1.15). Citrullation has an important role in the VSP switching mechanism; a mutation of this residue affects the switching frequency (Touz et al., 2008).

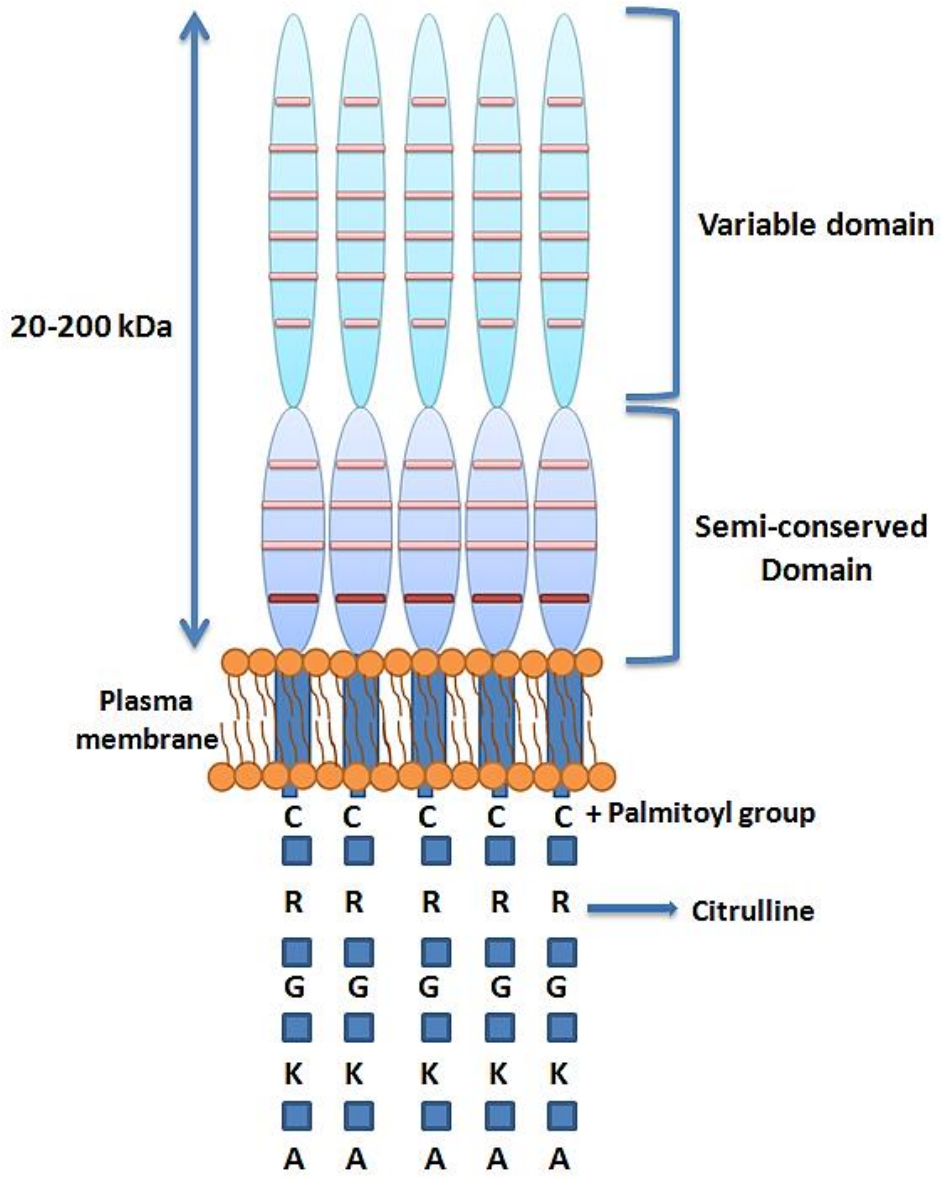


Fig 1. 15: Variant Surface Proteins (VSPs) schematic representation (Based on Ankarklev et al., 2010 with permission). VSPs produce a dense coat by covering the whole surface of the trophozoite. Only one type of VSP is found on the parasite surface (except during differentiation and the switching which occurs every 6 to 13 generations). One VSP usually dominates in a population of parasites; however, a few express other VSPs. VSPs contain about 11-12 % of cysteine AA, most of which is found numerous CXXC ( ) that build up disulfide bonds. The amino terminus is the most variable portion and is called variable domain; this domain seems to be at the interface between the parasite and its environment. The extracellular domain close to the plasma membrane is a semi-conserved domain and contains one or two GGCY motifs ( ). All VSPs have a conserved hydrophilic cytoplasmic tail composed of the amino acids CRGKA. This motif can be modified by palmitoylation of the cysteine residue and by citrullination of the arginine residue.

Our knowledge of the molecular mechanisms involved in the regulation of *Giardia* antigenic variation is still limited. There is no evidence of (i) gene rearrangements, (ii) DNA modifications along with the presence of expression-linked copies, (iii) sequence alterations of the DNA, (iv) telomere-linked transcription requirements. All of these regulation mechanisms have been linked to antigenic variation in others protozoan parasites (such as *Trypanosoma brucei* and *Plasmodium falciparum*) (Lopez-Rubio et al., 2007). However, Kulakova et al suggested that epigenetic mechanisms control VSP regulation. They demonstrated that the expression of *vsp* was not because of special DNA rearrangement. They also showed that antigenic variation occurred in the absence of *vsp* gene movement but was associated with *in situ* chromosome changes in the immediate upstream sequences of the expressed VSP (Kulakova et al., 2006). This suggests that some epigenetic mechanisms may be involved in *vsp* gene transcription activation (Ankarklev et al., 2010, Kulakova et al., 2006). Other studies proposed an alternative hypothesis: *vsp* genes post-transcriptional silencing occurs via a micro RNA (miRNA)-mediated mechanism (Saraiya and Wang, 2008, Prucca et al., 2008). During their investigation, Prucca et al showed that the silencing machinery in *Giardia* processes specifically *vsp* RNAs *in vitro.* Hence, it is possible that chromatin modification and post-transcriptional processes collaborate to regulate VSP expression.

Their transport to the surface represents a vital trafficking pathway in *Giardia*. An immunoelectron microscopy study showed the presence of VSPs in the ER, at the plasma membrane and in PVs, but not in ESVs of encysting cells (McCaffery et al., 1994). However, no intermediate compartments containing VSP cargo protein in the export pathway between the ER and the plasma membrane were identified. It was suggested that, because of a C-terminal transmembrane anchor and a short five AA cytoplasmic domain (CRGKA), these proteins were in direct contact with the cytoplasm and thus potentially able to mediate their own export from the ER. Marti et al showed that it was possible to generate an inducible chimeric reporter that trafficked to the plasma membrane by combining the exodomain of a *Toxoplasma gondii* surface antigen (SAG1) and the conserved C-terminus of a VSP under the control of the CWP1 promoter. This indicates that the

C-terminus of VSPs may be sufficient to direct a foreign antigen to the plasma membrane (Marti and Hehl, 2003). On the other hand, by using the entire VSP exodomain in a reporter domain, another study showed that the transport of VSP to the plasma membrane despite the absence of the cytoplasmic domain (Touz et al., 2003). However, due to the fact that a control experiment using a heterologous exodomain was not included in this study, it remained unclear whether the reporter lacking a C-terminus targeting signal was exported to the plasma membrane simply by co-transport through interaction with endogenous VSPs. A number of reporter proteins with different heterologous components were unable to exit the ER and progress along the secretory pathway without a VSP C-terminus which indicates that the presence of a general default pathway seems unlikely (Hehl and Marti, 2004, Marti and Hehl, 2003). It was suggested that coatomer proteins II (COPII) were recruited directly via the conserved cytoplasmic tail for the export of VSPs (Hehl and Marti, 2004). Once VSPs have left the ER, their transport to the plasma appears to follow a direct pathway with fast kinetics (Marti and Hehl, 2003).

VSPs are the most studied of *Giardia*-pathogenic factors; they are the main mechanism for *Giardia* evasion of the host humoral immune response. Therefore, how exactly antigenic variation occurs in *Giardia* in absence of any immune pressure has to be determined. How a VSP is replaced by another during antigenic switching requires also further investigations, as regulation mechanisms of VSPs expression.

1.5.3.4/ Cysteine proteases

Cysteine proteases (CPs) are proteins degrading polypeptides; they catalyse the cleavage of amide linkages in macromolecular proteins and oligomeric peptides (Fig 1.16). They are commonly encountered in fruits but are also found in animals, viruses or protozoa. CPs are involved in many cellular processes such as growth, PCD (apoptosis and senescence), accumulation and mobilisation of storage proteins (such as in seeds). Moreover, they have roles in signalling pathways and in the response to biotic and abiotic stresses. In parasites, cysteine proteases have key roles in immunoevasion, enzyme activation, virulence, tissue and cell invasion, excystment, hatching and moulting (Sajid and McKerrow, 2002). CPs size ranges

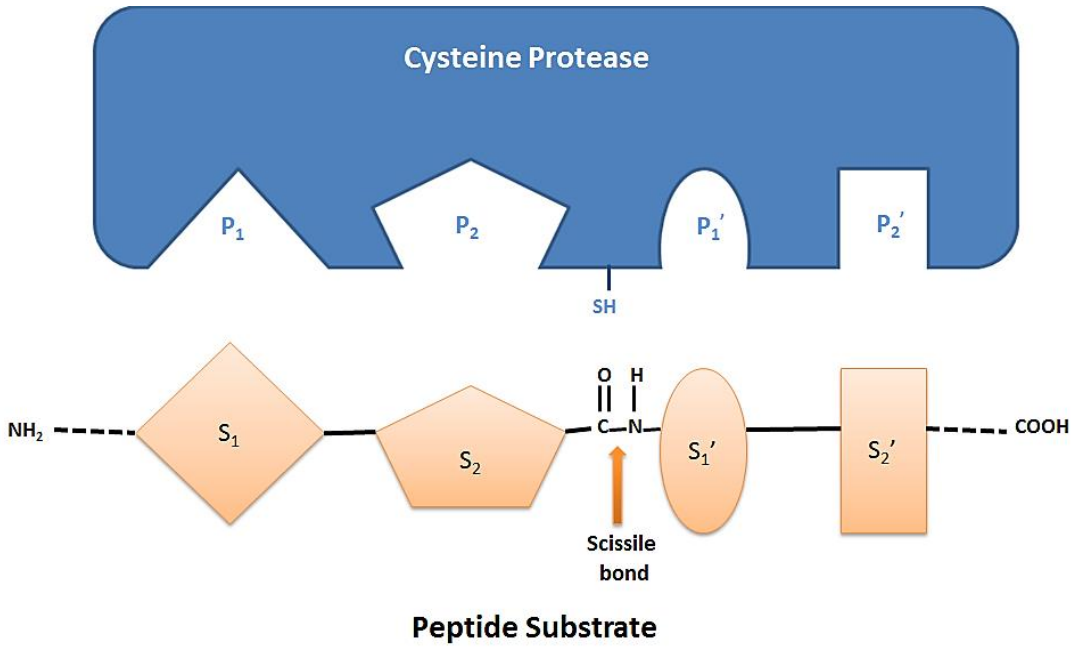


Fig 1. 16: Schematic representation of peptide substrate interaction with an active cysteine protease (Based on Sajid and Mckerrow, 2002, with permission). AA residues from the peptide are denoted by "S" (for substrate) and the CP sub-sites that the peptide interacts with are given the letter "P" (for active site pocket). Sulfhydryl nucleophile, the active site of the CP, is represented as –SH; the peptide bond is attacked prior to hydrolysis (scissile bond). The carboxyl side and corresponding subsites are referred as the prime side and are respectively designated: S_1', S_2', S_n' and P_1', P_2', P_n'. The amino side of the peptide and corresponding subsites are referred as the non-prime side and respectively termed: S_1, S_2, S_n and P_1, P_2, P_n.

from 10 kDa monomers to multimeric complexes of several hundred kDa. Catalysis can be induced either via endoprotease activity (within polypeptide) or via exopeptidase activity (from amino/carboxyl ends). CPs are classified into clans, the most abundant of these clans is CA which is divided into families. C1 is one of these families and is represented by cathepsins B, C, H and L. Cathepsins are synthesised as zygomens; they are activated by proteolytic removal of their N-terminal propeptide (DuBois et al., 2006, Turk et al., 2002). Orthologs of many of the genes encoding for these CPs are present in *Giardia* genome.

The first cathepsin-like protease genes identified in *Giardia* were CP1, CP2 and CP3 which encode cathepsin B-like proteases (Ward et al., 1997). Twenty-seven cathepsin-like protease genes have now been identified in *Giardia* genome (DuBois et al., 2008); they encode three main CP classes: cathepsin B-like, cathepsin C-like and cathepsin K/L-like proteases. These genes, which are not found in groups or tandem arrays, are located throughout *Giardia*'s genome. According to sequence analysis, most of the *Giardia* cathepsins, like their homologues, possess a pro-domain and are likely synthesised as inactive precursors (DuBois et al., 2008, DuBois et al., 2006). It has been demonstrated that the expression of most CPs increases upon excystation. The cathepsin B-like proteases encoding genes seem to be expressed at higher level than the cathepsin C-like or K/L-like protease genes (DuBois et al., 2006). As reported by Sajid and McKerrow, *Giardia* cathepsin B-like proteases lack the "occluding loop" present in cathepsins B of higher eukaryotes. This occluding loop gives cathepsin B proteases their peptidyldipeptidase and dipeptidyl carboxypeptidase activity; as it stabilises the carboxyl end of substrate peptides upon substrate-protease binding (DuBois et al., 2006, Sajid and McKerrow, 2002). Therefore, *Giardia* cathepsin B-like proteases are still believed to exhibit only endopeptidase activity against peptides (DuBois et al., 2008).

Several studies have investigated the possible biological roles for some of the *Giardia* cathepsin-like proteases. Ward et al have shown a decrease of the number of trophozoites during excystation of *G.muris* cysts when CPs inhibitors were added to the media suggesting that CPs were involved in the excystation process during colonisation of a new host by the parasite. They also demonstrated that CP2 may be the agent responsible for breaking down the cyst wall during

excystation. DuBois and collaborator also showed that CP2 gene is expressed at a high level compared to most other genes during vegetative growth and encystation; suggesting that CP2 may have more functions in *Giardia* intracellular processes. Another study also presented data suggesting that a cathepsin C-like protease (encystation-specific cystein protease: ESCP) was the intracellular CP involved in the encystation of trophozoites (Touz et al., 2002). Indeed, the cyst wall proteins, forming the *Giardia* cyst, require proteolytic processing of an alkaline tail before being released from encystation-specific vesicles and added into the cyst wall. In 2006, a study showed for the first time that CPs are involved in the attachment of trophozoites to the host epithelial cells (Rodriguez-Fuentes et al., 2006). By using CPs inhibitors, they observed a decrease of the parasites adhesion. This discovery supports the hypothesis that protease secretion plays a role in the adhesion of *Giardia* to the epithelial cells. It has been proposed that, *in vivo*, CPs may play a role in the degradation of mucin, facilitating either the adhesion of the parasite to host cells or the parasite motility.

Members of CA clan, cathepsin B in particular, have been shown to be either targeted to intracellular vesicle compartments or secreted which suggests that, besides their implications in encystation, excystation and trophozoite attachment to host intestinal cells, *Giardia* cathepsin B-like proteases may be involved in other host-pathogen interaction pathways (Sajid and McKerrow, 2002). A recent study suggested that *Giardia* cathepsin B-like proteases degrade intestinal interleukin-8 (IL8) and are responsible for the host low inflammation reaction to *Giardia* infection (Cotton et al., 2014).

Although 27 genes encoding CPs have been located throughout *Giardia* genome and few of CPs major roles in trophozoite excystation, encystation and attachment were identified, their implication in the disease pathogenesis and host immunity still remains elusive. Analysis of assemblages A and B coding regions indicated an amino acid identity of only 78 % (Franzen et al., 2009). This suggests that CPs may play different roles, specific to one assemblage, in the modulation and/or attenuation of the host immune response (Cotton et al., 2014). Whether CPs have some assemblage-specific roles needs to be assessed and further investigate.

1.5.4/ *Giardia*-secreted factors

Since its discovery 300 hundred year ago, *Giardia* still remains an enigma. Even if at least five of *Giardia* pathogenic factors have been identified (Table 1.6), none has been directly associated with giardiasis so far. Earlier studies have raised the possibility of protein release into the growth medium during *in vitro* incubation of *Giardia* trophozoites (Katelaris and Farthing, 1992). Others, more recently, have tried to identify some possible *Giardia*-secreted proteins which could be the cause of the physiological changes within epithelial cells (Ringqvist et al., 2011, Ankarklev et al., 2010, Ringqvist et al., 2008, Rodriguez-Fuentes et al., 2006).

Cysteine proteases were shown to be secreted by *Giardia* trophozoites upon incubation with different intestinal epithelial cell lines suggesting that CPs may play a role in the attachment of trophozoites to the intestinal epithelia (Rodriguez-Fuentes et al., 2006). It was also suggested that secreted CPs may play a role in the degradation of intestinal mucin as shown by an *in vivo* study (Paget and James, 1994). Another study showed that *Giardia* released three metabolic enzymes: (i) Enolase; (ii) arginine deiminase (ADI); and (iii) ornithine carbamoyltransferase (OCT) in response to interaction with intestinal cells (Ringqvist et al., 2008). Enolase has been shown to be involved in pathogenesis of other pathogens colonising mucosal surfaces such as prokaryotic *Streptococcus agalactiae* and eukaryotic *Candida albican* in which either enolases are either secreted or present at the surface of the organism (Fluegge et al., 2004, Sundstrom and Aliaga, 1994). However, the specific activity of *Giardia*-secreted enolase was not determined during this study; therefore, the role of enolase, if any, during host-pathogen interaction remains to be determined (Table 1.6). ADI and OCT are used in *Giardia* to actively metabolise arginine for energy. Arginine depletion was shown to reduce the ability of intestinal cells to produce nitric oxide (NO), anti-microbial innate defence molecule. NO inhibits both encystation and excystation in *Giardia* (Eckmann et al., 2000). NO was previously shown to inhibit giardial growth but not viability, *in vitro*, and also both encystation and excystation (Eckmann et al., 2000). Therefore, luminal OCT and ADI might reduce the levels of intestinal arginine leading to a lower production of NO by intestinal epithelial cells. Arginine depletion is also known to induce apoptosis in

Table 1. 6: *Giardia* virulence factors and their function in the host-pathogen interaction

Virulence factor	Function	References
Lectin [a]	Attachment to intestinal epithelia	Farthin et al., 1986
Taglin	Microvillus shortening	Weiland et al., 2003; Farthin et al., 1986; Lev et al., 1986
Giardins [a]		
α1- and α2-giardin	Attachment to intestinal epithelia	Steuart et al., 2008; Peattie et al., 1989
	Lectin activity	Steuart et al., 2008; Weiland et al., 2003
other α-giardins	Attachment to intestinal epithelia	Steuart et al., 2008; Palm et al., 2005
β-giardin	Attachment to intestinal epithelia	Palm et al., 2005; Peattie et al., 1990; Crowley and Holberton, 1985
δ-giardin	Attachment to intestinal epithelia	Jenkins et al., 2009
γ-giardin	Undetermined	Feliziani et al., 2011; Palm et al., 2005
VSP [a]	Cellular signalling	Touz et al., 2005; Hiltpold et al.,2000
	Antigenic variation	Adam, 2010 and 2001
	Immune evasion	Singer et al., 2001
	Potential protection against luminal proteases, oxygen and free-radicals	Nash, 2002
	Host-parasite interaction	Touz et al., 2005
Cysteine proteases [a,b]	Excystation	DuBois et al., 2006; Sajid and McKerrow, 2002;Ward et al., 1997
	Encystation	DuBois et al.,2006; Touz et al., 2002
	Attachment to intestinal epithelia	Rodriguez-Fuentes et al., 2006
	Degradation of mucin	Paget and James, 1994
Cathepsin B-like protease[b]	Degradation of intestinal IL-8 leading to attenuation of host pro-inflammatory response	Cotton et al., 2014
Enolase [b]	Undetermined	Ringqvist et al., 2008
ADI [b]	Reduction of intestinal arginine levels and NO	Ringqvist et al., 2008
	Alteration of host innate defence	Ringqvist et al., 2008; Roxstrom-lindquist et al., 2006; Rodriguez-Fuentes et al., 2006
OCT [b]	Reduction of intestinal arginine levels and NO production by intestinal cells	Ringqvist et al., 2008

[a] surface proteins

[b] secreted/released proteins

human cell lines (Philip et al., 2003); human giardiasis patients show an increase rate of intestinal epithelial cell apoptosis (Troeger et al., 2007). A recent study showed the implication of *Giardia*–secreted cathepsin B in the degradation of intestinal interleukin-8 (IL-8) leading to an attenuation of the IL-8-induced neutrophil chemotaxis (Cotton et al., 2014). This shows that *Giardia* is capable of attenuating aspects of the host's pro-inflammatory response. However, whether both assemblage A and B trophozoites are able to do so remains unclear and has to be investigated. The mechanisms involved in the attenuation of IL-8 production by *Giardia*-secreted cathepsin B require also further investigations. All these studies suggest that *Giardia* secretes molecules having potential roles in host immune system evasion, parasite attachment and anti-inflammatory modulation. Yet, all these different analyses were done on *Giardia* lysates after incubation with different intestinal cell lines which could lead to transcriptional changes in the trophozoites during interaction with the host intestinal epithelia. Whether or not trophozoites secrete proteins prior to interacting with intestinal cells, or during colonisation of a new environment, has not been determined yet. Identifying the protein secreted by *Giardia* at early stage of infection could give some insights on how the infection begins and may help understand how the different giardiasis symptoms occur.

During secretory diarrhoea, chloride secretion is increased by pathogen enterotoxins such as the cholera toxin secreted by *Vibrio cholera*. The electrogenic chloride ion secretion from enterocytes is thought to be responsible for the osmotic imbalance that draws fluid into the lumen. The normal physiological secretion can be altered by enterotoxins; thus, the secretion accelerates due to a fluid entry overwhelming the normal chloride absorption (Lucas, 2010). Even if *Giardia* infection causes non-secretory diarrhoea, it has been thought to increase the anion secretion, in a similar fashion to pathogens causing secretory diarrhoea (Hodges and Gill, 2010, Troeger et al., 2007). However, a recent study in our laboratory has shown that a *Giardia*-secreted factor inhibits the enterocyte chloride secretion (Al-Naimi et al, unpublished data). In that study, an Ussing chamber system was used to determine whether or not *Giardia* cultures and supernatants of both human assemblages could influence the *trans*-Epithelial Electrical Resistance (TEER) and

short-circuit current (Isc) of the colonic cell line CaCo-2. After 24 hours incubation with *Giardia,* a significant decrease of the CaCo-2 TEER was observed (Fig 1.17). In order to study the effect of *Giardia* co-culture and supernatants on the CaCo-2 Isc, CaCo-2 cells were incubated with *Giardia* culture/supernatants and either chloride channel inhibitors or activators were added (Fig 1.18). After 24 hours, the cellular Isc for the chloride channel inhibitors, 4,4'-disothiocyanatostibene-2,2'-disulfonic acid (DIDS) and Glybenclamide (Gly 101), was suppressed. Moreover, there was a reduction in both the forskolin and the Uridine 5'-Triphosphate (UTP) on the Isc stimulating reaction (Fig 1.19). A reduced effect was also observed when CaCo-2 cells were incubating with diluted *Giardia* supernatants (Fig 1.19 B.). This effect on chloride secretion by *Giardia* supernatants indicates that *Giardia* secretes a soluble factor which affects the intestinal epithelial cells. CaCo-2 cells were shown to be viable by performing a viability assay (data not shown). This strongly supports the possibility of *Giardia*-secreted factors acting on the intestinal cells chloride secretion. The fact that intestinal ion transport is abrogated by *Giardia* trophozoite-secreted factor suggests that intestinal water transport may also be disrupted by the parasite secreted factor. Moreover, it was previously suggested that an inhibition of chloride secretion lead to a decrease in cell volume (Walters et al., 1992), which suggests that *Giardia*-secreted factor may also have an effect on the intestinal cell volume and therefore morphology. However, what this soluble factor is, the mechanism involved and the effect on intestinal cell morphology require further investigation.

1.6/ Aims of the study

Having a better understanding of how *Giardia* causes diarrhoea is crucial to developing interventions to fight this waterborne disease. To investigate *Giardia* virulence factors, many studies tried to answer an important question: Does *Giardia* secrete soluble mediators which disrupt the effective functions of the intestinal epithelium? (Ringqvist et al., 2011, Ringqvist et al., 2008, Al-Naimi et al., Unpublished data). *Giardia* possesses peripheral vesicles and ESVs suggesting that trophozoites may release soluble factors in the intestinal lumen.

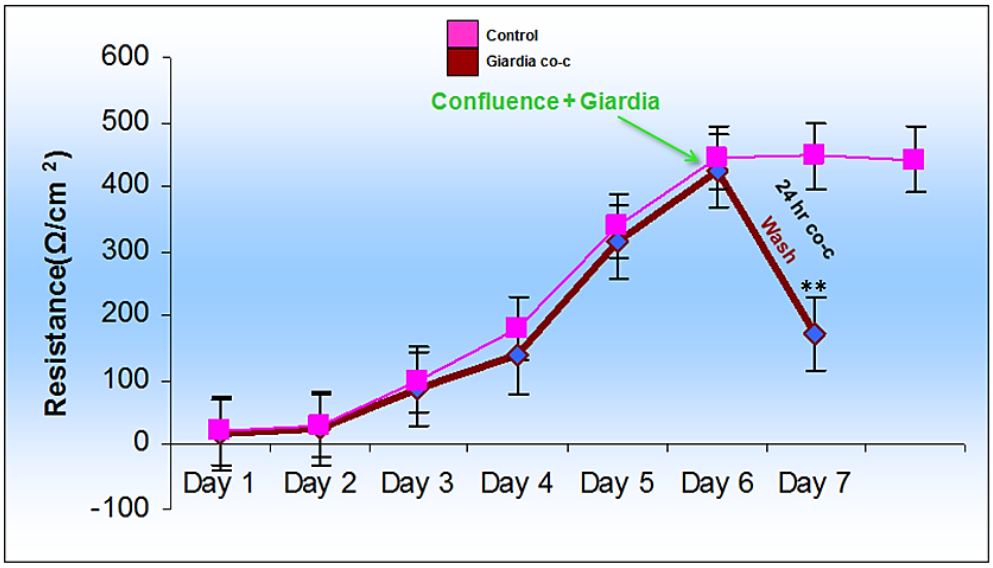


Fig 1. 17: Effect of 24 hours *Giardia* co-culture on the CaCo-2 TEER. CaCo-2 cells were mounted in Ussing chamber. (n=6) Values are expressed relative to controls. The results were assessed by student's t-test and expressed as mean values ±standard error mean (SEM). Significant difference **P<0.01 compared with controls

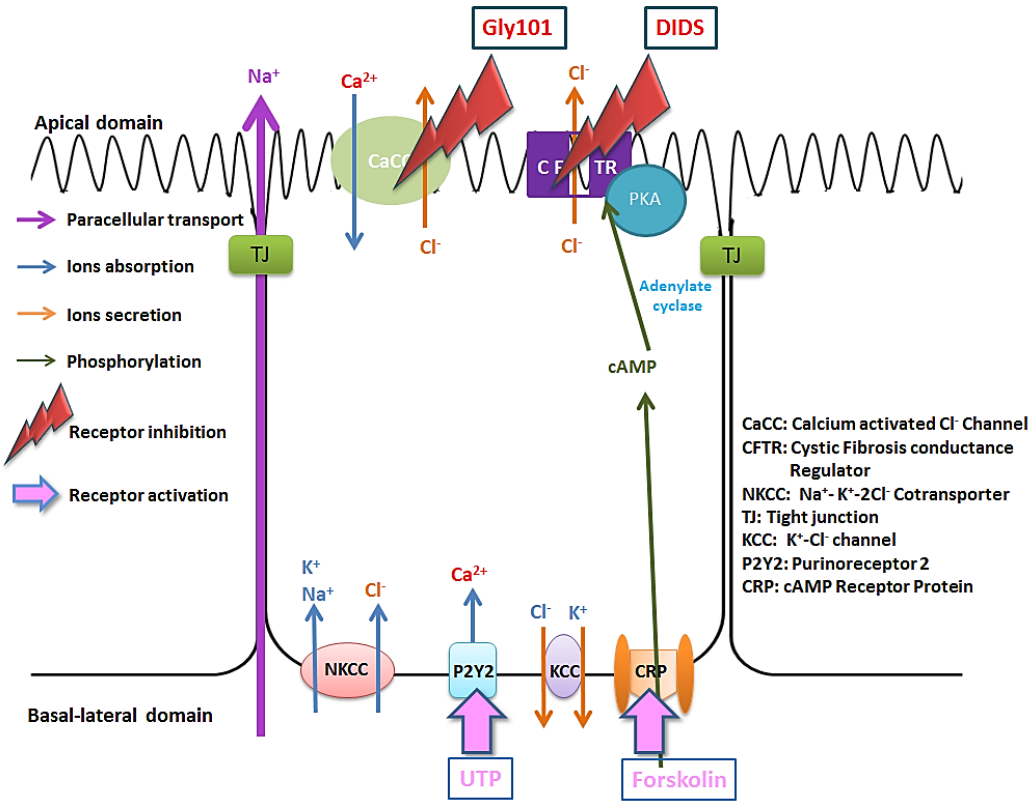


Fig 1. 18: Schematic representation of Chloride secretion activators/inhibitors sites of action. The chloride secretion was inhibited by Glybenclamide (GlyH 101) on the calcium-activated Cl^- channel (CaCC) and by 4,4'-disothiocyanatostibene-2,2'-disulfonic acid (DIDS) on the cystic fibrosis conductance regulator (CFRT). The ion secretion was stimulated by Uridine 5'-triphosphate (UTP) on the Purinoreceptor 2 (P2Y2); and by Forskolin on the cAMP Receptor Protein (CRP).

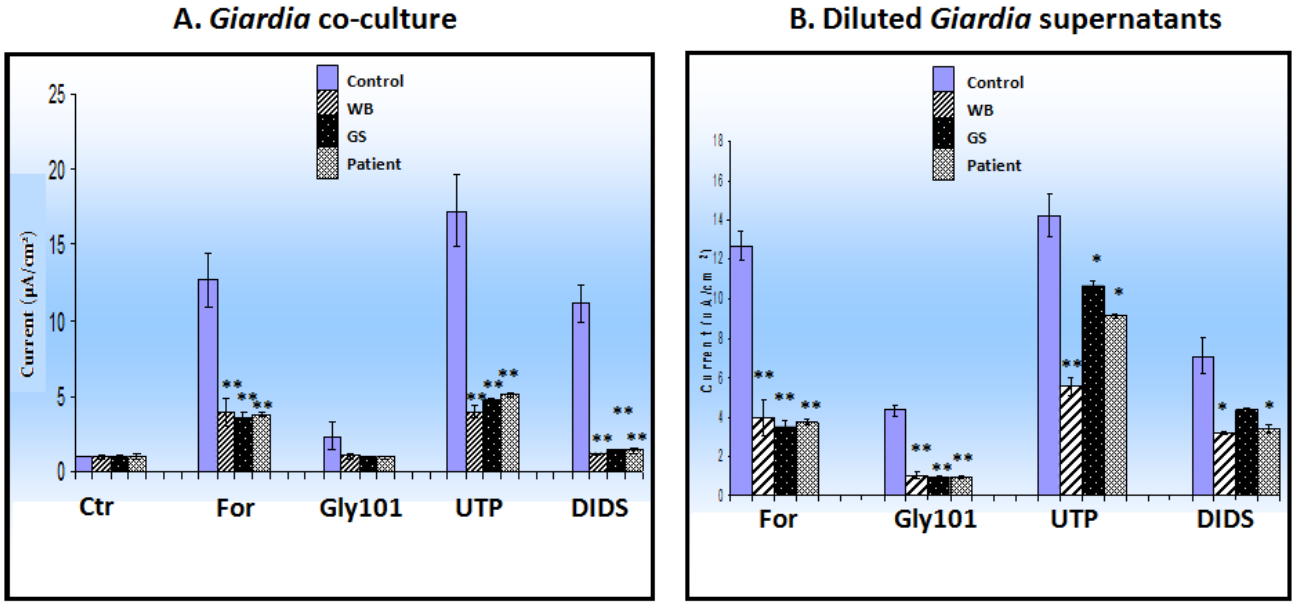


Fig 1. 19: Effect of Forskolin, UTP, Gly 101 and DIDS on human colon tissue (CaCo-2) Isc. A. 24 hours incubation with *Giardia* culture (n=3). **B.** 24 hours incubation with diluted *Giardia* supernatants (n=3). Values are expressed relative to controls. The results were assessed by student's t-test and expressed as mean values ±SE. *P<0.05, **P<0.01 compared to CaCo-2 without *Giardia* co-culture or supernatants.

What these soluble mediators are remains partly unknown and need to be further investigated. Few *in vitro* studies have shown that *Giardia* trophozoites are able to release several enzymes having an effect on intestinal cells during host-pathogen interaction (Table 1.6) (Cotton et al., 2014, Ringqvist et al., 2011, Ringqvist et al., 2008, Rodriguez-Fuentes et al., 2006). However, which proteins are released by *Giardia* at the early stage of the infection, and in what quantity, prior to host-pathogen interaction have never previously been investigated.

The aims of this thesis are:

(i) To investigate the effects of both *Giardia* assemblage A and B supernatants, obtained from *in vitro* culture, on intestinal cells morphology and turgidity via immunofluorescence and microscopic analyses.

(ii) To investigate the effects of both *Giardia* assemblage A and B supernatants, obtained from *in vitro* culture, on intestinal cell main water and ion transporters (AQP3, SGLT-1 and CFTR) via immunofluorescence microscopy and western blot analyses.

(iii) To obtain high quality supernatant free from exogenous proteins for subsequent proteomics analysis.

(iv) To determine quantitative secretion profile for both assemblage A and B via mass spectrometry and bioinformatics analyses.

Chapter II

Material and Methods

2.1/ Materials

All materials were supplied by either Sigma-Aldrich (UK); Sera Laboratories International (SLI, UK); Molecular Probes (UK); Invitrogen (UK), or Santa Cruz Biotechnology (USA) unless otherwise stated.

2.1.1/ Antibiotics, TYI-S-33 and Dulbecco's Modified Eagle Medium (DMEM)

2.1.1.1/ Antibiotics and reagents

Penicillin/Streptomycin (P/S) stock solution – 100 units/ml penicillin G; 100 µg/ml streptomycin. Stored at -20°C

10* Phosphate buffered saline (PBS) – Dissolved 80 g NaCl, 2 g KCl, 26.8 g Na_2HPO_4 – $7H_2O$ and 2.4 g KH_2PO_4 in 800 ml H_2O. pH was adjusted to 7.4 with HCl and volume to 1 l with H_2O.

The solution was sterilized by autoclaving (20 min, 121°C, liquid cycle) and stored at room temperature.

PBS – 100 ml of 10x PBS in 900 ml of dH_2O. The diluted solution was also sterilized by autoclaving and then was stored at room temperature.

B12 Vitamin mix – dissolved 8.5 g B12 Vitamin mix (Sigma-Aldrich, UK) into 100 ml ultrapure water. Solution was sterilised, kept away from light and stored at -20°C.

2.1.1.2/ TYI-S-33 medium (Keister's modification) (Table 2.1)

888 ml sterilised distilled water, 1 g dibasic potassium (K_2HPO_4), 0.6 g monobasic potassium (KH_2PO_4), 1 g NaCl, 5 g neutralized pancreatic digest, 25 g yeast extract, 0.5 g dehydrated bovine bile, 10 g D-glucose, 2 g L-cysteine, 0.2 ascorbic acid (vitamin C), 22.8 mg ferric ammonium citrate, 100 ml albumin bovine serum-heat inactivated (ABS-HI), 20ml B12 vitamin mix, 5 ml P/S antibiotics. In order to reach a 7.0-7.1 pH, NaOH was added to the solution (Table 2.1).

Medium was stored at -20°C, in 50ml falcon tubes.

.

Table 2.1: TYI-S-33 Medium formula for in vitro growth of *Giardia* (Keister, 1983)

Component	Quantity/litre	Final concentration
K_2PO_4	1 g	4.4 mM
KH_2PO_4	600 mg	4.4 mM
NaCl	1 g	
Neutralized Pancreatic digest	5 g	
Yeast extract	25 g	
Dehydrated Bovine Bile	500 mg	
D-glucose	10 g	56 mM
L-cysteine HCL	2 g	16.5 mM
Ascorbic acid	200 mg	1.1 mM
Ferric NH_4-citrate	22.8 mg	
NaOH	Bring pH to 7.0-7.1 (1.65 ml of 10 M solution)	
Heat-inactivated Serum [a] (Bovine or fetal Calf)	100 ml	10%
B12 Vitamin mix [a]	20 ml	

[a] The vitamin B12 mix and heat-inactivated serum were kept as solution at - 20°C. All the components except the heat-inactivated serum and the vitamin B12 mix were dissolved in water, brought to 900 ml; heat-inactivated (56°C for 20 min) serum and vitamin B12 mix were then added. The solution was sterilized on a 0.45 µm-pore-size filter, aliquoted in 50 ml falcon tubes and stored at -20°C. During *in vitro* growth, the organisms were grown in sealed glass tubes nearly filled with medium.

2.1.1.3/ Dulbecco's Modified Eagle Medium (DMEM)

Supplemented DMEM – containing Non-Essential Amino Acid (NEAA) and 4.5 g/l glucose, 1% L-glutamine, 1% P/S, and 10% Foetal Calf Serum (FCS). Stored at +4°C.

2.1.2/ Cell lines

_ The human colonic adenocarcinoma derived epithelial cell line: Caco-2, which was kindly given by Dr Yongping Bao (BioMedical Research Centre, UEA, Norwich, UK) (Roxstrom-Lindquist et al., 2005, Grasset et al., 1984).

_ *Giardia*: WB (assemblage A) and GS (assemblage B) strains and culture supernatants, which were kindly provided by Dr Jorge Tovar et al , from the School of Biological Sciences of The Royal Holloway, University of London (Smith et al., 1982, Meyer, 1970) (Fig 2.1).

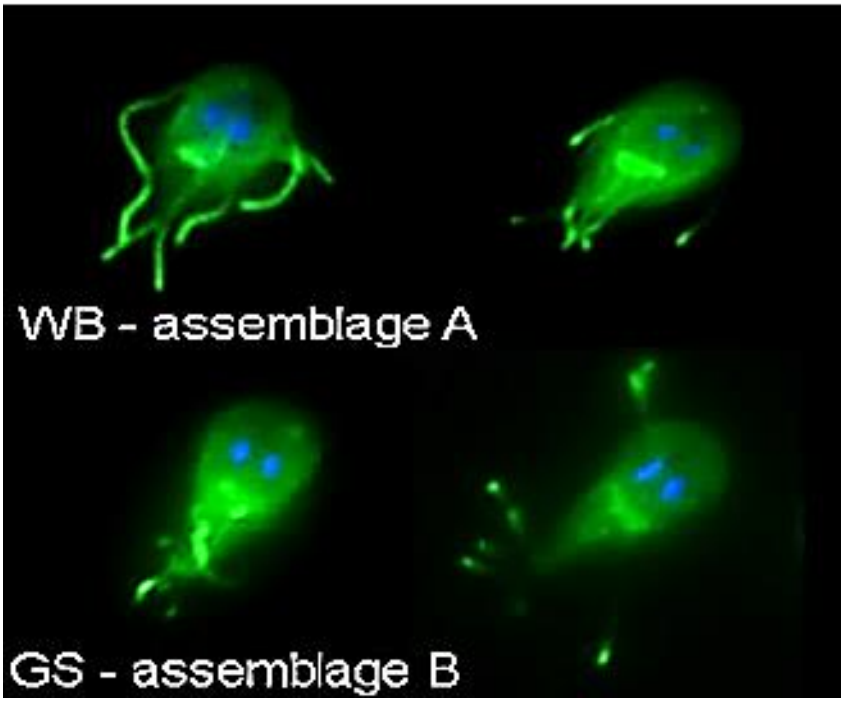


Fig 2. 1: *Giardia* isolates: WB (assemblage A) and GS (assemblage B) (Al-Naimi et al, unpublished)

2.1.3/ *Giardia* Supernatant preparation

Non-supplemented DMEM – containing NEAA [+], 4.5 g/l Glucose [+], L-glutamine [-], Sodium pyruvate [-].Stored at +4°C.

2.1.4/ Protein analysis

2.1.4.1/ Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)

❖ **SDS-PAGE buffers**

30 % acrylamide/Bis solution 37.5:1 (2.6%) Acrylamide – N,N'-methylene-bis-acrylamide Electrophoresis Purity Reagent (Bio-Rad Laboratories, UK).

1.5 M Tris-HCl pH 8.8 – 54.45 g Tris base, 150 ml distilled water (ddH_2O), pH adjusted with HCl.

0.5 M Tris-HCL pH 6.8 – 6 g Tris base, 60 ml distilled water, pH adjusted with HCl.

10% Sodium Dodecyl Sulfate (SDS) – 50 g dissolved in 500 ml distilled water.

10% ammonium persulfate (APS) – 100 mg dissolved in 1 ml distilled water.

1 L of 10 x Gel running buffer pH 8.3 – 30.3 g Tris base, 144 g glycine, 10 g SDS.

1L of 1 x Gel running buffer – 100 ml 10 x Running buffer, 900 ml distilled water.

Loading buffer – 4 ml ddH_2O, 1 ml 0.5 M Tris-HCl pH 6.8, 1.6 ml 10 % SDS, 0.8 ml glycerol, 0.2 ml bromophenol blue. Stored at -20°C.

❖ **SYPRO staining buffers and reagents**

SYPRO Fixing solution – 50 % methanol, 7 % acetic acid in a 200 ml solution.

SYPRO Washing solution – 10 % methanol, 7 % acetic acid in a 100 ml solution.

SYPRO® Ruby protein gel stain – 500 ml solution, Molecular probes, Invitrogen, UK.

❖ **Gels for SDS-PAGE electrophoresis**

Gels were prepared the day before running the electrophoresis, and kept at 4°C. The percentage of acrylamide used varied according to protein size (Table 2.2). The soluble factor size being unknown, SDS-PAGE electrophoresis were run under standard conditions, 10 and 12 % acrylamide separation gels and 5 % acrylamide stacking gels were prepared.

Table 2.2: Separation Polyacrylamide gel recipes (for two gels).

Gel components final volume= 15 ml	Percentage of acrylamide				
	6%	8%	10%	12%	15%
ddH_20	7.9 ml	6.90 ml	5.9 ml	4.9 ml	3.4 ml
30% Bis Acrylamide	3.0 ml	4.0 ml	5.0 ml	6.0 ml	7.5 ml
1.5 M Tris HCl pH 8.8	3.8 ml	3.8 ml	3.8 ml	3.8 ml	3.8 ml
10 % SDS	150 µl	150 µl	150 µl	150 µl	150 µl
10 %APS	150 µl	150 µl	150 µl	150 µl	150 µl
TEMED	12 µl	9 µl	6 µl	6 µl	6 µl
Range of separation of proteins	60 kDa-250 kDa	40 kDa- 100 kDa	20 kDa - 70 kDa	20 kDa- 60 kDa	10 kDa- 40 kDa

Separation gels:

Recipe for 10% Acrylamide gel (15 ml):

1.5 M Tris pH 8.8	3.8 ml
30% Acryl-bisacrylamide (BioRad, UK)	5.0 ml
10% SDS (Melford, UK)	0.15 ml
10% Ammonium Persulfate (APS)	0.15 ml
TetraMethylEthylenDiamineTEMED (BioRad, UK)	0.006 ml
Sterile H_2O	4.9 ml

Recipe for8 % Acrylamide gel (15 ml):

1.5 M Tris pH 8.8	3.8 ml
30% Acryl-bisacrylamide (BioRad, UK)	4.0 ml
10% SDS (Melford, UK)	0.15 ml
10% APS	0.15 ml
TEMED (BioRad, UK)	0.009 ml
Sterile H_2O	6.9 ml

500 µl of isopropanol were added on top of the gels during polymerization. Once gels polymerized, isopropanol was removed, top of the gels were washed once with ddH_2O, and stacking gels were poured.

Stacking gel:

Recipe for a 5% Acrylamide gel (3 ml):

0.5 M Tris pH 6.8	0.38 ml
30% Acryl-bisacrylamide	0.5 ml
10% SDS	0.03 ml
10% APS	0.03 ml
TEMED	0.003 ml
Sterile H_2O	2.1 ml

2.1.4.2/ Western blot

❖ **Western blot buffers and reagents**

10x Transfer Buffer/stock pH 9.2 –390 mM glycine, 480 mM Tris, 0.3 % SDS in a 1 litre solution.

1x Transfer buffer – 100 ml 10x Transfer buffer, 200 ml Methanol in a 1 litre solution.

10x Tris-Buffered Saline and Tween20 solution (TBST) – 200mM Tris, 1.5 mM NaCl. 1 % Tween in a 1 litre solution. pH was set to 7.6.

1x TBST – 100 ml 10x TBST in a 1 l solution.

20x NuPAGE® MES SDS Running buffer– 500 ml solution, Life technologies Ltd, Invitrogen (Paisley, UK).

1x NuPAGE® MES SDS Running buffer – 100 ml 20x NuPAGE® MES SDS Running buffer in 1.9 L distilled water.

Odyssey blocking buffer –. 500 ml solution, Li-Cor inc (UK).

❖ **Western blot gels**

As the sizes of the target proteins vary widely (from 40 to 150 kDa), precast gels with a concentration gradient were purchased from Life technologies Ltd, Invitrogen (Paisley, UK).

NuPAGE® Novex® – 4-12 % Bis-Tris Gels 1.0 mm 10 wells.
4-12 % Bis-Tris Gels 1.0 mm 12 wells.
4-12 % Bis-Tris Gels 1.0 mm 15 wells.

2.1.4.3/ Immunofluorescence Assay Material

4 % Paraformaldehyde (PFA)/PBS – 16 g of PFA was added to 200 ml of just boiled water. NaOH was added until the liquid lost turbidity. 100 ml was added to 80 ml of H_2O and 20 ml of 10x PBS then adjusted to pH 7-8. The solution was aliquoted and frozen at -20°C.

0.5% Bovine Serum Albumin (BSA) – 0.5 mg BSA in 100 ml 1x PBS. Stored at -20°C.

4',6-diamidino-2-phenylindole (DAPI): – 1:500 dilution, 5 µl DAPI in 2.5 ml of 1x PBS. Wrapped in foil and stored at -20°C.

2.1.4.4/ Antibodies and fluorescent dyes

Rabbit anti-AQP3 (Clone H-80), Rabbit anti-SGLT1 and one monoclonal Mouse anti-CFTR (Clone A-3) were purchased from Santa-Cruz Biotechnology (Dallas, USA). A monoclonal Mouse anti-CFTR was purchased from R&D systems (Minneapolis, USA); and another monoclonal Mouse anti-CFTR antibody was purchased from AbCam Plc (Cambridge, UK). Rabbit anti-ZO1 was purchased from Invitrogen (Paisley, UK). Mouse anti-β actin was purchased from Sigma (St Louis, USA). All the antibodies were used as directed below (table 2.3).

Table 2.3: Working dilutions for Antibodies used for Immunofluorescence and Western blot

Primary Antibody	Clone	Working dilution		Company
		Western blot	IFA	
Rabbit anti-AQP3	H-80	1:1,000	1:50	Santa-Cruz Biotechnology (sc-20811)
Rabbit anti-SGLT1	-	1:1,000	1:300	Santa-Cruz Biotechnology (sc-98974)
Rabbit anti-ZO1	-	-	1:400	Invitrogen (40-2200)
Mouse anti-CFTR	13-1	1:250	1:50	R&D systems (MAB-1660)
Mouse anti-CFTR	CF3	1:750	1:50	Abcam (ab2784)
Mouse anti-CFTR	A-3	1:500	1:50	Santa-Cruz Biotechnology (sc-376683)
Mouse anti-β actin	AC-15	1:25,000	-	Sigma (A5441)

Alexafluor 488nm and Alexafluor 594nm were purchased from Molecular Probes (Invitrogen, Paisley, UK). Goat anti-mouse 680 nm and Goat anti-rabbit 800 nm were purchased from Li-Cor (Lincoln, USA). DAPI was purchased from Sigma (St Louis, USA). All antibodies and fluorescent dyes were used as directed below (table 2.4).

All the microscopy was performed using the Axioplan 2 imaging microscope and the Axiovision software version 4. 8.2 unless otherwise stated.

Table 2.4: Working dilutions for Secondary Antibodies and Fluorescent dyes used for Immunofluorescence and Western blot

Secondary Antibody/Fluorescent Dyes	Working dilution		Company
	Western blot	IFA	
Goat anti-Rabbit Alexa fluor® 488	-	1:1,000	Molecular probes (A-11008)
Donkey anti-Mouse Alexa fluor® 594	-	1:1,000	Molecular Probes (A-21203)
IRDye 680LT Goat anti-Mouse	1:7000	-	Li-Cor (926-68020)
IRDye 800LT Goat anti-Rabbit	1:7000	-	Li-Cor (926-32211)
DAPI	-	1:100	Sigma (D9542)

2.1.4.5/ Protein extraction

Phosphatase inhibitor cocktail set II – Calbiochem, Germany.

M-PER® Mammalian Protein Extraction Reagent – Thermo Scientific, Waltham, USA.

2.1.5/ Flow cytometry viability assay

10* PBS – Dissolve 80 g NaCl, 2 g KCl, 26.8 g Na_2HPO_4 – $7H_2O$ and 2.4 g KH_2PO_4 in 800 ml H_2O. pH was adjusted to 7.4 with HCl and volume to 1 l with H_2O.

The solution was sterilized by autoclaving and stored at room temperature.

PBS – 100 ml of 10x PBS in 900 ml of dH_2O. The diluted solution was also sterilized by autoclaving and then was stored at room temperature.

Non-supplemented DMEM – containing NEAA [+], 4.5 g/l Glucose [+], L-glutamine [-], Sodium pyruvate [-].Stored at +4°C.

Trigene – Desinfectant cleaner for laboratory applications, Tristel Solutions Ltd, Snailwell, Cambridgeshire, UK. Trigene was diluted in sterile water at a 1:5 dilution.

Propidium Iodide (PI) – 1mg/ml. Invitrogen, UK.

2.1.6/Proteomic assay

2.1.6.1/ Samples preparation

30 % Trichloroacetic acid (TCA) – 6 ml TCA in 20 ml H_2O (Sigma-Aldrich, Dorset, UK).

100 % Acetone– Sigma-Aldrich (Dorset, UK).

3,000 MWCO PES Membrane Vivaspin Columns – Vivaproducts (Littleton, USA).

25mM Ammonium Bicarbonate (Ambic) – 19.77 g Ambic in 10 ml Ultrapure H_2O.

1% RapiGest™ – Waters MS techonologies (Starstedt, Leicester, UK).

3mM dithiothreitol – Sigma-Aldrich (Dorset, UK).

9mM iodoacetimde – Sigma-Aldrich (Dorset, UK).

Proteomic grade trypsin – Sigma-Aldrich (Dorset, UK).

NuPAGE® Novex® – 4-12 % Bis-Tris Gels, from Life technologies Ltd, Invitrogen (Paisley, UK).

Trypsion – Sigma-Aldrich (Dorset, UK).

2.1.6.2/ Proteomic analysis

- **Mass Spectrometry (MS)**

nanoACQUITY-nLC system – Waters MS technologies (Manchester, UK).

LTQ-Orbitrap Velos mass spectrometer – ThermoFisher Scientific (Bremen, Germany).

- **Data analysis**

Progenesis LC-MS software – version 4.1, Nonlinear Dynamics (UK).

Mascot software – version 2.3.02, Matrix Science.

2.1.7/ Statistical analysis software

Student Package for Social Science software (SPSS) version 18 – SPSS Inc, Chicago, Illinois, USA.

2.2/ Methods

2.2.1/ Tissue culture

All the aspects of tissue culture including media preparation were carried out under sterile conditions in a class II tissue culture cabinet.

2.2.1.1/ *Giardia* trophozoites

❖ **Routine culture of *Giardia* trophozoites**

Trophozoites were cultured in 15 ml glass tubes and maintained in TYI-S-33 medium modified according to Keister (Keister, 1983), and supplemented with 0.5% penicillin / streptomycin. Cells were grown to confluence at 37°C in an atmosphere of 5% CO_2 and 95% air, and were subcultured every 3 to 4 days. To subculture, tubes were put on ice for 15 up to 20 min and then slowly, manually shaken to remove cells from tube wall and 10 µl of cells were added in new glass tubes containing 13 ml of fresh TYI-S-33 medium.

❖ **Preparation of *Giardia* trophozoite supernatant**

Parasites were grown until they reached confluence, and put on ice for 20 min. Then, both WB and GS strains were transferred into 15 ml falcon tubes and separated by centrifugation for 10 min at 3,000 rotations per minute (rpm). Supernatants were collected and stored at -20°C until use for cellular assay on CaCo-2 cells. Pellets were either resupsended in 1:5 Trigene solution (diluted in PBS) for flow cytometry analysis or used for proteomic analysis (Fig 2.2 A and C.).

In order to have *Giardia* WB and GS supernatants as free from contamination with the bovine serum proteins as possible, a "supernatant cleaning" protocol was created. After 20 min on ice, both WB and GS strains were transferred into 15 ml falcon tubes and separated by centrifugation for 10 min at 3,000 rpm. Supernatants were removed and pellets were washed three times with pre-warmed PBS (4 ml, 2 ml and 1 ml respectively) to remove any bovine proteins from the milieu of culture. Pellets were centrifuged for 10 min at 3,000 rpm between each wash. After the third wash, pellets were resuspended in pre-warmed non-supplemented DMEM, to boost the cellular growth and secretion, and incubated for approximately 60 min at 37°C. After 60 min incubation, samples were chilled for 5 min; and supernatants were separated from pellets by centrifugation for 10 min at 3,000 rpm. Supernatants were collected and stored at -20°C until use for either proteomics analysis or cellular assay on Caco-2; Pellets were either resupsended in 500 µl PBS for flow cytometry analysis or used for proteomics

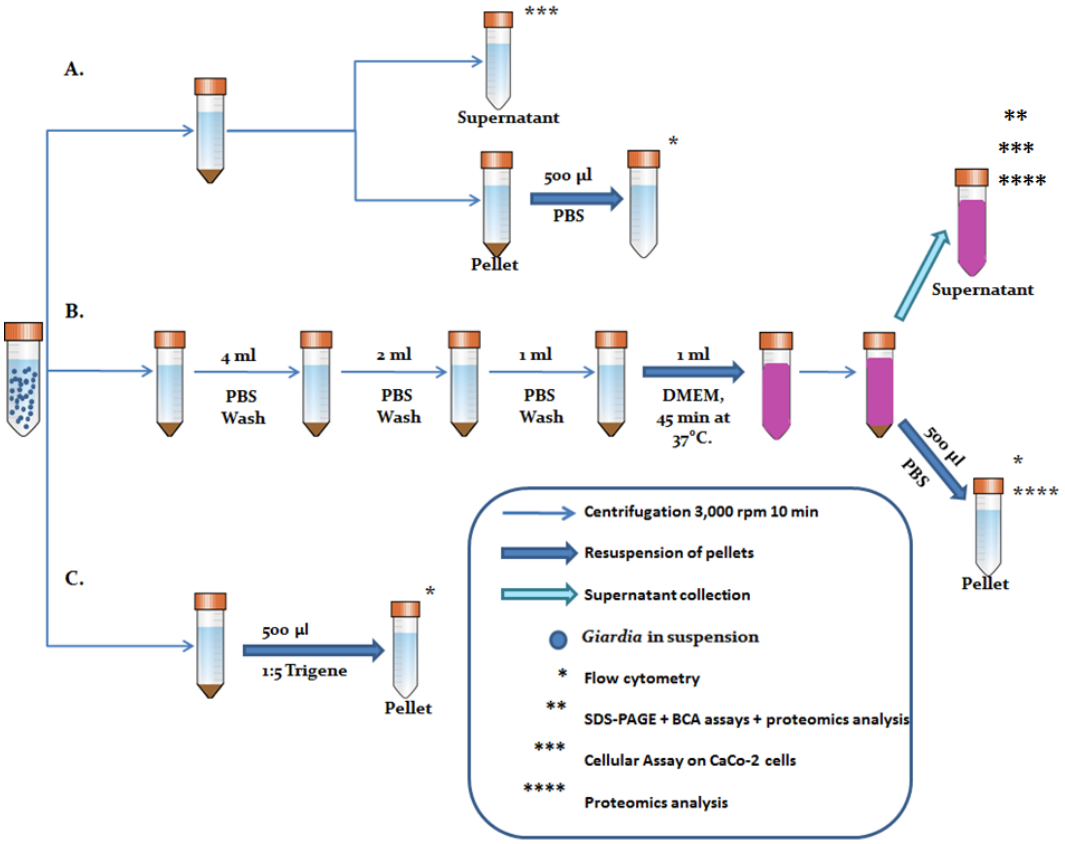


Fig 2.2: Preparation of *Giardia* samples for SDS-PAGE/Western blot assays, flow cytometry, proteomics and cellular assays on CaCo-2 cells. All samples were grown in normal milieu of culture (TYI-S-33), then chilled for 20 min and transferred in 15ml plastic tubes and centrifuged 10 min at 3,000 rpm ( ). **A.** *Giardia* samples prepared directly from culture tubes. Pellets were resuspended ( ) in 500 µl PBS to perform flow cytometry assay; and supernatants were tested against CaCo-2 cells as described in section 2.2.2.3. **B.** *Giardia* cells were washed three times with PBS, incubated in FCS-free DMEM for up to 60 min at 37°C, chilled for 5 min and centrifuged. Pellets were either used for flow cytometry after resuspension in 500µl PBS or stored at -20°C for proteomics analysis. Supernatants were stored at -20°C for either proteomics analysis or cellular assay on CaCo-2. **C.** Samples were killed with 1:5 Trigene (diluted in PBS); to be used as a "dead cell" control in flow cytometry.

analysis (Fig 2.2 B.).

When used for cellular assay, *Giardia* supernatants were diluted in supplemented DMEM either at 1: 250 and 1:1,000 for immunofluorescence, or 1:250, 1:500, 1:1,000 and 1:5,000 for western blot.

2.2.1.2/ Caco-2 cells

❖ **Routine culture of CaCo-2cells**

Cells were grown in 75 cm^2 flasks and maintained in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% foetal calf serum (FCS), 1% L-glutamine, 1% penicillin/streptomycin (100 units/ml penicillin G / 100 µg/ml streptomycin). Cells were cultured at 37°C in a humidified atmosphere of 5 % CO_2 in air. The medium was changed every 48 to 72 hours to remove debris from flasks.

Once 80 % confluence was reached, Caco-2 cells were washed three times with 1x PBS; 1 ml of trypsin was added to the flask which was left at 37°C for 5 minutes. Then 9 ml of media were added to inactivate Trypsin; 1 to 2.5 ml of cells were transferred to a new T75 flask with fresh media up to 15 ml. Medium was changed approximately 24 hours after the passage of the cells and then every 48 to 72 hours.

❖ **Preparation of CaCo-2 cells for protein analysis**

Cells were cultured in DMEM, at 37°C and 5% CO_2, on chambered glass coverslips or in multiwall-plates for immunofluorescence and protein extraction, respectively, in DMEM (Table 2.5). A Neubauer chamber was used to determine the number of cells needed per microliter (Fig 2.3). The following formula was then then employed to obtain the average number of cells per ml:

$$\text{Average number of cells} \times \text{dilution (if used)} \times 10^4$$

The new cell suspension was prepared with a 1:3 relation; for 1ml of cells, 3 ml of DMEM.

Table 2.5: Cell densities for different culture plates for protein extraction and immunofluorescence assay.

Culture Vessel	Growth Size (cm^2)	Cell Density (Per Well)	Plating Media (ml)	Relative Area
48 well-plate	1	10,000	0.2	0.5x
24 well-plate	2	20,000	0.4	1x
12 well-plate	3.8	50,000	1	2x
6 well-plate	9.5	250,000	2	5x
6 cm dish	21	300,000	3	10x

Once CaCo-2 cells reached 80% confluence, cells were washed once with PBS and incubated with different dilutions of both *Giardia* human strains (as described in Chapter 2.1.1.2) for 1, 8 and 24 hours.

At the end of the incubation period, cells were washed once with PBS and then either fixed on glass coverslips with 4% PFA for immunofluorescence, or their proteins were extracted in order to perform the western blot analysis (as described in sections 2.2.3.1 and 2.2.2.1 respectively).

2.2.2/ Protein analysis

2.2.2.1/ SDS-PAGE

❖ **Protein concentration assessment**

The level of protein was determined by Bicinchoninic Acid Assay (BCA protein assay kit, Thermo Scientific, UK) and read on DynamTech plate reader or OMEGA plate reader (BMG Labtech, Germany) (λ= 550 nm). The level of protein was between 120 and 220 µg/µl. The supernatants were used without any dilution. Only the control (TYI-S-33) was diluted to 1/100 and 1/500, its concentration being too high to be read on the plate reader.

❖ **SDS-PAGE**

Proteins were mixed with loading buffer and then loaded in a ratio of 3:1 and boiled for 5 min (Laemmli et al., 1970). Gels were placed in a vertical gel tank system and 500 ml of running buffer 1x were placed between them until it overflowed into the surrounding tank. Gels were loaded with 20 µl of diluted and undiluted supernatants alongside a protein standard molecular weight marker (Precision Plus Dual Colour Protein Standard, BioRad, UK) and run at 100 Volts (V),

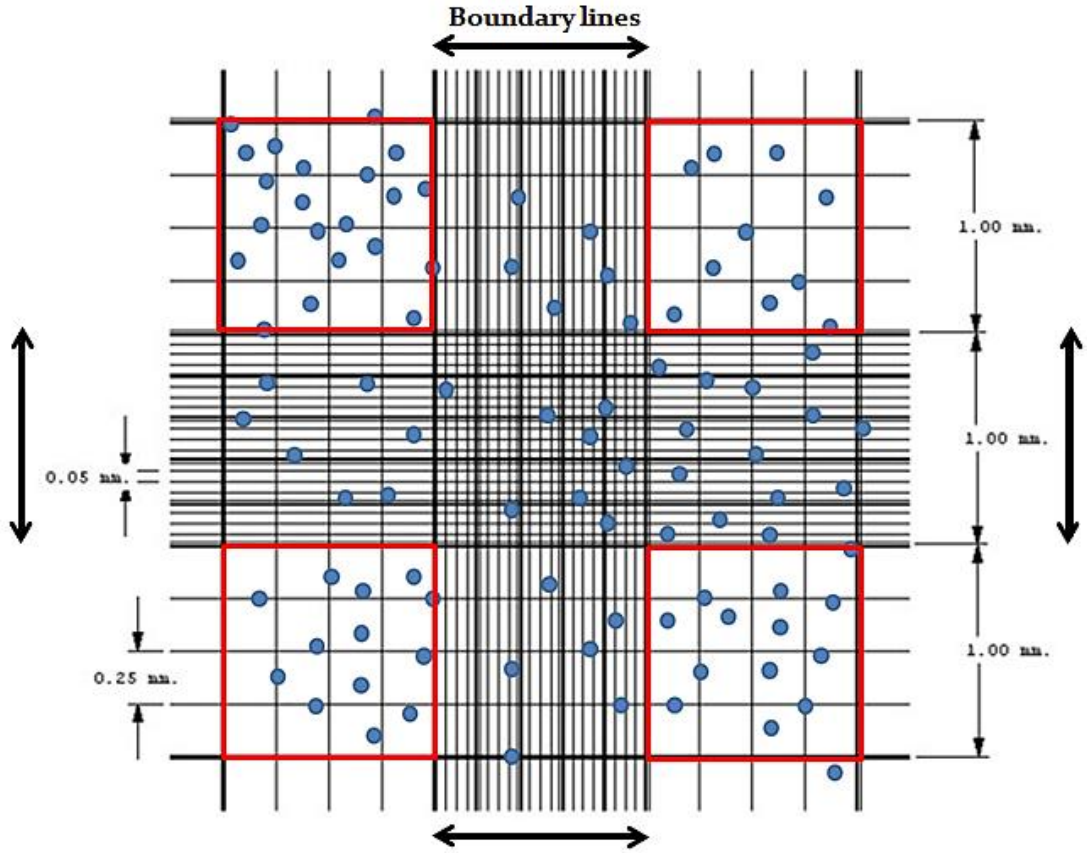


Fig 2. 3: Example of a Neubauer counting chamber. The counting grid is 3 mm x 3 mm in size, subdivided in 9 squares of 1 mm width. Boundary lines of the Neubauer chamber are the centre lines (⟷). The four squares on each side of the Neubauer chamber represent the counting area (□) where cells in suspension (●) are counted. They are subdivided into 16 small squares. The ruled surface below the cover glass is 0.1 mm, giving a volume over each of the 16 small squares of 0.00025 mm^3.

for approximately 2 hours, or until the protein standard indicated good separation at room temperature.

❖ **Gel fluorescent staining with SYPRO**

Gels were immersed twice in a fixing solution bath for 30 min and stained with 40 ml of a pre-diluted SYPRO solution, overnight, in a container protected from the light (Lopez et al., 2000). The next day, gels were washed once with a washing solution for 30 min and then twice with ultrapure water for 5 min.

2.2.2.2/ Western blot

❖ **Protein extraction**

After being seeded in 12 well-plates and incubated with different concentrations of *Giardia* assemblage A and B supernatants (1:250, 1:500, 1:1,000, and 1:5,000). CaCo-2 cells were incubated with M-PER (Thermo fisher, US) for 5 minutes at +4°C in order to extract proteins.

To obtain the best protein concentration during the protein extraction, the optimum volume of M-PER needed was assessed according to Thermo Fisher suggested volume table (Table 2.6), starting at 250 µl decreasing down to 50 µl for a 12 well-plate culture vessel. The optimum protein concentrations were obtained by incubating CaCo-2 cells with 50 µl of M-PER. Once incubated with M-PER, cells were scraped using cell scraper and transferred into 1.5 ml eppendorf tubes. Tubes were centrifuged at 13,000 rpm for 10 minutes. Supernatants were collected into 0.5 ml eppendorf tubes, protein concentrations were assessed by BCA, and tubes were stored at -20°C until use for western blot analysis.

Table 2.6: Suggested volume of Thermo Scientific M-PER Reagent to use for different sizes of standard culture plates.

Culture Vessel/ Plate Size	M-PER Reagent volume
100 mm*	500 – 1,000 µl
60mm *	250 -500 µl
6 well-plate	200 -400 µl
24 well-plate	100 -200 µl
96 well-plate	50 - 100 µl

* Cells grown in 100 mm plates typically contain 10^7 cells (50 mg) and yield ~3 mg total protein depending on cell type.

❖ **Protein concentration assessment**

The level of protein was assessed as described above in chapter 2.2.1.1. Samples were prepared accordingly to their protein concentration in order to load 10 µg of proteins.

❖ **SDS-electrophoresis**

Proteins were mixed with loading buffer (3:1 ratio) and boiled for 5min at 100°C (Laemmli et al., 1970). Precast gels were placed in a vertical gel tank system and 500 ml of 1x NuPAGE® MES SDS Running buffer were placed between them until it overflowed into the surrounding tank. Gels were loaded with 15 µl of samples alongside a protein standard molecular weight marker (Precision Plus Dual Colour Protein Standard, BioRad, UK) and run at 150 V for approximately 1 hour, or until the protein standard indicated good separation, at room temperature.

❖ **Protein blotting**

After SDS-electrophoresis, PVDF membranes were activated in methanol and gels were semi-dry blotted for 30 minutes at 15 V for one gel or 25 V for two gels. Membranes were washed and blocked in Odyssey blocking buffer (Li-Cor, UK) for one hour. Primary antibodies were prepared in Odyssey blocking buffer and left overnight at +4°C on rotary shaker (Table 2.3). Membranes were washed four times in TBST, 5 minutes per wash. Secondary antibodies were also prepared in Odyssey blocking buffer and left at room temperature for 1.5 hours (Table 2.4). Membranes were washed in TBST four times 5 minutes and once in PBS for 5 minutes.

Imaging of infrared tagged antibodies (Li-Cor, UK) was performed on the Odyssey machine (Li-Cor, UK) and associated imaging software.

2.2.2.3/ Microscopy and Immunofluorescence analysis

❖ **Incubation with *Giardia* supernatant**

Once seeded on chambered glass coverslips, with an 80% confluence, CaCo-2 cells were incubated at 37°C with either 1:250 or 1:1,000 dilution of *Giardia* assemblages A and B supernatants, DMEM was used as a negative control showing the standard condition of growth.

❖ **Immunofluorescence assay**

At the end of the incubation period, cells were washed once in PBS and fixed on glass coverslips using 4% paraformaldehyde (PFA) for 20 min. Then, cells were permeabilised with 1% NP_4O for three minutes; and incubated with 30 to 100 µl of primary antibody/BSA block for one hour. 0.5% BSA block for AQP3 antibody and 3% BSA block for SGLT-1, ZO-1 and CFTR antibodies respectively. Coverslips were washed three times in PBS and incubated with 30 to 100 µl of Ig G secondary antibody for 30 min. To finish, cells were washed three times with PBS and incubated with 30 to 100 µl of DAPI for three min. After the nuclei staining of the cells by DAPI, cover slips were hydromounted (Hydromount, National Diagnostics inc, USA) on microscope slides and left to dry for 24 hours at +4°C. This protocol was modified from the procedure firstly described by Pinto et al (Pinto et al., 1983).

❖ **Microscopic images analysis**

All cell images were obtained on a Zeiss Axioplan 2 microscope at either x400 or x630 magnification, unless otherwise stated (Table 2.7). Images were deconvolved using the Axioplan software (version 4.8.2); and analysed using both Axiovision software (version 4.8.2) and ImageJ (version 1.45s).

Table 2. 7: Excitation and emission wavelength ranges of filter sets

Filter Set (Zeiss)	Excitation Wavelength (nm)	Emission Wavelength (nm)
15	546	590
25	400, 495, 570	460, 530, 625
38	470	525
43	545	605
49	365	445

2.2.3/ Flow cytometry viability analysis

2.2.3.1/ Samples preparation

Giardia samples were prepared as stated above in chapter 2.2.1.1 (Fig 2.2).

2.2.3.2/ Flow cytometry

A flow cytometry assay was performed to estimate the ratio of living/dead cells within *Giardia* colonies under different culture conditions (Fig 2.2).

Due to a high number of trophozoites, 250 µl of samples were diluted in 1 ml of PBS (1:4 dilution); and, according to the manufacturer's instruction, 10 µg/ml of propidium iodide (PI) were added in order to assess trophozoites viability via the BD Accuri™ C6 flow cytometer (BD Biosciences, Franklin Lakes, New Jersey, USA), using a blue laser (λ= 488 nm) and an optical filter 585/40. Data were analysed using BD Accuri C-flow software (version 1.0.227.4).

2.2.4/ Proteomic assay

2.2.4.1/ Sample preparation

Giardia supernatant samples were prepared as stated above in chapter 2.2.1.1, using either non-supplemented DMEM containing phenol red (Fig 2.4 A.) or non-supplemented phenol red- free DMEM (Fig 2.4 B.)

Pellets were stored at - 20°C, and the level of protein within supernatant samples was assessed as described in chapter 2.2.1. Once determined the level of protein within supernatant samples, proteins were concentrated using Vivaspin columns with a 3000 Molecular Weight cut-off (MWCO). Samples were centrifuged at 12,000 rcf (Relative Centrifugal Force) for 30 minutes; pellets were then collected. Proteins were washed in 1 ml 25 mM Ammonium biocarbonate (Ambic) at 12,000 rcf for 30 minutes, this step was repeated either once for phenol red-free samples or twice for samples containing phenol red. Proteins were then incubated at room temperature in 50 µl of 25 mM Ambic for one hour, and centrifuged at 3,000 rcf for 2 minutes (Fig 2.5).

The level of protein in concentrated samples was assessed by BCA assay and SDS-PAGE electrophoresis were run to verify that proteins did not degrade during the concentration assay.

Samples were then sent to the Institute of Infection and Global Health, at the University of Liverpool, for proteomic analysis. In order to run a mass spectrometry analysis, pellet samples were first dispensed into low protein-binding microcentrifuge tubes and made up to 160 µl by addition of 25 mM Ambic. Proteins from the pellet samples were denatured using 10 µl of 1% (w/v) RapiGest ™ in 25 mM Ambic; followed by three freeze-thaw cycles and two cycles of 10 min

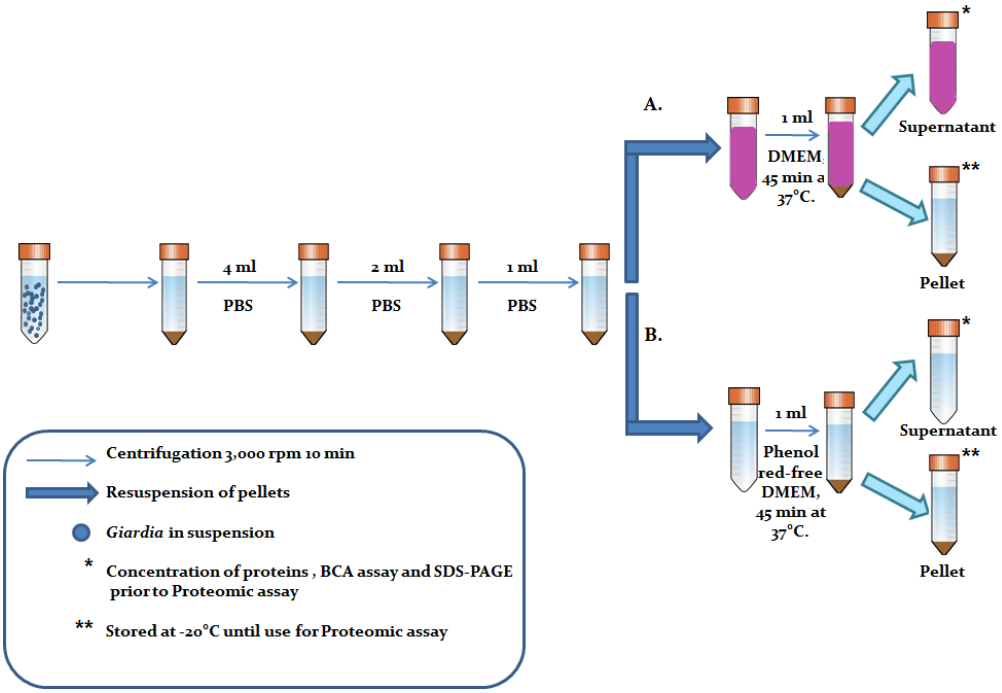


Figure 2.4: Preparation of *Giardia* supernatant and pellet samples for proteomic assay. Parasites were chilled for 20 minutes, transferred into 15 ml falcon tubes and centrifuged at 3,000 rpm for 10 minutes. Supernatants were discarded and pellets were washed three times in warmed 1xPBS (4 ml, 2 ml and 1 ml respectively). Pellets were then incubated for 45 minutes at 37°C either in **A.** non-supplemented DMEM containing phenol red or **B.** non-supplemented phenol red-free DMEM. After 45 minutes incubation, parasites were chilled for 5 minutes, transferred in 15 ml falcon tubes and centrifuged at 3,000 rpm for 10 minutes. Pellets were stored at -20°C. Proteins present in supernatant samples were concentrated prior to perform Proteomic assay.

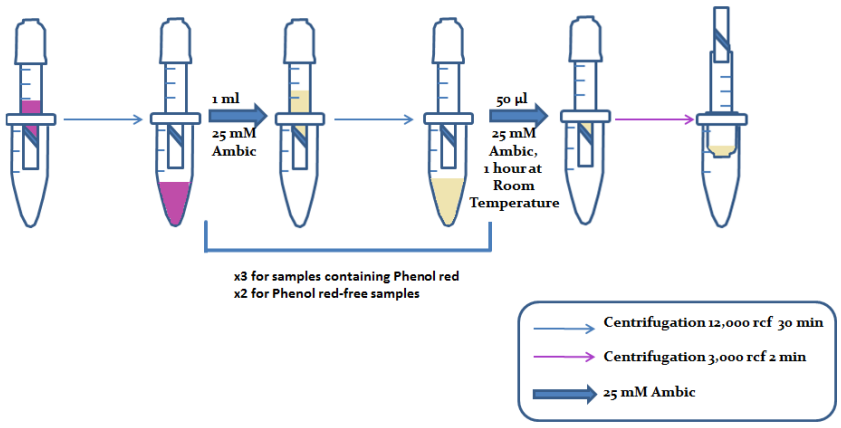


Figure 2.5: Protocol to concentrate proteins contained in *Giardia* supernatant samples prior to Proteomic assay. Supernatants were transferred in Vivaspin columns with a 3,000 MWCO and centrifuged at 12,000 rcf for 30 min. Proteins were washed up to 3 times with 25 mM Ambic (depending on presence of phenol red within DMEM) and centrifuged at 12,000 rcf for 30 min. Then, 50 µl of 25 mM Ambic were added and samples were left at room temperature for one hour; final spin at 3,000 rcf for 2 min to recover proteins. BCA.

sonication in a water bath. Samples were then incubated at 80°C for 10 min and reduced with 3 mM dithiothreitol at 60 °C for another 10 min. Next, samples were alkylated with 9 mM iodoacetamide at room temperature for 30 min in the dark. Proteomic grade trypsin was added a 50:1 protein:trypsin ratio, samples were incubated at 37°C overnight.

In parallel, to maximise protein separation, protein samples were also solubilised and resolved on NuPAGE® Novex®_4-12 % Bis-Tris Gels following manufacturer's instruction. Four gel slices were excised and each digested with trypsin.

2.2.4.2/ Proteomic analysis

❖ **Mass Spectrometry**

A nano-LC MS ESI (ElectronSpray Ionisation) MS/MS analysis was performed. Peptide mixtures from both in solution digestion and 1D-SDS PAGE were analysed by on-line nanoflow liquid chromatography (LC) using the nanoACQUITY- nLC system coupled to a linear ion trap quadrupole mass filter (LTQ)-Orbitrap Velomass spectrometer equipped with the manufacturer's nanospray ion source. The analytical column (nanoACQUITY UPLC^{TM} BEH130 C18 15 cm x 75 µm, 1.7 µm capillary column) was maintained at 35 °C and a flow-rate of 30 nl/min. The gradient consisted of 3-40 % acetonitrile in 0.1 % formic acid for 90 min then a ramp of 40-85% acetonitrile in 0.1 % formic acid for 3 min. Full scan MS spectra (mass to ion charge ratio, m/z, range 300-2000) were acquired by the Orbitrap at a resolution of 30,000. Analysis was performed in data dependant mode, the top 20 most intense ions from MS1 scan (full MS) were selected for tandem MS by collision induced dissociation (CID) and all product spectra were acquired in the LTQ ion trap. Ion trap and Orbitrap maximal injection times were set to 50 milliseconds (ms) and 500 ms respectively. assay and SDS-PAGE were performed to assess protein level and protein degradation prior to proteomic assay. This protocol was modified from the procedure described by Jackson et al (Jackson et al., 2014).

❖ **Data analysis**

Thermo RAW files were imported into Progenesis LC-MS (version 4.1, Nonlinear Dynamics, UK); runs were aligned using default settings and an auto selected runs as reference. Peaks were picked by the software and filtered to include only peaks with a charge state of between +2 and +6. Peptides intensities were normalised against the reference run by Progenesis LC-MS; these intensities were used to highlight differences in protein expression between control and treated samples with supporting statistical analysis (ANOVA and q-values) calculated by Progenesis LC-MS software, Spectral data were transformed to mgf files with Progenesis LC-MS software and exported for peptide identification using Mascot (version 2.3.02, Matrix Science) search engine.

Tandem MS data were searched against the sequence database. Search parameters were as followed: precursor mass tolerance set to 10 parts-per-million (ppm) and fragment mass tolerance set to 0.5 dalton (Da). One missed tryptic cleavage was permitted; carbamidomethylation (cysteine) was set as a fixed modification and oxidation (methionine) was set as variable modification. Mascot search results were further processed using the machine learning algorithm Percolator. The false discovery rate (FDR) was inferior to 1% (< 1%) and individual ion scores were superior to 13 (>13) indicated identity or extensive homology ($p < 0.05$). Results were imported to Progenesis LC-MS as .xml file. At least two unique peptides were required for reporting proteins that were differentially expressed. Results from in solution digestion and gel slices of 1D-SDS PAGE were grouped using "Combine analysed fractions" function in Progenesis LC-MS where statistical analyses were updated. Peptide sequences and their normalised abundances were associated with matching proteins using GiardiaDB online database (www.Giardiadb.org).

Once all the proteins present in the pellet and supernatant samples were identified, pellet and supernatant data were merged to establish which proteins were found in both pellet and supernatant samples, and which proteins were only found in either pellet samples or supernatant samples. This was performed for all datasets obtained both *Giardia* GS and WB isolates. Then, for proteins found in

both pellet and supernatant samples, a supernatant : pellet ratio was calculated to determine which proteins were more likely to be more present within supernatant than pellet. The top 50 of "ratioed" proteins was combined with the proteins only identified in the supernatant samples. Proteins were ranked according to their abundance in the supernatant; this was performed for the four datasets (GS Orbitrap and QeXactive, and WB Orbitrap and QeXactive). For all datasets, proteins were matched to their *Giardia* assemblage A (WB or A2 isolate) or B (GS isolate) orthologs using GiardiaDB online database. The different localisations of the proteins identified and their gene ontology (GO) functions were established via the Uniprot online database (http://www.uniprot.org) and GiardiaDB online database (http://www.giardiadb.org). Secretion profiles from the four datasets were then compared to determine similarities. Proteins of interest were aligned and compared using ClustalW and phylogenetic trees were built via maximum likelihood approach using MEGA software (version 6.06).

2.2.5/ Statistical analysis

Error bars on all graphs presented in this work represent the standard error mean (SEM). The normal distribution and homogeneity of variance of all data were assessed using the Kolmogarov-Smirnoff test and the Levene test respectively. For parametric data, an independent *t* test, a paired *t* test, or a one way ANOVA was performed. For non-parametric data, the Mann-Whitney U test, or the Kruskal-Wallis test was carried out. P-values of less than 0.05 were considered significant.

Chapter III

Effect of *Giardia* supernatant on intestinal epithelial cells and their water transporters.

3.1/ Introduction

The *in vitro* study of transport through the intestine is difficult to realise, due to the complex topology and the diverse cellular heterogeneity that this organ exhibits. Moreover, access to the serosal side of the epithelium is impaired by the presence of subjacent tissues which complicates study of interactions on the basolateral side. A solution to this problem is to use cultured human colon carcinoma cell lines as a model for human colonic tissue epithelia. In 1988, Chantret and collaborator compared 20 different human colonic carcinoma cell lines in order to determine which cell line can spontaneously develop the greatest degree of similarity to normal human intestinal cells. Characteristics such as cell polarity, expression of villin, and ability to undergo enterocytic differentiation under standard culture conditions were assessed (Chantret et al., 1988). The only human colonic carcinoma cell line that exhibited these features at the same time was the human colonic adenocarcinoma derived epithelial cell line-2 (CaCo-2). Indeed, CaCo-2 spontaneously acquired the following enterocytic differentiation characteristics: (i) organization of the cells into a polarized monolayer; (ii) presence of an apical brush border; and (iii) presence of brush border-associated hydrolases such as sucrase-isomaltase, lactase, or alkaline phosphatase (Chantret et al., 1988). Another study showed that CaCo-2 cells possessed similar electrochemical properties to enterocytes. A measurable TEER and a positive Isc which reflect ion transport across the cell membrane were evaluated (Grasset et al., 1984). In 2005, Roxstrom-Lindquist et al showed that the CaCo-2 cell line were a suitable model for the *in vitro* study of *Giardia* effects on intestinal cells (Roxstrom-Lindquist et al., 2005).

As stated in chapter I (section 1.3.2.1), giardiasis symptoms are watery diarrhoea, abdominal cramps, vomiting, malabsorption and weight loss (Cotton et al., 2011, Abel et al., 2001, Robertson et al., 2010, Flanagan, 1992, Wolfe, 1992, Beaumont and James, 1986). According to several studies, diarrhoea may be caused by disruption of the intestinal epithelial barrier (Yu et al., 2008, Troeger et al., 2007, Scott et al., 2002, Chin et al., 2002), increase of apoptosis (Cotton et al., 2011, Troeger et al., 2007, Roxstrom-Lindquist et al., 2005, Chin et al., 2002), and diffuse

shortening of the brush border microvilli leading to malabsorption and anion hypersecretion (Baldi et al., 2009, Troeger et al., 2007, Buret, 2007, Scott et al., 2004, Faubert, 2000, Scott et al., 2000, Cevallos et al., 1995) (Chapter I, Fig 1.14). However, whether or not these effects are due to the trophozoites or any molecules secreted by *Giardia* remains unclear as these studies were performed either on patients biopsies (obtained at the site of infection), or *in vitro*, on intestinal cells incubated with *Giardia* trophozoites. A previous study undertaken in our laboratory showed that prolonged exposure to *Giardia* supernatant reduces chloride efflux of CaCo-2 cells and inhibits intestinal cell Isc (data unpublished) (Chapter I, Fig 1.17-19). This loss of efflux may create a loss in water within intestinal cells leading to a reduction of cell size; suggesting a dysregulation of intestinal water transporters. Indeed, during electrophysiological analysis, cells appeared to be shrunken which led us to look for a surrogate microscopic assay to investigate the effect of *Giardia* supernatant on intestinal cell morphology and water transporter expression.

The aims of the studies described in this chapter were:

(i) to investigate the effects of *Giardia* supernatant on intestinal cell morphology.

(ii) to investigate the effects of *Giardia* supernatant on the main intestinal water transporters: aquaporin 3 (AQP3) (Fig 3.1), sodium-glucose co-transporter (SGLT-1) (Fig 3.2), and cystic fibrosis conductance regulator (CFTR) (Fig 3.3) .

(iii) to determine the chronology of any effects observed.

This was undertaken by incubating CaCo-2 cells with different dilution of *Giardia* WB (assemblage A) and GS (assemblage B) strains supernatants for 1, 8 and 24 hours. After incubation, either immunofluorescence assay followed by microscopic analysis (IFA) or western blot assay were performed on the intestinal cells.

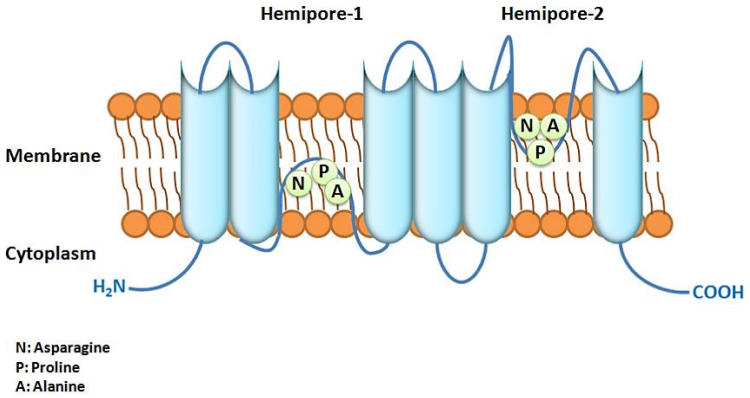


Fig 3. 1: Schematic representation of AQP3 structure. AQP3 is present on the apical side of colonic cells; it transports water and small molecules from the lumen of the intestine to the cell cytoplasm via osmosis. AQP3 is composed of a single chain of approximately 270 AAs and spans the membrane 6 times.

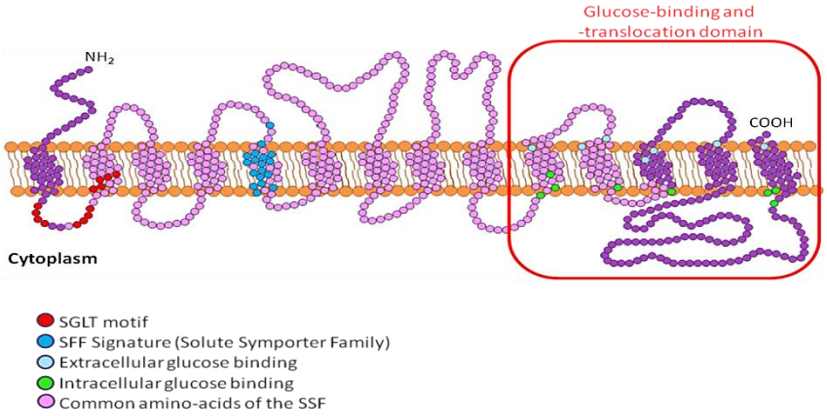


Fig 3. 2: Schematic representation of SGLT-1. SGLT-1 is the main apical co-transporter for active glucose uptake; it is also involved in the water and small molecules (such as Na^{2+}) uptake via osmosis and co transport. SGLT-1 is composed of 14 transmembrane α-helical domains (664 AAs). The glucose-binding and –translocation domain (C5) is located at the COOH-terminal end of the protein.

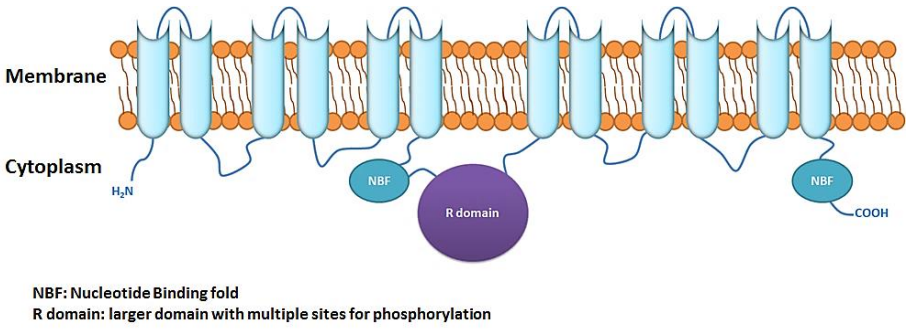


Fig 3. 3: Schematic representation of CFTR. CFTR is a membrane channel that actively transports ions. It contains 1480 AAs. CFTR has several distinct regions, the majority of which span the cell membrane. On the inside of the cell, connected to the membrane spanning domains, are two tightly folded regions that bind nucleotides (NBF). A larger 'R' domain is located between the NBFs; it is a region with multiple sites for phosphorylation. The R domain and the NBFs regulate chloride conductance through the channel.

3.2/ Effect on intestinal cell morphology

3.2.1/ Pilot study to determine potential CaCo-2 shrinkage after 24 hours incubation with *Giardia* supernatant

To investigate the effect of both *Giardia* WB and GS strains supernatants on intestinal cell morphology, CaCo-2 cells monolayers grown on coverslips were incubated for a further 24 hours in the presence or absence of *Giardia* culture medium diluted to 1:250 and 1:5,000 (as stated in chapter II, section 2.2.1.1, Fig 2.2 A.), and compared with controls. After 24 hours, coverslips were fixed with 4 % PFA, and cells were visualized by differential interference contrast (DIC) microscopy using Axiovision (v 4.7.1). Cell length was defined as greatest length between each cell margins, and width as the distance between cell margins at 90° to the length (Fig 3.4 A.). A hundred cells were measured and statistically compared to the controls.

Results showed that significant differences were observed between controls and treated groups. In the case of cells treated with WB supernatant higher concentration, the reduction in size compared to DMEM only was superior to 40 %, for other groups the reductions was more modest but nevertheless significant (Fig 3.4 B.). It was noted however that the TYI-S-33 media itself might have a minor effect on intestinal cells.

These results suggest that even at fairly high dilution, *Giardia* supernatants obtained directly from culture induce cell shrinkage. The inhibition of intestinal chloride secretion previously observed and these results strongly suggest that the intestinal water uptake may be disrupted by *Giardia* supernatant leading to a reduction in size of intestinal cells. However, due to the effect that *Giardia* growth medium may have on intestinal cells, the changes in intestinal cell morphology could not be accurately assessed. Further experiments aim to remove TYI-S-33 from supernatants and improve the visualization of cell margins through the use of markers of cell periphery such as tight junction (TJ) proteins. TJs seal the paracellular space between epithelial cells and maintain the distinct composition of

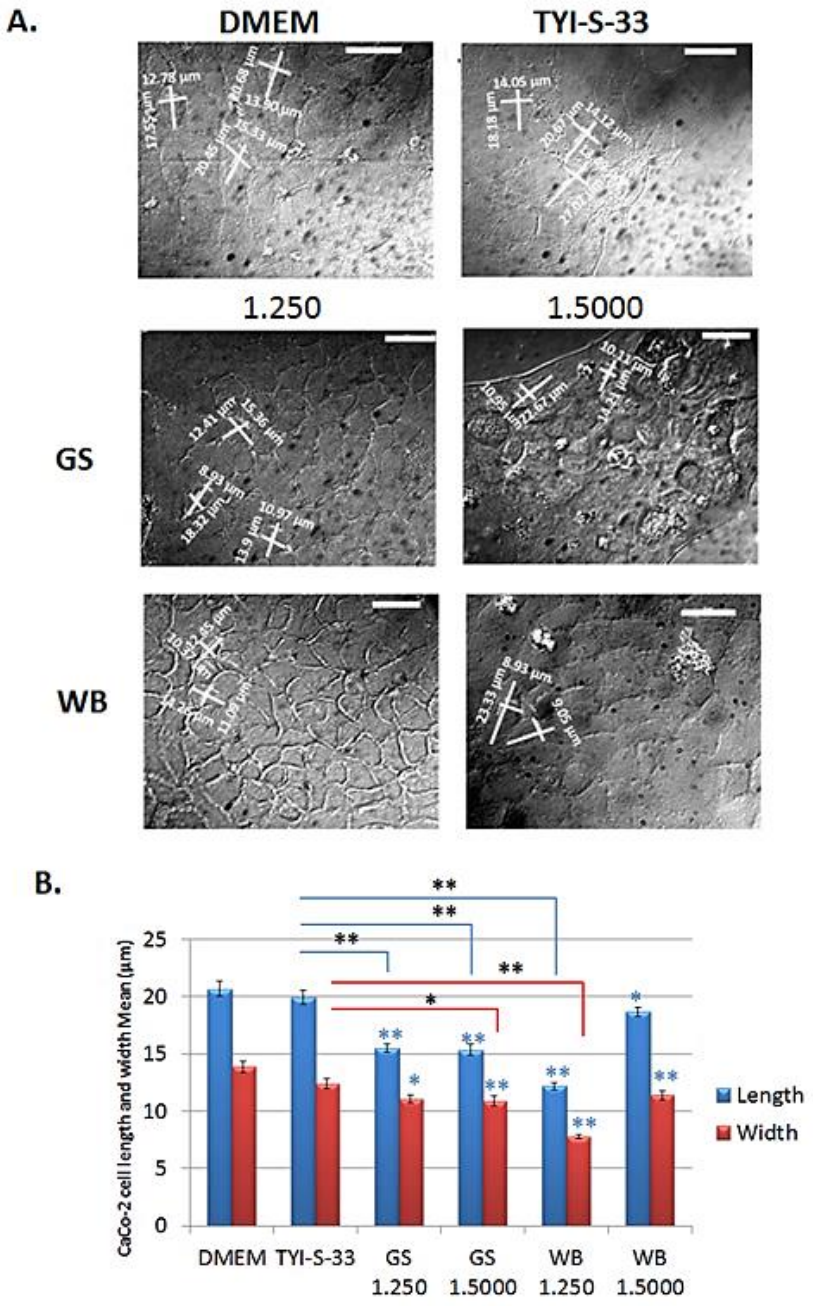


Fig 3. 4: *Giardia* supernatant induces shrinkage of Caco-2 cells after 24 hours incubation. CaCo-2 cells were incubated with either *Giardia* GS or WB supernatant at 1:250 and 1:5,000 dilutions, incubation in both DMEM and TYI-S-33 (1:1,000 dilution) was used as control. After 24 hrs, cells were fixed with 4% PFA and analysed on microscope using DIC light. Cells length and width were measured blindly using the Axiovision software (v 4.7.1). **A.** DIC microscopy showed a decrease of CaCo-2 cells length and width when cells were incubated with *Giardia* supernatant. Scale bar= 20 µm. **B.** Caco-2 cells length and width means showed a significant shrinkage of the intestinal cells when compared to the control: DMEM. *Giardia* milieu of culture, TYI-S-33, appeared to also have some effects on the cells morphology when compared to *Giardia* supernatant.
Values were expressed compared to the control, DMEM (*) and to TYI-S-33 (⎯*⎯). Results were assessed by paired *t*-test and expressed as mean values ± SEM. * p-value < 0.05 and ** p-value < 0.01 (Appendix II). Data represents one independent experiment.

adjacent tissue compartments, they are composed of 8 different proteins forming a complex: occluding, claudin-1, e-cadherin, junctional adhesion molecule-1 (JAM-1), catenin, actin and zonula occludens-1 and 2 (ZO-1 and ZO-2) (Anderson and Van Itallie, 1995) (Fig 3.5). Immunostaining one of the proteins composing the TJs may allow a more accurate measurement of intestinal cell surface area and perimeter as these proteins are present on the whole cell circumference. ZO-1 was shown to be a good indicator of epithelial cells morphology and was used in this study (Umeda et al., 2004).

3.2.2/ *Giardia*-derived supernatant compounds induce shrinkage of intestinal cells after 8 hours.

To investigate the effect of soluble factors from healthy *Giardia* trophozoites on intestinal cells, an improved supernatant compound preparation was undertaken by harvesting trophozoites from confluent culture, washing and resuspending in non-supplemented DMEM for 45 minutes at 37°C (as stated in Chapter II, section 2.2.1.1, Fig 2.2 B.). At this incubation length, all *Giardia* remain viable as shown via cell viability flow cytometry assay (Chapter IV, section 4.2.2, Fig 4.3). CaCo-2 monolayers were prepared as in the previous experiment and were incubated with the diluted supernatants diluted at 1:250 and 1:1,000 and with control media. Three time points for incubation were assessed: 1 hour, 8 hours and 24 hours for each dilution and for each *Giardia* strain. Following incubation, cells on coverslips were fixed and stained for ZO-1 to establish the intestinal cell perimeter. Image J software (v 1.45s) was utilized for measurement and samples were blinded during measurements of surface area and perimeter (Fig 3.6 A., 3.7 A., 3.8 A.). Apoptotic and mitotic cells were excluded from the analysis to have only non-dividing living cells. Between 400 and 800 cells were measured and statistically compared to the control.

Results showed no significant differences between control and treated group for both surface area and perimeter after one hour incubation (Fig 3.6). After 8 hours incubation, results showed significant differences between control and treated groups (Fig 3.7). GS supernatant higher concentration appeared to have a

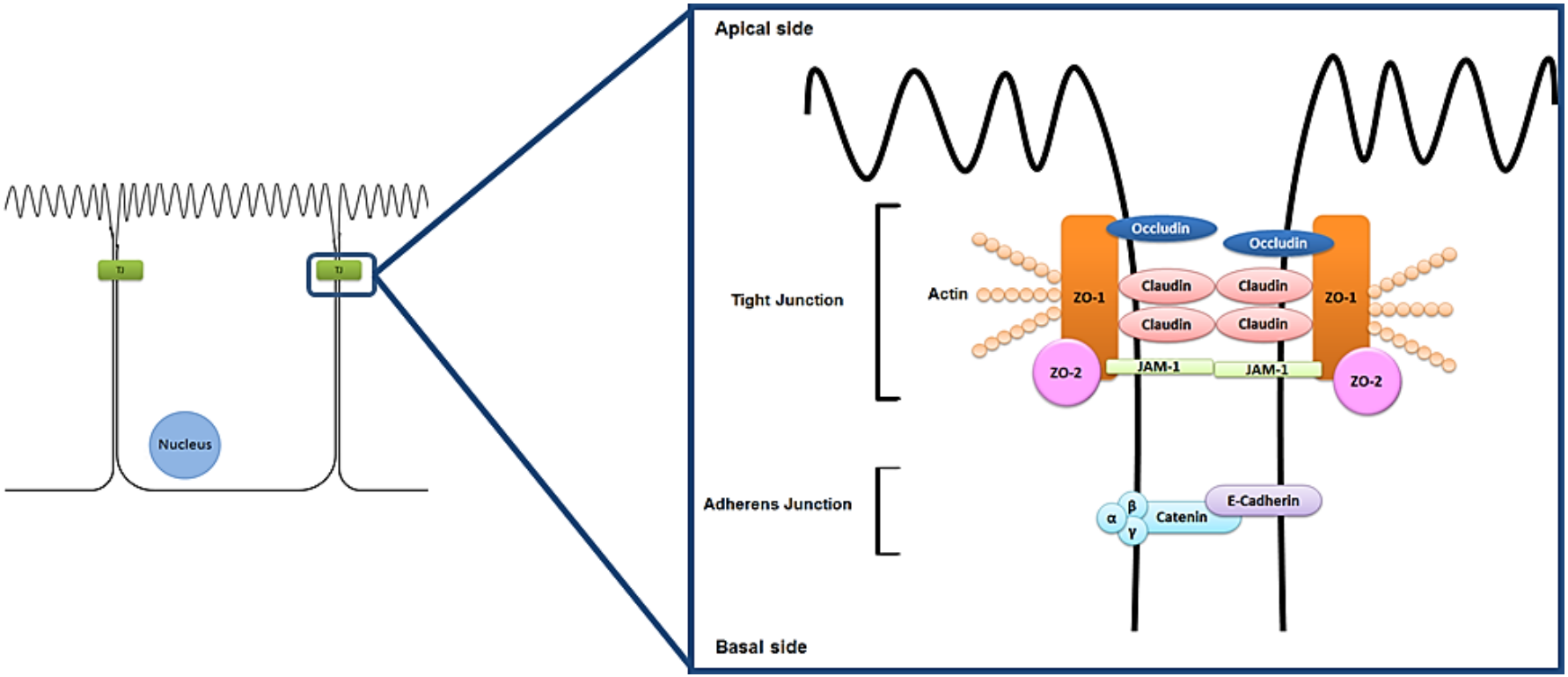


Fig 3. 5: Composition of epithelial tight junctions (TJs) and adherens junctions. Both type of junctions are positioned as continuous contacts at the apical-lateral membrane borders between enterocytes. TJs are formed by homotypic contacts of transmembrane proteins: occluding and claudin which are bound on the cytoplasmic surface directly to ZO-1. The ZO-1/ZO-2 heterodimer binds the cytoskeletal protein actin. Adherens junctions are formed by homotypic proteins: e-cadherins which assemblageociate directly to cytoplasmic proteins α-, β-, and γ-catenins.

more pronounced effect with a reduction in size superior to 15 % compared to the control. GS and WB supernatants lower concentration appeared to have a similar effect with a reduction in size of approximately 12 %. In the case of intestinal cells treated with WB supernatant higher concentration, the effect was less pronounced than for the other groups, a reduction in size inferior to 10 % was observed; yet this difference in size was statistically significant. After 24 hours incubation, results showed significant differences between control and treated group (Fig 3.8), but the effect appeared to be less pronounced than after 8 hours treatment with *Giardia* supernatant. As above, WB supernatant at higher concentration appeared to have a more pronounced effect than other groups, with a reduction in size of 10 %. Other groups had a lesser effect, but still statistically significant, on intestinal cell morphology, with a reduction in size ranging between 5 % and 7 %.

These results suggest that *Giardia*-derived supernatant compounds induce shrinkage of intestinal cells after 8 hours incubation. This reduction in size was still occurring after 24 hours, even if less pronounced than after 8 hours incubation, suggesting that this effect may be a long lasting event. As previously suggested, this change in intestinal cell morphology may be due to a disruption in the cellular water uptake; the loss of water may lead to a reduction in size. However, the effect observed in this study was lesser than the one observed during the pilot study, suggesting that *Giardia* growth medium was increasing the observed intestinal cell size reduction.

3.2.3/ *Giardia*-derived supernatant compounds can induce ruffling of CaCo-2 cell margins.

Apart from the size reduction observed while investigating *Giardia* supernatant effect on intestinal cell morphology (as stated in section 3.2.1.2), cells also appeared to be more wrinkled compared to control. This wrinkling effect is similar to the effect observed in host cells during bacterial infection and called ruffling effect. This ruffling effect was described as a disruption of host cell membrane and cytoskeleton due to the attachment of bacteria or bacterial secreted factors (Guichard et al., 2013). To obtain an index of this ruffling effect,

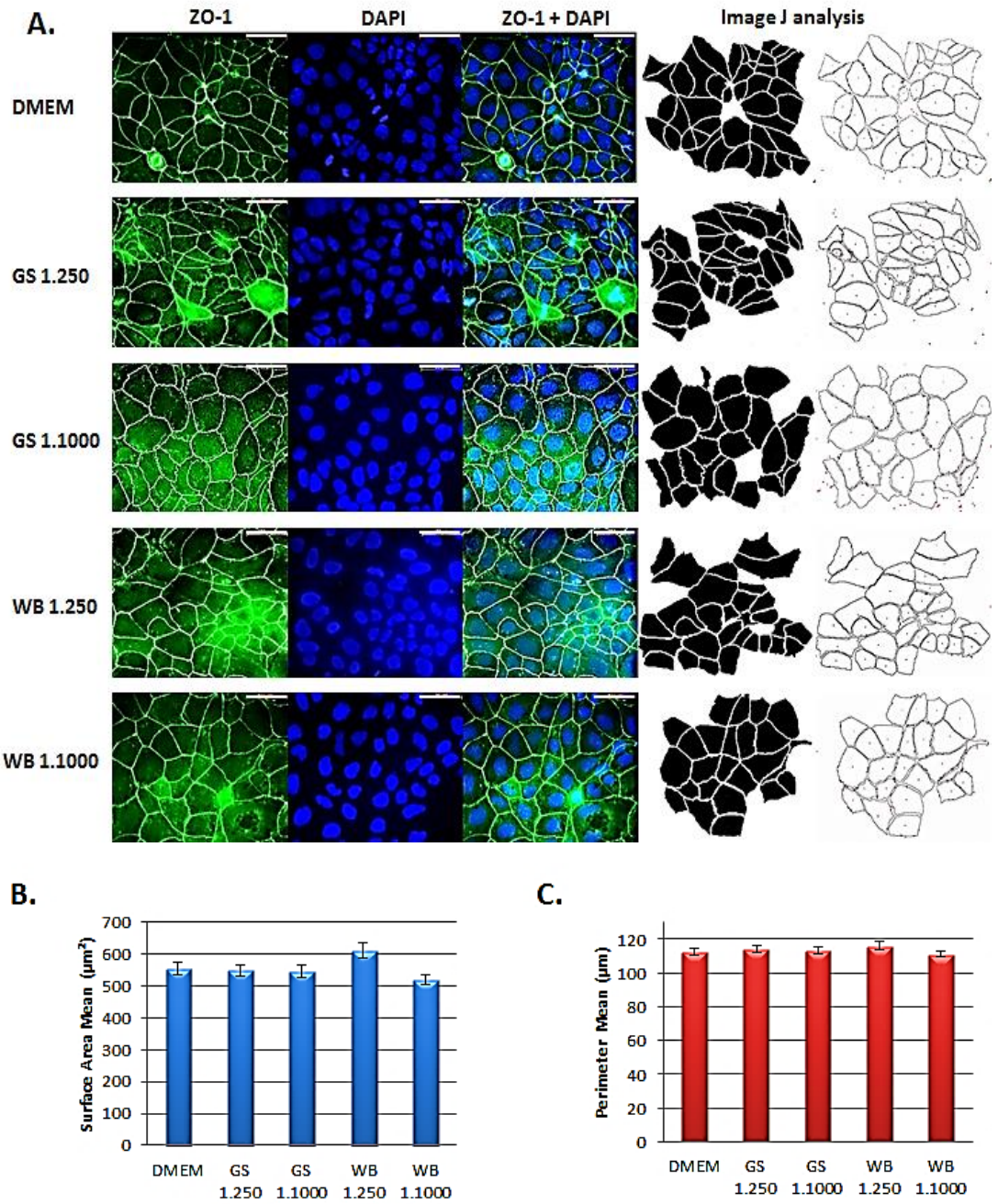


Fig 3. 6: *Giardia* supernatant does not induce any changes in CaCo-2 cells morphology after 1 hour incubation. Caco-2 cells were incubated with either GS or WB strain supernatant, at 1:250 and 1:1,000 dilutions. DMEM was used as a control. After 1 hr, cells were fixed with 4%PFA and stained with anti-ZO-1 antibody to detect changes in the cells morphology; the nuclei were stained with DAPI (blue). ZO-1 was visualised using secondary antibodies conjugated with Alexa fluor 488 (green). **A.** Surface area and perimeter of CaCo-2 cells were measured blindly using Image J software (v. 1.45s) after obtaining images via Axiovision (v 4.8.2), x40 magnification, scale bar = 50 µm. **B.** Variation in CaCo-2 cells surface area compared to the control: DMEM. **C.** Variation in CaCo-2 cells perimeter compared to the control: DMEM. For both surface area and perimeter, results were assessed by non-parametric paired t-tests and expressed as mean values ± SEM, * p-value < 0.05 and ** p-value <0.001 (Appendix III A.). Data represents three independent experiments with biological replicates.

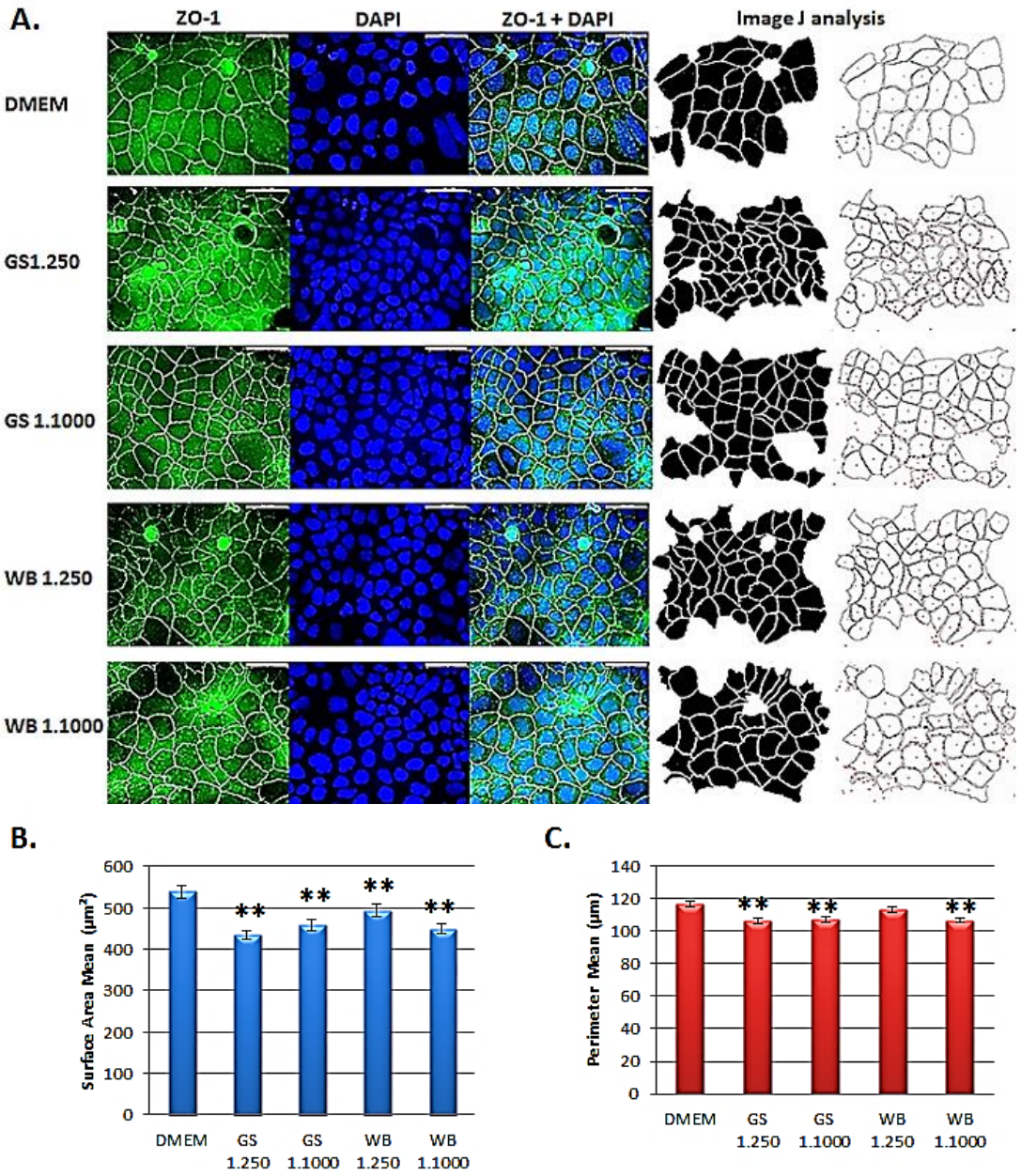


Fig 3. 7: *Giardia* supernatant induces a decrease in CaCo-2 cell size after 8 hours incubation. Caco-2 cells were incubated with either GS or WB strain supernatant, at 1:250 and 1:1,000 dilutions. DMEM was used as a control. After 8 hrs, cells were fixed with 4%PFA and stained with anti-ZO-1 antibody to detect changes in the cells morphology; the nuclei were stained with DAPI (blue). ZO-1 was visualised using secondary antibodies conjugated with Alexa fluor 488 (green). **A.** Surface area and perimeter of CaCo-2 cells were measured blindly using Image J software (v. 1.45s) after obtaining images via Axiovision (v 4.8.2), x40 magnification, scale bar = 50 µm. **B.** Variation in CaCo-2 cells surface area compared to the control: DMEM. **C.** Variation in CaCo-2 cells perimeter compared to the control: DMEM. For both surface area and perimeter, results were assessed by non-parametric paired t-tests and expressed as mean values ± SEM, * p-value < 0.05 and ** p-value <0.001 (Appendix III B.). Data represents three independent experiments with biological replicates.

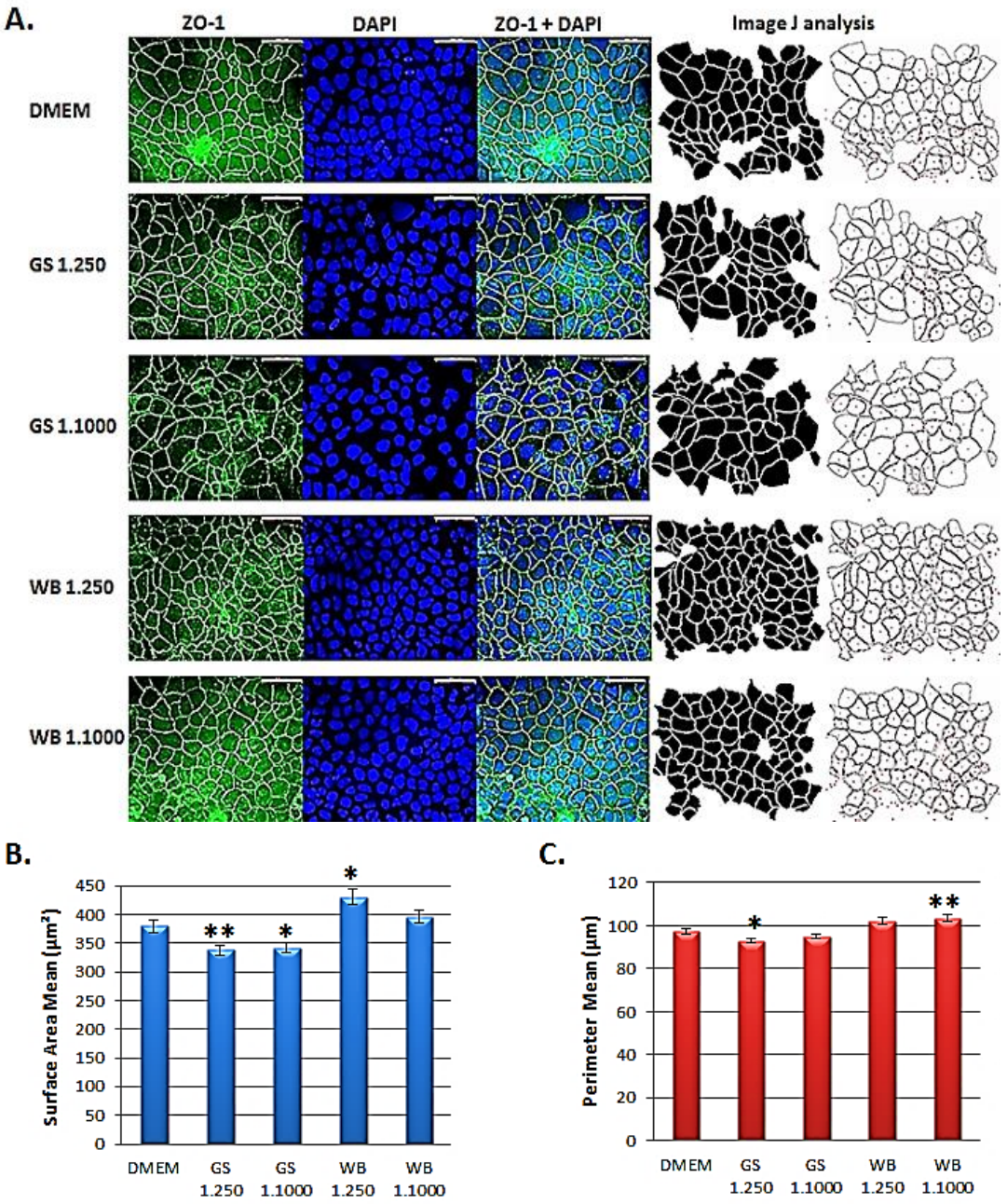


Fig 3. 8: *Giardia* supernatant induces a decrease in CaCo-2 size after 24 hours incubation. Caco-2 cells were incubated with either GS or WB strain supernatant, at 1:250 and 1:1,000 dilutions. DMEM was used as a control. After 24 hrs, cells were fixed with 4%PFA and stained with anti-ZO-1 antibody to detect changes in the cells morphology; the nuclei were stained with DAPI (blue). ZO-1 was visualised using secondary antibodies conjugated with Alexa fluor 488 (green). **A.** Surface area and perimeter of CaCo-2 cells were measured blindly using Image J software (v. 1.45s) after obtaining images via Axiovision (v 4.8.2), x40 magnification, scale bar = 50 µm. **B.** Variation in CaCo-2 cells surface area compared to the control: DMEM. **C.** Variation in CaCo-2 cells perimeter compared to the control: DMEM. For both surface area and perimeter, results were assessed by non-parametric paired t-tests and expressed as mean values ± SEM, * p-value < 0.05 and ** p-value <0.001 (Appendix III C.). Data represents three independent experiments with biological replicates.

the following formula: $\frac{\textbf{Perimeter}}{\sqrt{(\textbf{Surface area})}}$ was used to measure the intestinal cell ruffle (Fig 3.9). Perimeter and surface area are measured in µm and in μm^2 respectively, an index is not define by a measurement unit; therefore, to obtain the ruffling index in this case perimeter was divided by the square root of the surface area, this was done for each treatment group and each time point (Fig 3.9).

Results showed no significant ruffling effect after one hour incubation, yet, there were some significant differences in ruffling between groups treated with WB higher concentration and lower concentration (Fig 3.9 A.). After 8 hours, significant differences in ruffling between control groups and groups only treated with WB supernatant higher concentration were observed. Even if WB supernatant lower concentration did not induce a significant ruffling of intestinal cells, some significant differences were observed between the two WB dilutions (Fig 3.9 B.). After 24 hours, only cells treated with WB supernatant lower concentration showed significant differences in ruffling than the control group. In this case too, there were significant differences between groups treated with WB supernatant higher and lower dilutions (Fig 3.9 C.).

These results show that only WB-derived supernatant compounds induce a significant ruffling of Caco-2 cells after 8 hours and 24 hours. The previous study on the effect of *Giardia*-derived supernatant effect on intestinal cell morphology showed that GS-derived supernatant had a more pronounced effect than WB supernatant on intestinal cell size. These results and the fact that WB supernatant has a significant ruffling effect on intestinal cells suggest that WB supernatant may cause more mucosal damages than GS-derived supernatant. These results also showed significant differences between groups treated with WB higher and lower concentrations at both 8 hours and 24 hours; which suggests that there may be a dose-effect response in this case.

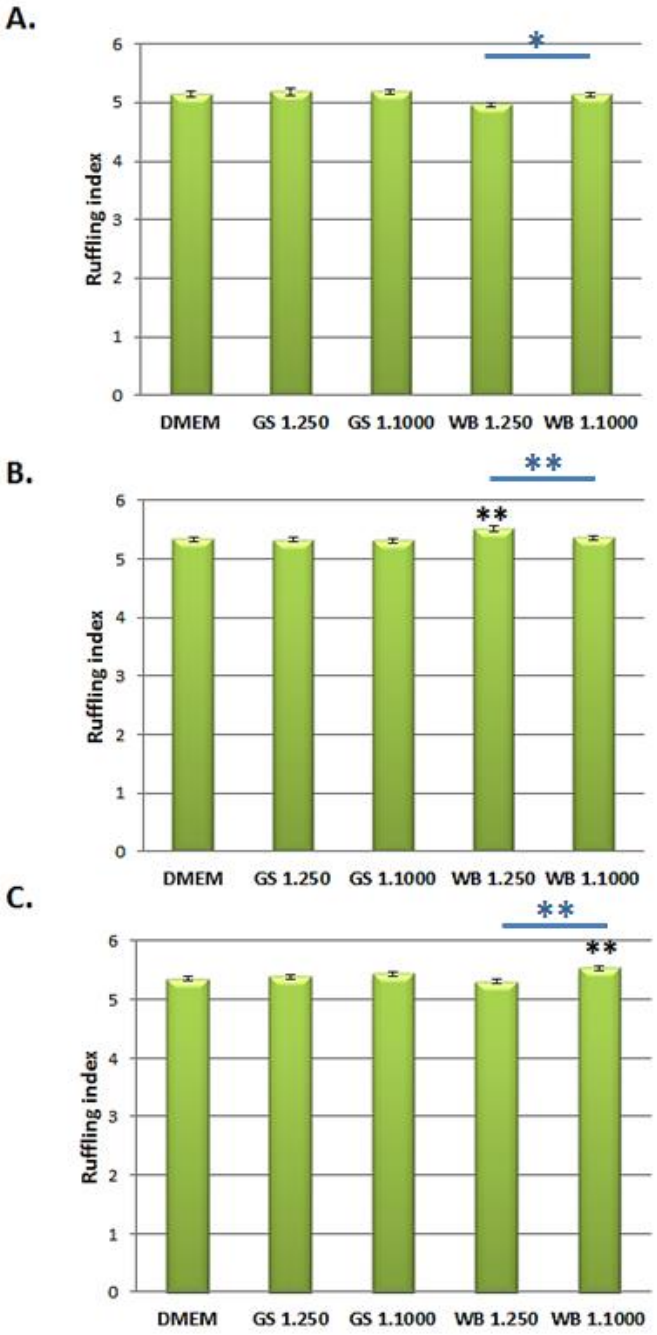


Fig 3. 9: *Giardia* WB supernatant induces some ruffling of Caco-2 cells after 8 and 24 hours incubation. Caco-2 cells were incubated with either GS or WB strain supernatant, at 1:250 and 1:1,000 dilutions for 1, 8 and 24 hours. DMEM was used as a control. After incubation, cells were fixed with 4%PFA and stained with anti-ZO-1 antibody to detect changes in the cells morphology; the nudei were stained with DAPI obtained via Axiovision (v 4.8.2 and the ruffling index was calculated after measuring the surface area and perimeter of intestinal cells using image J. **A.** Ruffling effect after 1 hour incubation. **B.** Ruffling effect after 8 hours incubation. **C.** Ruffling effect after 24 hours. Indexes were compared to the control: DMEM (**), and between different concentration for each strain (——*——). Results were assessed via a panel data approach on Stata software (v 13.1) and expressed as mean values ± SEM, * p-value < 0.05 and ** p-value <0.001 (Appendix IV). Data represents three independent experiments with biological replicates.

3.3/ Effect on intestinal cells water and ion transporters

As previously stated (Chapter I, section 1.5.1.2.), intestinal water uptake occurs by osmosis, co-transport and co-transport + osmosis (Zeuthen, 2010). AQP3 and SGLT-1 are two of the main water transporters present in intestinal cells (Fig 3.1, 3.2). The preceding study on intestinal morphology showed that intestinal water efflux was disrupted by *Giardia* supernatant after 8 hours. This suggests that one or several intestinal water transporters may interact with *Giardia* supernatant. Another study showed that after 24 hours exposure to *Giardia* supernatant, intestinal chloride secretion was inhibited suggesting that *Giardia* supernatant may have an effect on one of the intestinal chloride channels such as CFTR (Fig 3.3) (Al-Naimi et al, unpublished, Chapter I, section 1.5.4.). In this study, the effect of *Giardia* supernatant on the expression and translocation of AQP3, SGLT-1 and CFTR was investigated throughout time course IF microscopy and western blot assays.

3.3.1/ Pilot experiment to evaluate translocation of intestinal AQP3 after incubation with *Giardia* supernatant

To investigate the effect of *Giardia* assemblage A (WB strain) and B (GS strain) supernatants on AQP3. Caco-2 cell monolayers grown on coverslips were incubated for 1hour, 8 hours, and 24 hours in presence or absence of *Giardia* culture medium diluted to 1:250, 1:500, 1:1,000 and 1:5,000 in non-supplemented DMEM (as stated in Chapter II, section 2.2.1.1., Fig 2.2 A), and compared with controls. Following incubations, coverslips were fixed with 4 % PFA, stained for AQP3 and AQP3 distribution throughout cells was visualized via Axiovision software (v 4.8.2). AQP3 was shown to be the main AQP protein present on the apical side of colonic cells such as CaCo-2 cell line (Matsuzaki et al., 2004) (Fig 3.1); CaCo-2 cells were used in this study, AQP3 was therefore chosen to realize the assay.

Results showed that a diffuse staining was observed under all conditions, at the early time point (Fig 3.10 1hr). After 8 hours, AQP3 proteins appeared to localise towards the periphery of the intestinal cells, particularly at higher

concentrations (1:250 and 1:500) for both *Giardia* isolates (Fig 3.10 8 hrs). A similar peripheral protein distribution was observed after 24 hours (Fig 3.10 24 hrs). It was also noted that TYI-S-33 and non-supplemented DMEM media induce the translocation of AQP3 proteins after 8 hours (Fig 3.10 A and D).

These results suggest that *Giardia* supernatants obtained directly from culture may induce the translocation of intestinal AQP3 protein after 8 hours incubation. The inhibition of chloride secretion and the changes in intestinal cell morphology previously observed suggested that the intestinal water uptake may be disrupted by *Giardia* supernatant. This study strongly supports this hypothesis and also suggests that AQP3 may be the intestinal water channel involved in this disruption. However, due to the effect that both growth media may have on AQP3, the changes in AQP3 translocation could not be accurately assessed. In addition, some CaCo-2 cells were clearly not in a confluent monolayer during the experiment which may have affected the reaction between *Giardia* supernatant and the intestinal cells (Fig 3.12 B and C). Further experiments aim to remove TYI-S-33 from *Giardia* supernatant. Moreover, to avoid any starvation of the CaCo-2 cells, supplemented DMEM was used instead of non-supplemented DMEM for the different incubation periods and intestinal cell monolayers were grown to reach confluence.

3.3.2/ *Giardia* supernatant induces the translocation of AQP3 protein

To investigate the effect of soluble factors from healthy *Giardia* trophozoites on intestinal AQP3 protein, an improved supernatant compound preparation was undertaken by harvesting trophozoites from confluent culture, washing and resuspending in non-supplemented DMEM for 45 minutes at 37°C (as stated in Chapter II, section 2.2.1.1, Fig 2.2 B.). CaCo-2 monolayers were prepared as in the previous experiment and were incubated with the supernatants diluted to 1:250 and 1:1,000 and with control media for 1 hour, 8 hours and 24 hours. Following incubation, cells on coverslips were fixed and stained for AQP3. Axiovision software (v 4.8.2) was utilized to observe AQP3 translocation (Fig 3.11 A., 3.12 A., 3.13 A.). In addition, CaCo-2 monolayers grown on 12-well plates were

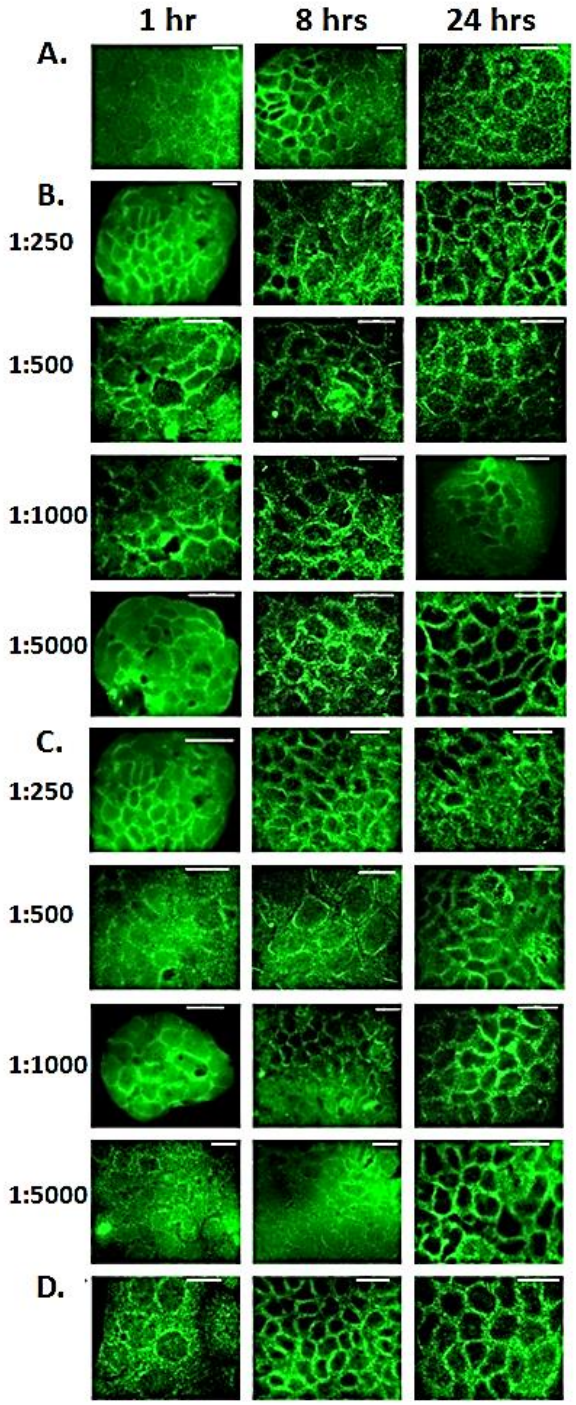


Fig 3. 10: *Giardia* supernatant induces the translocation of intestinal AQP3 after 8 hrs incubation. Caco-2 cells were incubated with different dilutions of *Giardia* GS and WB supernatants for 1, 8 and 24 hrs. Supernatants were collected form *Giardia* culture and diluted at 1:250, 1:500, 1:1,000 and 1:5,000 in non-supplemented DMEM. CaCo-2 cells were fixed with 4% PFA and stained with anti-AQP3 antibody to detect changes in the protein location in intestinal cells. AQP3 was visualised using secondary antibodies conjugated with Alexa fluor 488 (green). **A.** CaCo-2 cells were incubated with non-supplemented DMEM as a control. **B.** Incubation with *Giardia* GS isolate supernatant. **C.** Incubation with *Giardia* WB isolate supernatant. **D.** Incubation with TYI-S-33 at 1:1,000 dilution. Images were obtained using Axiovision software (v 4.7.1), x100 magnitude. Scale bar = 20 µm. Data represents one independent experiment.

incubated in presence or absence of supernatants diluted to 1:250, 1:500, 1:1,000 and 1:5,000 for 1 hour, 8 hours and 24 hours. Following incubation, proteins were extracted and blotted for AQP3. Odyssey machine (Li-Cor, UK) with associated software was utilized to assess AQP3 protein expression (Fig 3.11 B., 3.12 B., 3.13.). Results showed a small but significant decrease in AQP3 expression in treated groups after one hour incubation (Fig 3.11 B.). No significant changes in AQP3 expression were observed after 8 hours incubation (Fig 3.12 B.). Results also showed a slight but significant increase in AQP3 expression for groups treated with WB supernatant diluted to 1:500 and 1:5,000 after 24 hours incubation (Fig 3.13 B.). Microscopic analysis results showed differences in AQP3 distribution throughout intestinal cells between control and groups treated with GS supernatant, after one hour incubation; particularly for groups treated with supernatant higher concentration (Fig 3.11 A.). A similar peripheral distribution was observed after 8 hours incubation with WB supernatant; AQP3 proteins appeared to be more peripheral in groups treated with GS supernatant (Fig 3.12 A.). Results showed a greater peripheral distribution of AQP3 for groups treated with WB supernatant than for the groups treated with GS supernatant, after 24 hours (Fig 3.13 A.). It was also noted that AQP3 distribution was slightly more peripheral in control groups after 24 hours (Fig 3.13 A.).

These results suggest that *Giardia*–derived supernatant may have a weak marginal effect on AQP3 expression. However, a high level of variation in AQP3 expression between biological replicates was observed which suggests that *Giardia*–derived supernatant may not have any effect on AQP3 expression. The microscopic analysis results suggest that *Giardia*–derived supernatant induces the translocation of AQP3 protein, even shortly after inoculation. The fact that AQP3 are more translocated when intestinal cells are in contact with *Giardia* supernatant suggests that AQP3 may be involved in the water efflux disruption which may cause the changes in intestinal cell morphology and the inhibition of chloride secretion previously observed. Yet, the quantification of protein translocation observed in this study could not be assessed via simple immunofluorescence microscopy and has to be investigated.

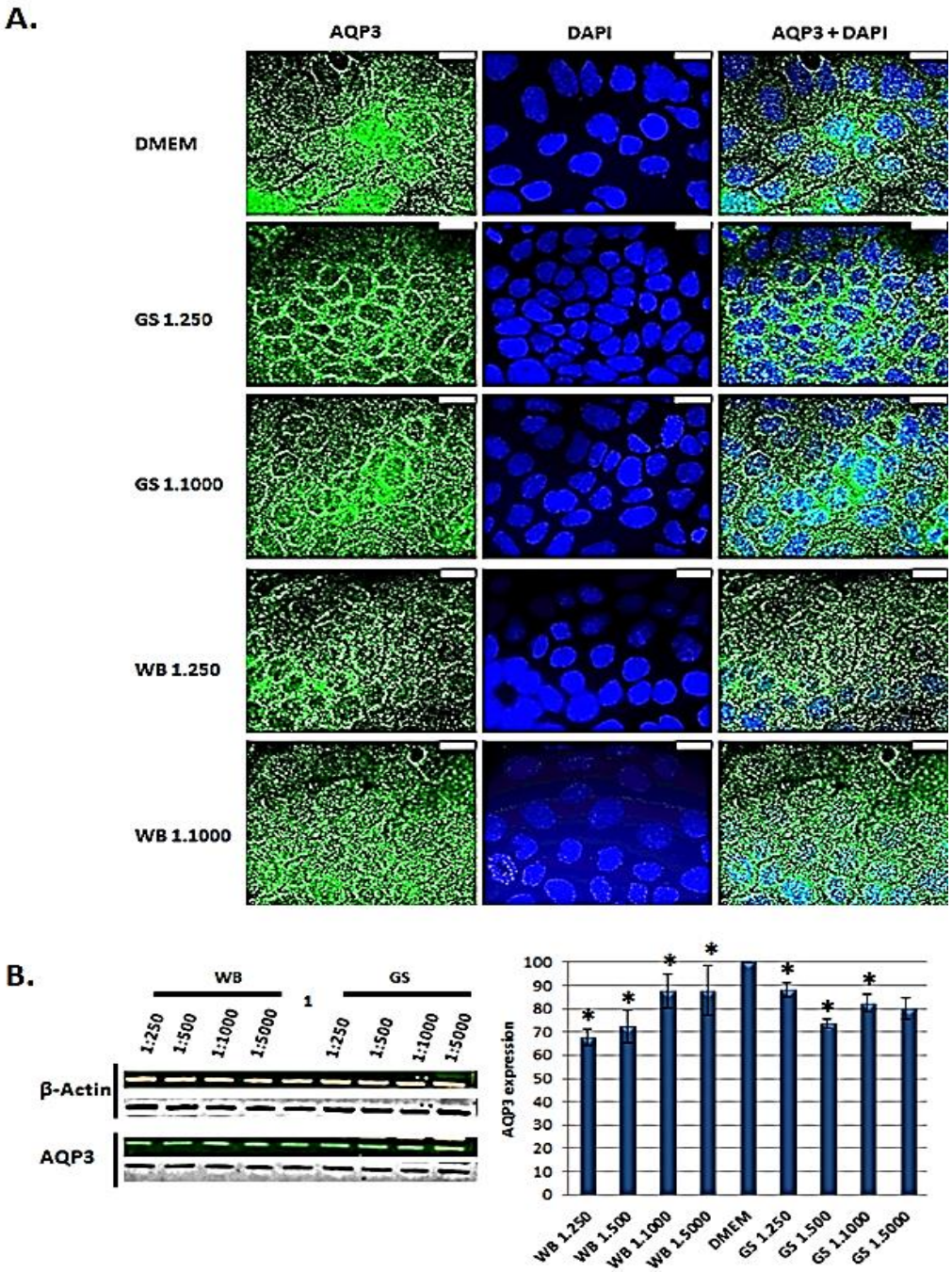


Fig 3. 11: *Giardia* GS isolate (assemblage B) supernatant induces the translocation of intestinal AQP3 after 1 hr incubation. A. CaCo-2 cells were incubated with *Giardia* GS and WB strains supernatants at 1:250 and 1:1,000 for 1 hr. After incubation, CaCo-2 cells were fixed with 4%PFA and stained with anti-AQP3 antibody; the nuclei were stained with DAPI (blue). AQP3 was visualised using secondary antibodies conjugated with Alexa fluor 488 (green). Images were obtained using Axiovision (v 4.8.2), x100 magnitude. Scale bar= 20µm. **B.** CaCo-2 cells were incubated with *Giardia* GS and WB strains supernatant at 1:250, 1:500, 1:100 and 1:5,000 for 1 hr. After incubation, proteins were extracted using M-PER, as stated in Chapter II (section 2.2.2.1). 1: Control incubated with DMEM. AQP3 expression was assessed by western blot analysis and normalised to β-actin using the ODYSSEY software. Results were assessed by non-parametric Mann-Whitney-U tests (compared to control) and expressed as mean values ± SEM, * p-value < 0.05 and ** p-value <0.001 (Appendix V). Data represents three independent experiments with biological replicates.

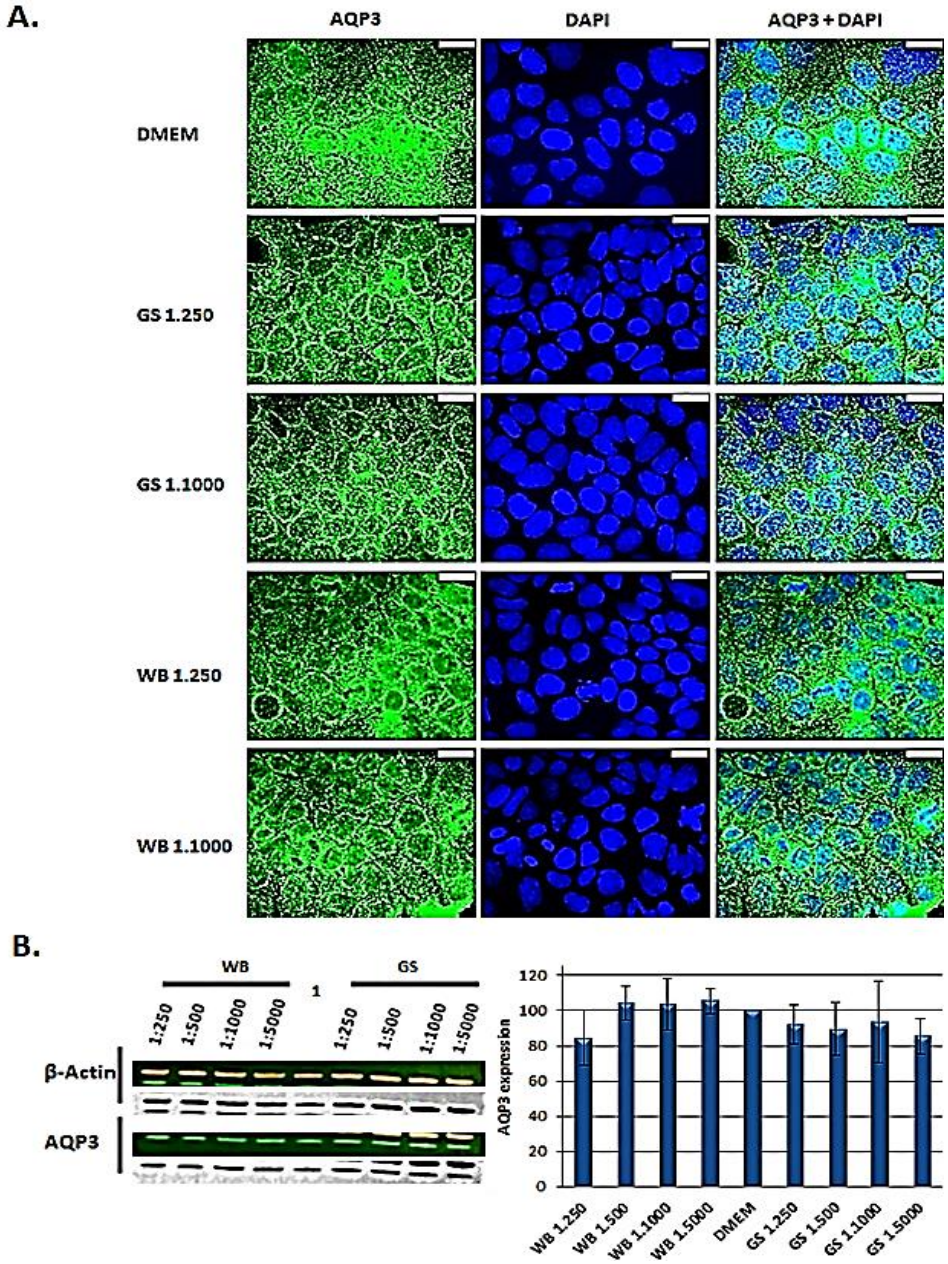


Fig 3. 12: *Giardia* WB and GS isolates (assemblage A and B) supernatants induce the translocation of AQP3 after 8 hrs incubation. A. CaCo-2 cells were incubated with *Giardia* GS and WB strains supernatants at 1:250 and 1:1,000 for 8 hrs. After incubation, CaCo-2 cells were fixed with 4%PFA and stained with anti-AQP3 antibody; the nuclei were stained with DAPI (blue). AQP3 was visualised using secondary antibodies conjugated with Alexa fluor 488 (green). Images were obtained using Axiovision (v 4.8.2), x100 magnitude. Scale bar= 20µm. **B.** CaCo-2 cells were incubated with *Giardia* GS and WB strains supernatant at 1:250, 1:500, 1:100 and 1:5,000 for 8 hrs. After incubation, proteins were extracted using M-PER, as stated in Chapter II (section 2.2.2.1). 1: Control incubated with DMEM. AQP3 expression was assessed by western blot analysis and normalised to β-actin using the ODYSSEY software. Results were assessed by non-parametric Mann-Whitney-U tests (compared to control) and expressed as mean values ± SEM, * p-value < 0.05 and ** p-value <0.001 (Appendix V). Data represents three independent experiments with biological replicates. The duplicate bands represent either the AQP3 below F-actin or F-actin above AQP3. Due to their similar MW, it was challenging to crop images to only see the targeted bands.

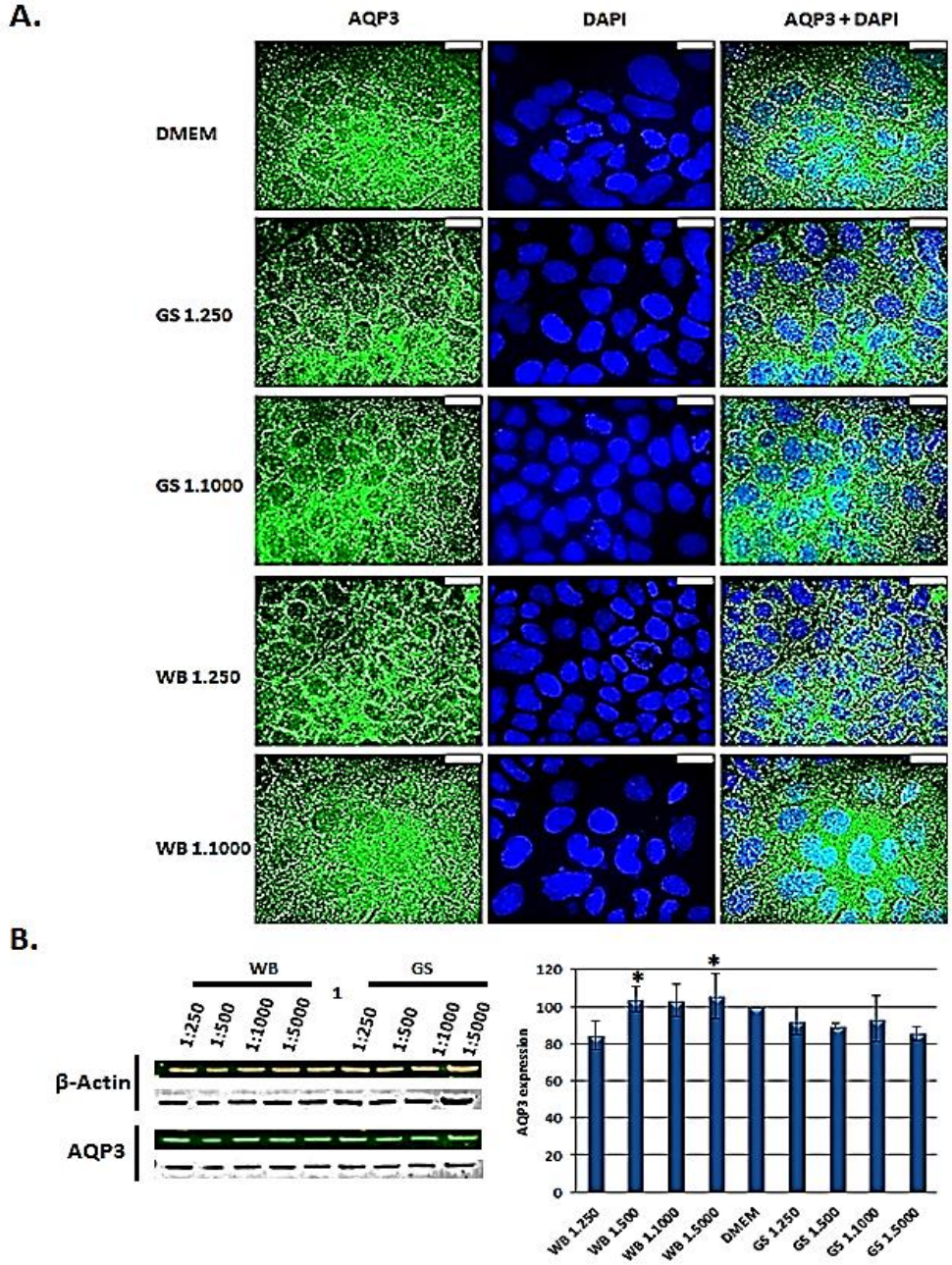


Fig 3. 13: *Giardia* WB and GS isolates (assemblage A and B) supernatants induce the translocation of AQP3 after 24 hrs incubation. A. CaCo-2 cells were incubated with *Giardia* GS and WB strains supernatants at 1:250 and 1:1,000 for 24 hrs. After incubation, CaCo-2 cells were fixed with 4%PFA and stained with anti-AQP3 antibody; the nuclei were stained with DAPI (blue). AQP3 was visualised using secondary antibodies conjugated with Alexa fluor 488 (green). Images were obtained using Axiovision (v 4.8.2), x100 magnitude. Scale bar= 20µm. **B.** CaCo-2 cells were incubated with *Giardia* GS and WB strains supernatant at 1:250, 1:500, 1:100 and 1:5,000 for 24 hrs. After incubation, proteins were extracted using M-PER, as stated in Chapter II (section 2.2.2.1). 1: Control incubated with DMEM. AQP3 expression was assessed by western blot analysis and normalised to β-actin using the ODYSSEY software. Results were assessed by non-parametric Mann-Whitney-U tests (compared to control) and expressed as mean values ± SEM, * p-value < 0.05 and ** p-value <0.001 (Appendix V). Data represents three independent experiments with biological replicates.

3.3.3/ *Giardia* GS supernatant may decrease SGLT-1 expression level

To determine the effect of *Giardia* supernatant on intestinal SGLT-1, expression and translocation, CaCo-2 monolayers were grown either on coverslips or in 12-well plates and incubated with different dilutions of *Giardia*-derived supernatant for 1 hour, 8 hours and 24 hours, as previously described in the AQP3 study (section 3.2.2). Intestinal cells were fixed with 4 % PFA and stained for SGLT-1 or proteins were extracted and blotted for SGLT-1.

Results showed a significant decrease of SGLT1 expression only for groups treated with WB supernatant diluted to 1: 1,1000 and GS supernatant diluted to 1:250 (Fig 3.14 B.). After 8 hours, results showed a significant increase of SGTL1 expression for groups treated with WB supernatant diluted to 1:1,000; and a significant decrease of SGLT1 expression for groups treated with GS supernatant diluted to 1:250, 1:500 and 1:1,1000 (Fig 3.15 B.). After 24 hours, results showed a similar significant increase in SGLT1 expression for groups also treated with WB supernatant diluted to 1:1,000; and a significant decrease of SGLT1 expression for groups treated with GS supernatant diluted to 1:250 (Fig 3.16 B.). It was noted that, after 8 and 24 hours, groups treated with WB supernatant diluted to 1:5,000 showed a SGLT1 expression profile similar to the control group however, the statistical analysis showed no significant differences between the control and these treated groups. Results of the microscopic analysis showed that SGLT1 proteins of treated groups had a distribution similar to control groups, under all conditions and at all time (Fig 3.14 A., 3.15 A., 3.16 A.).

These results suggest that a high concentration of *Giardia* GS supernatant might induce a slight reduction of SGLT-1 expression. However, this reduction of expression appears to be less than 40 % which indicates that SGLT-1 expression may not be dramatically affected by *Giardia* supernatant. Moreover, like for the AQP3 data previously discussed, the important variation in SGLT-1 expression between biological replicates suggests that, the whole effect of GS supernatant on SGLT-1 expression might have not been fully observed and may be stronger than these data suggest. These results also suggest that WB supernatant lower

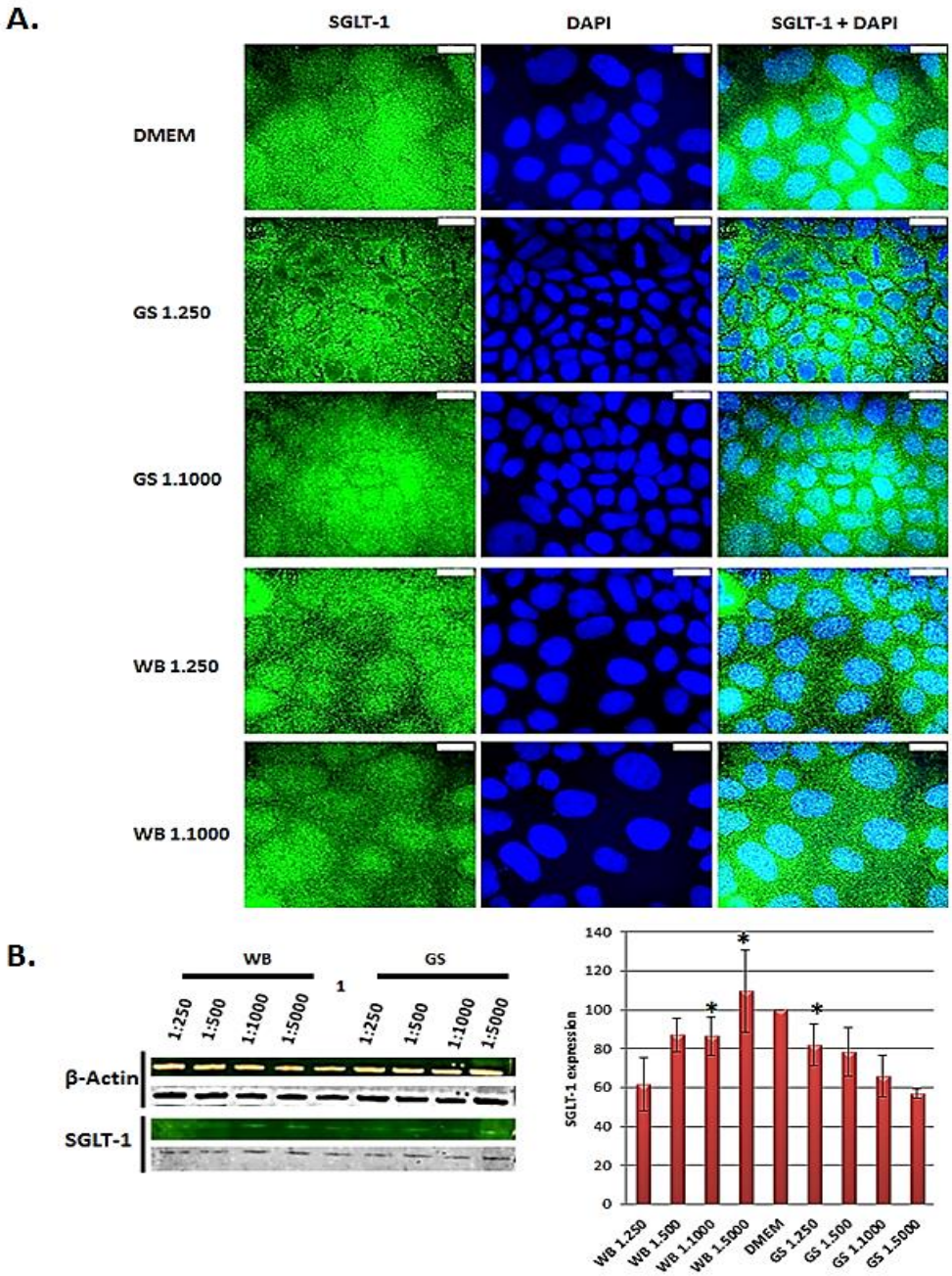


Fig 3. 14: *Giardia* WB and GS isolates (assemblage A and B) supernatants have no effect on the translocation and expression of SGLT-1 after 1 hr incubation.. A. CaCo-2 cells were incubated with *Giardia* GS and WB strains supernatants at 1:250 and 1:1,000 for 1 hr. After incubation, CaCo-2 cells were fixed with 4%PFA and stained with anti-SGLT-1 antibody; the nuclei were stained with DAPI (blue). SGLT-1 was visualised using secondary antibodies conjugated with Alexa fluor 488 (green). Images were obtained using Axiovision (v 4.8.2), x100 magnitude. Scale bar= 20µm. **B.** CaCo-2 cells were incubated with *Giardia* GS and WB strains supernatant at 1:250, 1:500, 1:100 and 1:5,000 for 1 hr. After incubation, proteins were extracted using M-PER, as stated in Chapter II (section 2.2.2.1). 1: Control incubated with DMEM. SGLT-1 expression was assessed by western blot analysis and normalised to β-actin using the ODYSSEY software. Results were assessed by non-parametric Mann-Whitney-U tests (compared to control) and expressed as mean values ± SEM, * p-value < 0.05 and ** p-value <0.001 (Appendix V). Data represents three independent experiments with biological replicates.

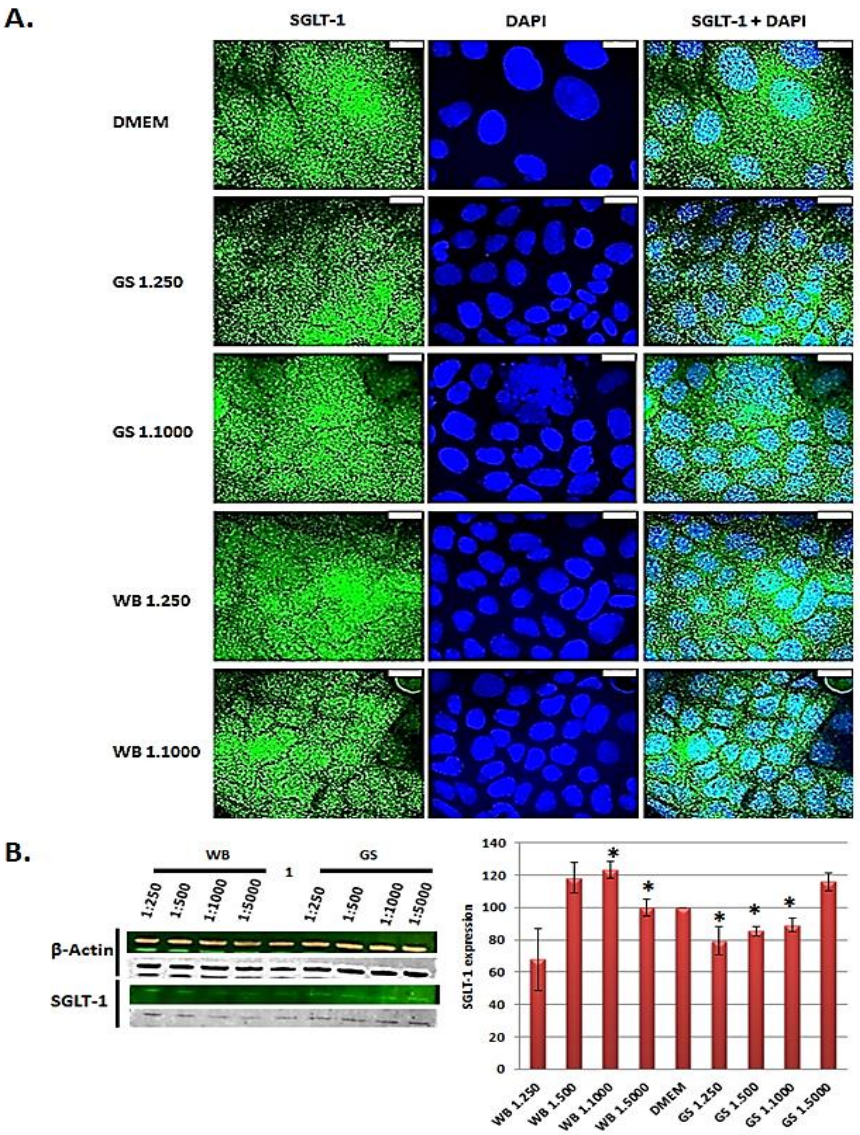


Fig 3. 15: *Giardia* WB and GS isolates (assemblage A and B) supernatants have no effect on the translocation and expression of SGLT-1 after 8 hrs incubation. A. CaCo-2 cells were incubated with *Giardia* GS and WB strains supernatants at 1:250 and 1:1,000 for 8 hrs. After incubation, CaCo-2 cells were fixed with 4%PFA and stained with anti-SGLT-1 antibody; the nuclei were stained with DAPI (blue). SGLT-1 was visualised using secondary antibodies conjugated with Alexa fluor 488 (green). Images were obtained using Axiovision (v 4.8.2), x100 magnitude. Scale bar= 20µm. **B.** CaCo-2 cells were incubated with *Giardia* GS and WB strains supernatant at 1:250, 1:500, 1:100 and 1:5,000 for 8 hrs. After incubation, proteins were extracted using M-PER, as stated in Chapter II (section 2.2.2.1). 1: Control incubated with DMEM. SGLT-1 expression was assessed by western blot analysis and normalised to β-actin using the ODYSSEY software. For WB/β-actin, the extra represent the SGLT-1 bands which could not be removed on image due to the oblique migration of proteins. Results were assessed by non-parametric Mann-Whitney-U tests (compared to control) and expressed as mean values ± SEM, * p-value < 0.05 and ** p-value <0.001 (Appendix V). Data represents three independent experiments with biological replicates. The duplicate bands represent either the SGLT-1 below F-actin or F-actin above SGLT-1. Due to their similar MW, it was challenging to crop images to only see the targeted bands.

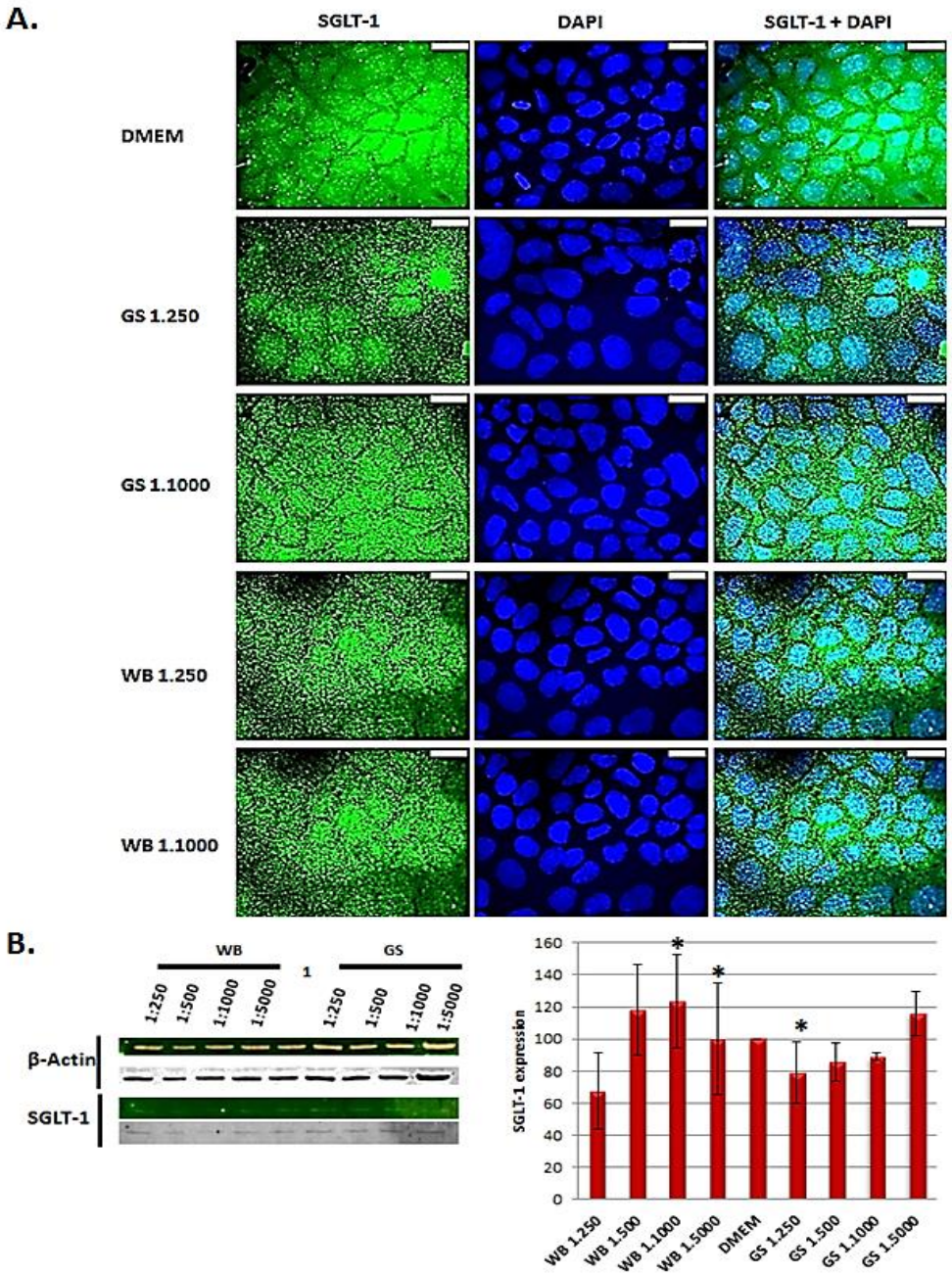


Fig 3. 16: *Giardia* WB and GS isolates (assemblage A and B) supernatants have no effect on the translocation and expression of SGLT-1 after 24 hrs incubation. A. CaCo-2 cells were incubated with *Giardia* GS and WB strains supernatants at 1:250 and 1:1,000 for 24 hrs. After incubation, CaCo-2 cells were fixed with 4%PFA and stained with anti-SGLT-1 antibody; the nuclei were stained with DAPI (blue). SGLT-1 was visualised using secondary antibodies conjugated with Alexa fluor 488 (green). Images were obtained using Axiovision (v 4.8.2), x100 magnitude. Scale bar= 20µm. **B.** CaCo-2 cells were incubated with *Giardia* GS and WB strains supernatant at 1:250, 1:500, 1:100 and 1:5,000 for 24 hrs. After incubation, proteins were extracted using M-PER, as stated in Chapter II (section 2.2.2.1). 1: Control incubated with DMEM. SGLT-1 expression was assessed by western blot analysis and normalised to β-actin using the ODYSSEY software. Results were assessed by non-parametric Mann-Whitney-U tests (compared to control) and expressed as mean values ± SEM, * p-value < 0.05 and ** p-value <0.001 (Appendix V). Data represents three independent experiments with biological replicates.

concentration may increase SGLT1 expression after 8 hours. However, due the great variation between biological replicates and the inconsistency of results between the different treated groups, this cannot be confirm nor deny. The microscopic analysis results suggest that *Giardia* supernatant does not induce SGLT1 translocation.

3.3.4/ *Giardia* supernatant induces CFTR translocation in intestinal cells in early stages of infection.

To determine the effect of *Giardia* supernatant on intestinal CFTR expression and translocation, CaCo-2 monolayers were grown either on coverslips or in 12-well plates and incubated with different dilutions of *Giardia*-derived supernatant for 1 hour, 8 hours and 24 hours, as previously described in the AQP3 and SGLT1 studies (section 3.2.2 and 3.2.3.). Intestinal cells were fixed with 4 % PFA and stained for CFTR or proteins were extracted and blotted for CFTR (Fig 3.17-3.19).

Western blots were performed on extracted proteins to quantify CFTR expression levels under different concentrations of *Giardia* supernatant at different times. Proteins were blotted with three different monoclonal mouse anti-CFTR antibodies from three different companies (AbCam, R&D systems, and Santa-Cruz Biotechnologies) as stated in chapter II (Table 2.3). All antibodies were tested at concentrations ranging from 1:500 to 1:250 and left overnight at 4°C, 0.2 % tween was added to improve the bond between proteins and antibodies; but no bands were observed for CFTR. Following these conditions, no bands were ever observed for CFTR apart from one western blot with AbCam antibody (Appendix VI). The gel showed either poor quality bands or no bands for CFTR. Due to this negative blotting on CFTR, the effect of *Giardia* supernatant on CFTR expression could not be assessed. Many different concentrations (up to 1:250) were tried but no reproducible data were obtained (Appendix VI). Even though the western blot assay was non-reproducible, this was considered not to affect the immunofluorescence assay; this was shown by the similarities in microscopic images observed after each experiment.

Results showed that CFTR distribution was more peripheral in groups treated with GS supernatant than control group after one hour; particularly groups treated with higher concentration (Fig 3.17). After 8 hours, a greater peripheral distribution of CFTR was observed for groups treated with *Giardia* supernatant (Fig 3.18). After 24 hours, groups treated with GS supernatant appeared to have a similar CFTR distribution to the control group, and groups treated with WB supernatant appeared to have a CFTR distribution more peripheral (Fig 3.19).

These results suggest that *Giardia* supernatant induces the translocation of CFTR towards the periphery of the intestinal cells. *Giardia* assemblage B (GS strain) appears to have an effect on CFTR distribution during the early stage of infection whereas assemblage A (WB strain) induces an effect sometimes between 1 hour and 8 hours (Fig 3.17-3.19). This suggests that *Giardia* assemblage A (WB strain) effect on intestinal CFTR may last longer than the effect of assemblage B (GS strain). These results also strongly suggest that CFTR may be the intestinal channel by which *Giardia* supernatant inhibits the intestinal chloride secretion. Even if CFTR translocation was observed in this study, its quantification could not be assessed via simple immunofluorescence microscopy and has to be investigated.

3.4/ Discussion

3.4.1/ Effect on intestinal cell morphology

Giardia has been shown to disrupt several intestinal cell functions such as the intestinal barrier, the anion absorption or apoptosis (Cotton et al., 2011, Troeger et al., 2007, Buret, 2007, Scott et al., 2002). However, no studies have yet demonstrated the effect of *Giardia* supernatant on the intestinal cells morphology. In this study, *Giardia* supernatant was shown to provoke a change in intestinal cell morphology and size; cells appeared shrunk and ruffled after exposure to *Giardia* supernatant (Fig 3.4, 3.6-3.11). It was previously shown that an inhibition of chloride secretion lead to a decrease of cell volume (Walters et al., 1992). This suggests that the changes in intestinal cell morphology after exposure to *Giardia* supernatant correlate with the changes in electrophysiology observed in our laboratory (Chapter I, section 1.5.4., Fig 1.17-1.19).

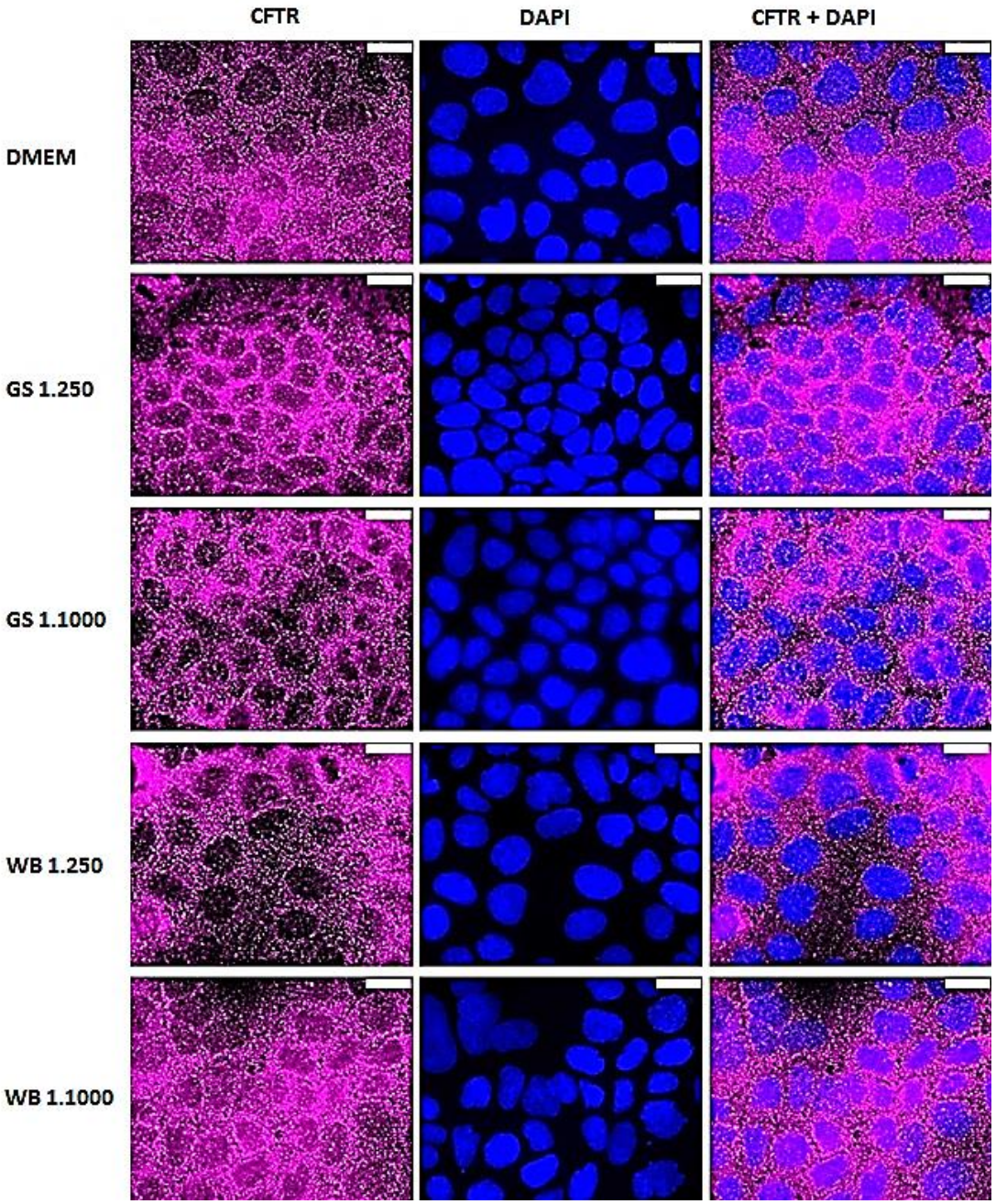


Fig 3. 17: *Giardia* GS isolate (assemblage B) supernatant induces the translocation of intestinal CFTR after 1 hr incubation. CaCo-2 cells were incubated with *Giardia* GS and WB strains supernatants at 1:250 and 1:1,000 for 1 hr. After incubation, CaCo-2 cells were fixed with 4%PFA and stained with anti-CFTR antibody; the nuclei were stained with DAPI (blue). CFTR was visualised using secondary antibodies conjugated with Alexa fluor 594 (red). Images were obtained using Axiovision (v 4.8.2), x100 magnitude. Scale bar= 20µm. Data represents three independent experiments with biological replicates.

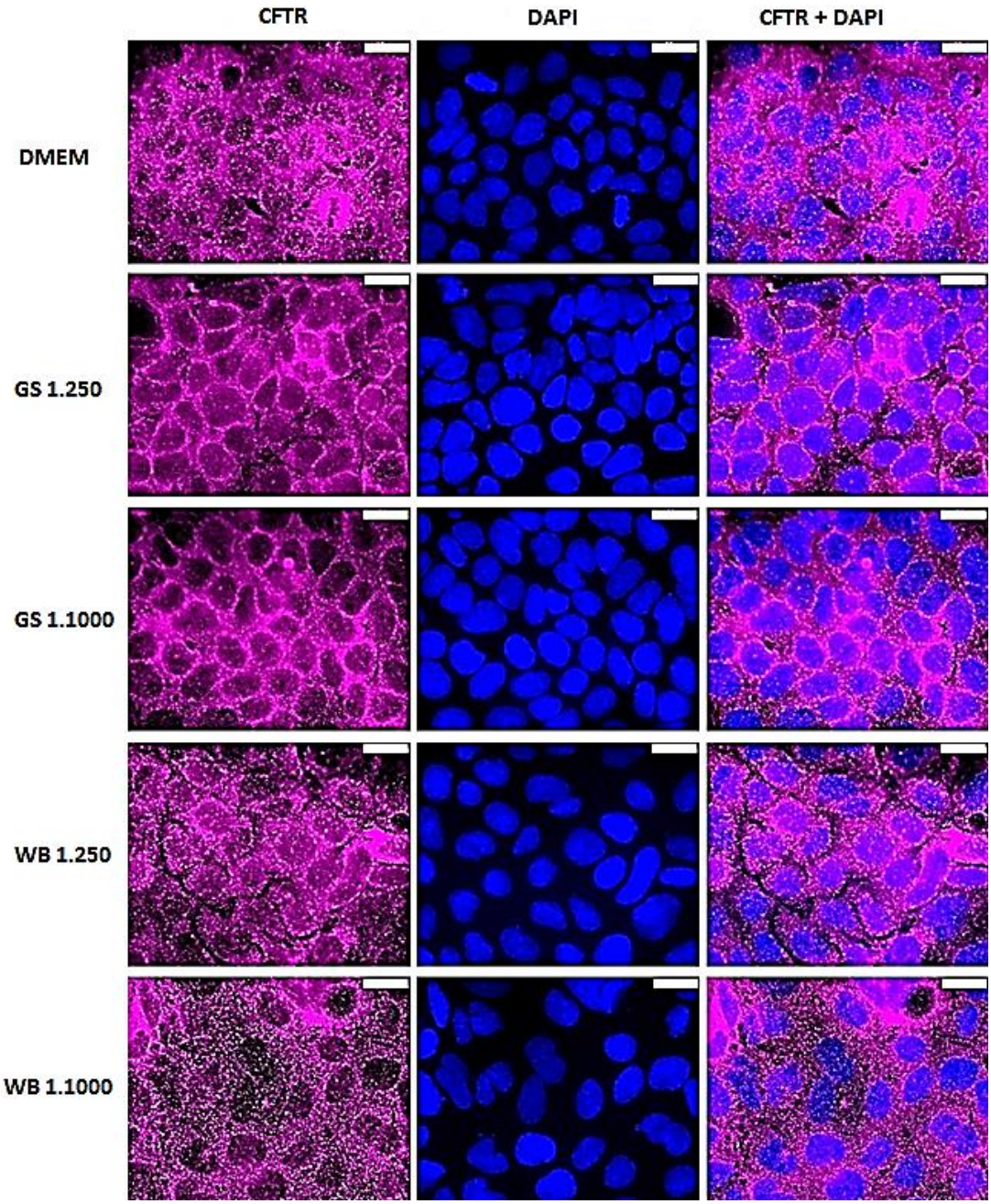


Fig 3. 18: *Giardia* WB and GS isolates (assemblage A and B) supernatants induce the translocation of CFTR after 8 hrs incubation. CaCo-2 cells were incubated with *Giardia* GS and WB strains supernatants at 1:250 and 1:1,000 for 8 hrs. After incubation, CaCo-2 cells were fixed with 4%PFA and stained with anti-CFTR antibody; the nuclei were stained with DAPI (blue). CFTR was visualised using secondary antibodies conjugated with Alexa fluor 594 (red). Images were obtained using Axiovision (v 4.8.2), x100 magnitude. Scale bar= 20µm. Data represents three independent experiments with biological replicates.

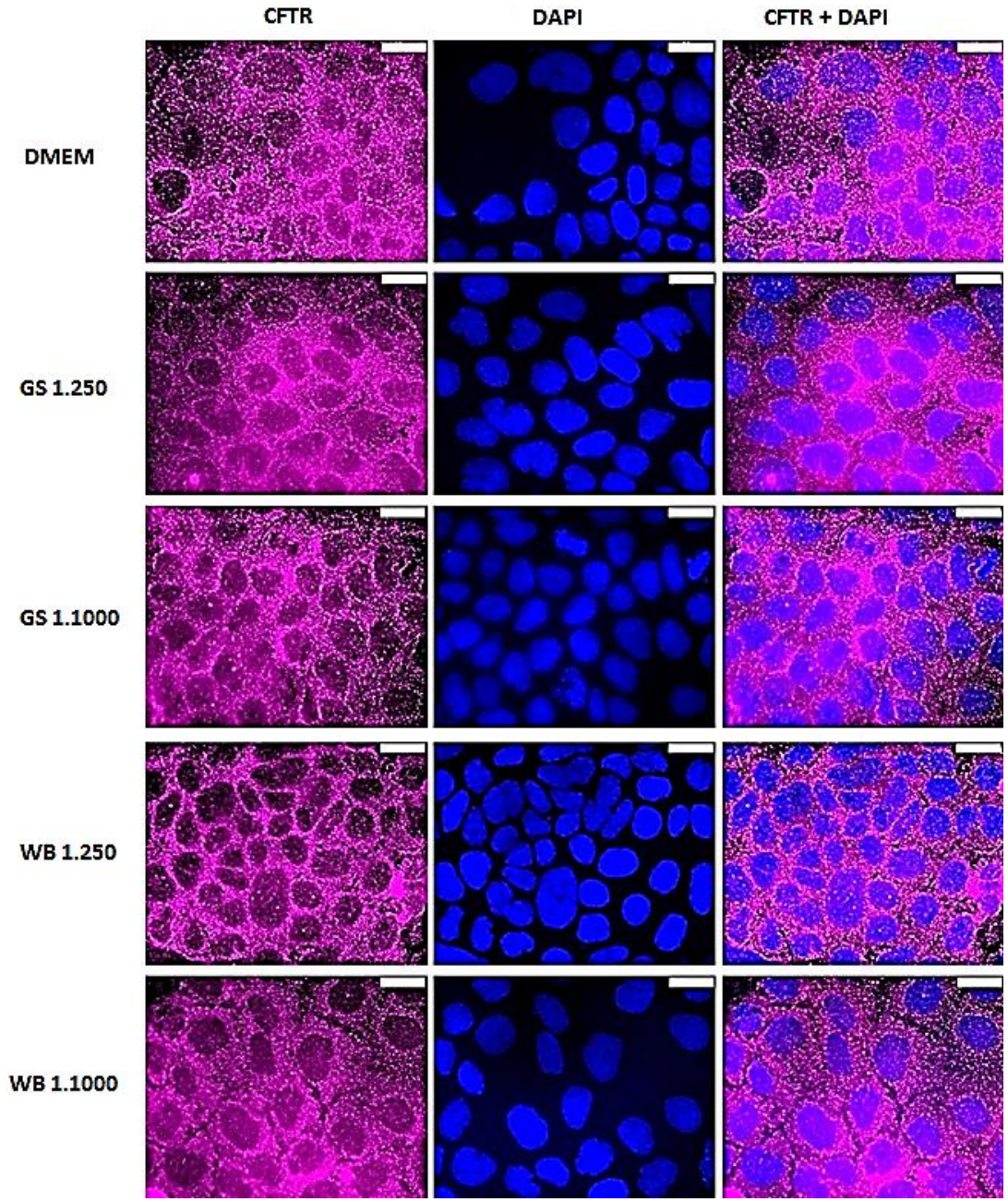


Fig 3. 19: *Giardia* WB and GS isolates (assemblage A and B) supernatants induce the translocation of CFTR after 24 hrs incubation. CaCo-2 cells were incubated with *Giardia* GS and WB strains supernatants at 1:250 and 1:1,000 for 24 hrs. After incubation, CaCo-2 cells were fixed with 4%PFA and stained with anti-CFTR antibody; the nuclei were stained with DAPI (blue). CFTR was visualised using secondary antibodies conjugated with Alexa fluor 594 (red). Images were obtained using Axiovision (v 4.8.2), x100 magnitude. Scale bar= 20µm. Data represents three independent experiments with biological replicates.

In this study, intestinal cells were first incubated with *Giardia* supernatant from culture for 24 hours. The cells length and width were measured to preliminary assess the effect of *Giardia* supernatant on the intestinal cell morphology. This pilot study showed apparent cell shrinkage when intestinal cells were incubated with *Giardia* culture supernatants (Fig 3.4). However, some effects from *Giardia* growth media were also observed (Fig 3.4 B.). As a growth media, TYI-S-33 components are present in supernatant in higher quantity than *Giardia* secreted component within culture tubes. This suggests that within diluted *Giardia* supernatant, more TYI-S-33 components were present than *Giardia* components. The fact that intestinal cells incubated with TYI-S-33 diluted at 1:1,000 were similar in size to intestinal cells under normal condition of growth showed that TYI-S-33 does not affect dramatically the intestinal cell size. However, intestinal cells under GS 1:5,000 treatment had a significant reduction in size than cells incubated with TYI-S-33 while intestinal cells incubated with GS 1:250 were not which indicates that TYI-S-33 does have some marginal effects on the intestinal cells morphology (Fig 3.4 B.).

The study was improved by removing the influence of *Giardia* growth media and by delineating cell margins (Fig 3.6-3.8). These results show a marginal effect of *Giardia* supernatant on cell morphology suggesting that TYI-S-33 contributed significantly to the initial observations. Nevertheless, small reduction in cell size and turgidity from 8 hours incubation with *Giardia* products is observed (Fig 3.6-3.9). Indeed, *Giardia* supernatant was shown to have a slight effect on the intestinal cells morphology and water flux, especially after 8 hours incubation with the parasite supernatant. *Giardia* assemblage B (GS strain) appears to have a more important effect on the intestinal cells surface area and perimeter than assemblage A (WB strain) which suggests that *Giardia* assemblage B may have a greater inhibition effect on the intestinal water uptake (Fig 3.7 and 3.8). However, intestinal cells appear to be more ruffled when incubated with *Giardia* assemblage A which indicates that assemblage A may cause more mucosal damage than assemblage B (Fig 3.9).

Studies on gerbils and rats showed that the diffuse loss of epithelial brush border and decrease of water absorption during giardiasis were *Giardia* strain-dependent (Benere et al., 2012, Cevallos et al., 1995). These studies also

demonstrated that assemblage B had a greater effect on brush-border enzymes deficiencies; and that these brush-border enzymes deficiencies were caused by a diffuse loss of epithelial brush-border microvillus surface area rather than alteration in villus structure (Benere et al., 2012). The fact that CaCo-2 cells surface area and perimeter were found to be more decreased when intestinal cells were incubated with *Giardia* GS supernatant correlates with these previous studies realised on gerbils and rats. Benere et al also suggested that assemblage A trophozoites can induce a high degree of mucosal and functional damages without causing obvious signs of diarrhoea (Benere et al., 2012). Here, assemblage A was shown to have a lesser effect on intestinal cell size suggesting that the intestinal water uptake may not be as affected by assemblage A supernatant as it is by assemblage B supernatant. Moreover, assemblage A supernatant causes a more important ruffling effect of intestinal cells than assemblage B which also correlates with the previous studies realised on gerbils and rats.

The immunostaining of ZO-1 allowed calculating the surface area and perimeter changes occurring in intestinal cells during infection with *Giardia* supernatant, however, the water loss could not be accurately assessed via microscopy. The realisation of an electrophysiology assay using Ussing chamber may allow assessing the actual loss of water in real time. It is important to note that the measurement of surface area and perimeter were done via Image J software by redrawing the outline of non-mitotic and non-apoptotic cells in order for the software to calculate these two data. This is not the most accurate method and it might have led to a slight loss of information. Due to this lack of accuracy in the measurement of surface area and perimeter, the ruffling index may not be accurate either suggesting that this effect might be more pronounced too. This could explain the non-significant data for GS supernatant when some ruffling of Caco-2 was microscopically observed. However, even with a slight loss of information, this study shows that *Giardia* supernatant induces significant but not dramatic morphological changes in intestinal cells.

3.4.2/ Effect on intestinal cell water and ions transporters

Some of the effects of *Giardia* on intestinal cells have long been well-established (Cotton et al., 2011, Roxstrom-Lindquist et al., 2005, Scott et al., 2000). Yet, the possible involvement of both *Giardia*-secreted factors and host cell transport mechanisms remains unknown. In this study, the implications of two main water channels, AQP3 and SGLT-1, and of one chloride secretion channel, CFTR were investigated (Fig 3.1-3.3).

Aquaporins are responsible for the water uptake by osmosis in many cells; but they can also transport small molecules through the lipid bilayer, AQP3 is also involved in the transport of glycerol, urea and other small solutes such as ammonia/ammonium (Krane and Goldstein, 2007, Holm et al., 2005). The pilot study shows apparent translocation of AQP3 protein under exposure to *Giardia*. However, *Giardia* growth media induces an even more pronounced effect (Fig 3.10).

Study was improved by removing the influence of *Giardia* growth media and by using supplemented DMEM for the incubation with intestinal cells as non-supplemented DMEM was initially used during incubation with *Giardia* supernatants (Fig 3.10-3.13). This study shows that *Giardia* supernatant may not have any effect on AQP3 expression due to the marginal effect observed. It also shows that *Giardia* supernatant induces the translocation of AQP3 protein. *Giardia* assemblage B induces the translocation of AQP3 shortly after inoculation as proteins appeared to be peripheral after one hour incubation (Fig 3.11 A., 3.12 A., 3.13 A.). *Giardia* assemblage A supernatant induces both AQP3 translocation between 1 and 8 hours incubation; and its effect appeared to last longer as more proteins were peripheral after 24 hours incubation with WB supernatant. This translocation appeared to be a dose-response effect as more proteins were observed to be peripheral with a 1:250 dilution for the parasite supernatants. However, the quantification of this translocation was not possible using an Axioplan 2 imaging microscope. The investigation of the density of vesicles in close proximity of the plasma membrane via the use of a confocal XZ-plane imaging microscope and a peripheral stain as a marker of cells shape, such as a tight junction protein,

should allow the quantification of the translocation observed (Domanova et al., 2011, Yu et al., 2008). The quantitative measurement of AQP3 translocation could also be assessed by flow cytometry as suggested by Koshy and collaborators in their study of GLUT4 translocation to the plasma membrane (Koshy et al., 2010).

SGLT-1 is responsible for the active transport of glucose across the brush border membrane of small intestine, via a coupled transport of two sodium ions for one glucose molecule (Wright et al., 2007); it can also be involved in the osmotic water uptake (Zeuthen, 2010). Like for the study on AQP3 proteins, CaCo-2 cells were incubated with both *Giardia* GS and WB supernatants at different dilutions, for 1, 8 and 24 hours; immunofluorescence microscopy assay and western blots were performed (Fig 3.14-3.16). This study shows that *Giardia* supernatant has a slight marginal effect on SGLT-1 expression suggesting that *Giardia* supernatant may not have any effect on this protein expression. This study also shows that *Giardia* supernatant has no effect on SGLT-1 translocation (Fig 3.14 A., 3.17 A., 3.18 A.). However, some studies described biological process in which SGLT-1 activation by *Giardia* may rescue enterocytes from lipopolysaccharide (LPS)-induced epithelial cell apoptosis by enhancing glucose uptake (Buret, 2008, Yu et al., 2008, Yu et al., 2006, Yu et al., 2005). Indeed, Yu et al showed that in a high glucose environment, *Giardia* induces the apical translocation of SGLT-1 transporters which leads to the inhibition of *Giardia*-induced apoptosis (Yu et al., 2008, Yu et al., 2006, Yu et al., 2005). Yu et al also suggested that a soluble heat-labile protein fraction in *Giardia* trophozoites may trigger the SGLT-1-mediated host response observed in their study (Yu et al., 2008). In this study, the lack of translocation observed could be explained by an environment low in glucose. Moreover, in their study, Yu and collaborators incubated *Giardia* trophozoites either in a low glucose concentration environment or in a high glucose concentration environment to specifically investigate the effect on SGLT-1 transporter which was not the aim of this study. Yet, it may be interested to repeat the assay on SGLT-1 in a high glucose concentration environment to investigate further the findings of Yu et al.

CFTR is a cAMP-activated channel responsible for the chloride secretion via an osmotic gradient (Murek et al., 2010). The effect of *Giardia* supernatant on CFTR expression and translocation was investigated following the same protocol as for

the study of AQP3 and SGLT-1. The influence of *Giardia* supernatant on CFTR expression could not be assessed due to the lack of reproducibility of the western blot assays for this protein (Appendix VI). CFTR is a heavy protein of approximately 150 kDa, the use of pre-cast gels with a lower gradient than the ones used during this study may allow a better reproducibility of the assay. CFTR proteins show a similar translocation effect to AQP3 when incubated with *Giardia* supernatants with an early and more pronounced translocation effect induced by *Giardia* assemblage B (Fig 3.17-3.19). This translocation also appears to be dose-dependent as more CFTR proteins were observed at the periphery of intestinal cells with a 1:250 dilution for *Giardia* supernatant. As for AQP3, the quantification of this translocation could not be evaluated and should be investigated either via confocal microscopy or flow cytometry as previously stated.

These studies suggest that *Giardia* supernatant induces the translocation of both AQP3 and CFTR which are involved in intestinal water uptake and chloride secretion respectively (Zeuthen, 2010, Murek et al., 2010). This suggests that these two intestinal proteins may be involved in the changes in cell morphology and turgidity, and chloride secretion inhibition caused by *Giardia* supernatant.

3.5/ Summary

These studies show that *Giardia* trophozoites secrete factors affecting the host cells morphology and cell trafficking mechanisms. *Giardia* supernatant induces a disruption of the intestinal water uptake suggesting an implication of the intestinal water channels in this biological process. The translocation of AQP3 observed indicates that AQP3 transporter may be involved in the decrease of cell turgidity observed, as AQP3 translocation occurs after one hour incubation and the first effects on morphology were observed after 8 hours incubation (Fig 3.4-3.9, 3.11-3.13). The fact that GS supernatant appears to have a greater effect than WB supernatant on both intestinal cell size and AQP3 translocation suggests that AQP3 may be involved in the loss of water occurring under treatment with assemblage B supernatant. *Giardia* assemblage A appeared to have a milder effect on both intestinal cell size and AQP3 translocation which suggests that assemblage A may have a lesser effect on intestinal water uptake.

A previous study on chloride secretion inhibition by *Giardia* supernatant realised in our laboratory suggests that CFTR might not be involved in the morphological changes as it was inhibited by *Giardia* supernatant (Chapter I, section 1.5.4, Fig 1.17-1.19). However, *Giardia* supernatant was shown to induce CFTR translocation after 1 hour which suggests that CFTR may be involved in some of the host cell disruptions caused by *Giardia*. In which of the different cellular disruptions caused by *Giardia* CFTR might be involved has yet to be investigated.

These studies show that *Giardia* does secrete soluble factors inducing some disruptions in host cells via their interaction with intestinal transporters. Whether these soluble mediators are proteins or other solutes remains unanswered. Reproducing all the different assays performed in this study with *Giardia* supernatant boiled prior to incubation would be an easy way to determine if these effects are due to *Giardia*-released proteins or not.

Chapter IV

Pilot study and protocol optimisation to investigate *Giardia* secretion profile.

4.1/ Introduction

As stated in Chapter I (section 1.5.2.3.), *Giardia* supernatant was shown to reduce intestinal chloride efflux and to inhibit CaCo-2 cells Isc suggesting that *Giardia* secretes mediators acting on intestinal cells functions (Al-Naimi et al, unpublished data). However, what kind of *Giardia*-secreted molecules are involved remains unknown and requires further investigation. Proteins are the most abundant cellular macromolecules and are involved in diverse cellular functions such as enzymatic reaction, transport of molecules across membranes or the conservation of the cellular cytoskeletal structure (Lodish et al., 2000). This suggests that the *Giardia*-secreted mediators involved in the cellular Isc inhibition may be proteins. Besides, only a few proteins were shown to be secreted or released by trophozoites upon interaction with intestinal cells; and a protein secretion profile for *Giardia* has yet to be determined (Cotton et al., 2014, Ringqvist et al., 2011, Ringqvist et al., 2008, Rodriguez-Fuentes et al., 2006).

As stated in Chapter II and III, the *in vitro* growth of *Giardia* trophozoites requires the use of numerous components such as BSA, pancreatic and bile digests, B12 vitamin, and yeast extract (Chapter II, Table 2.1). Therefore, TYI-S-33 proteins may interfere in the investigation of *Giardia*-secreted proteins from trophozoites cultured *in vitro*. To verify this hypothesis, supernatants form *Giardia* culture were collected for both assemblage A and B and a SDS-PAGE assay was performed to compare protein profile for *Giardia* supernatant to TYI-S-33 (Fig 4.1). Both assemblage supernatant showed a similar secretion profile to TYI-S-33; a similar profile was also observed with a 1:10 dilution (Fig 4.1 B.). This indicates that TYI-S-33 is too enriched in *Giardia* supernatant to use it directly from trophozoites *in vitro* culture.

The aims of this study were:

(i) To develop a method to obtain a supernatant as clear as possible from TYI-S-33

(ii) To optimise this method prior to proteomics analysis of *Giardia* samples.

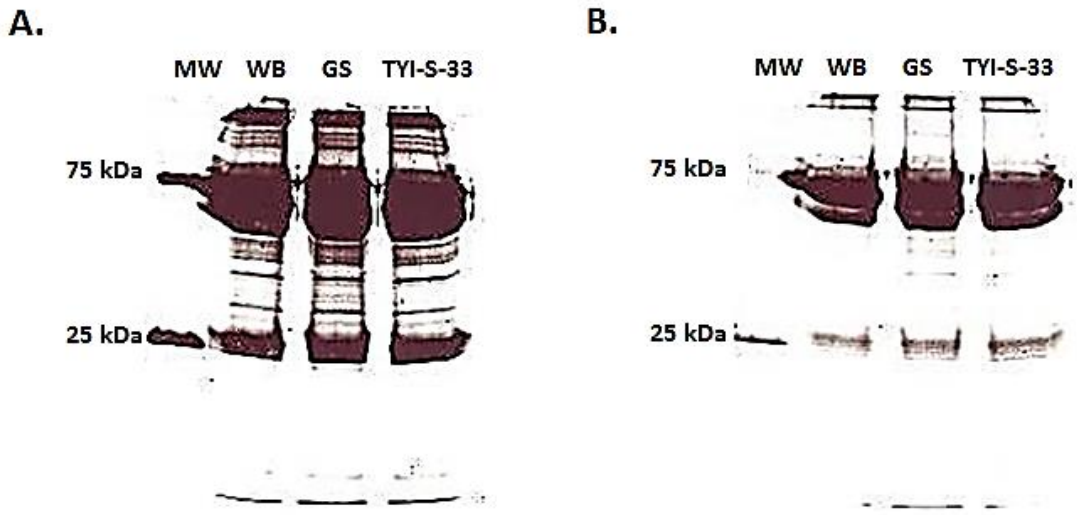


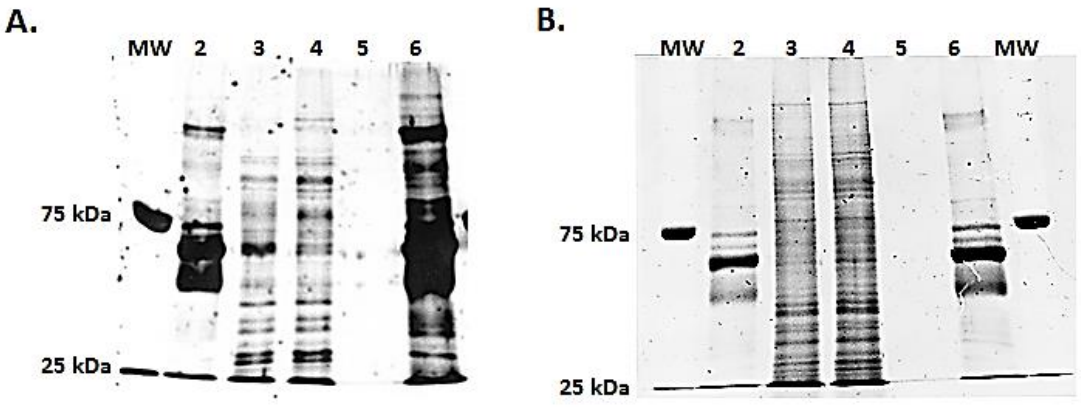
Fig 4. 1: *Giardia* supernatant from trophozoites culture is too enriched in TYI-S-33 proteins. Parasites were chilled on ice for 15 min, and centrifuged 10 min at 3,000 rpm. Supernatants were collected and SDS-PAGE on 12% gels, using a SYPRO staining, were realised. **A.** Protein profile of undiluted supernatant **B.** Protein profile in 1:10 dilution. Data represents two independent experiments
MW: Molecular weight.

4.2/ Pilot study to determine *Giardia* secretion profile

4.2.1/ *Giardia* secretes proteins

In order to obtain *Giardia* supernatants as clean as possible from TYI-S33 proteins (Fig 4.1), *Giardia* trophozoites were collected from culture tube and washed three times with PBS. At first, trophozoites were incubated in PBS for approximately 2 hours after PBS washes. However, only a few protein bands were observed on SDS-PAGE gel, which were also present in the TYI-S-33 control (data not shown). This suggested that *Giardia* trophozoites were not releasing any proteins in the milieu when incubated in PBS. PBS is not a growth medium; therefore no nutrients were available for the trophozoites in this environment which may explain the lack of *Giardia* proteins in the supernatants. To counteract this effect, *Giardia* was incubated in a growth medium different from TYI-S-33 with no bovine or calf proteins. DMEM, a standard *in vitro* growth medium used to culture several cell lines, was used. DMEM was not supplemented with BSA nor FCS, only antibiotics were added to keep an axenic milieu. Parasites were incubated for up to 60 minutes with DMEM. Parasites were then collected and centrifuged to collect supernatant from DMEM incubation; BCA protein assay and SDS-PAGE were performed to verify the protein profile (Fig 4.2). A maximum of 45 minutes incubation was shown to give optimum protein concentration and profile on SDS-PAGE gels as less debris were observed in the culture tubes (data not shown). Protein concentrations ranged between 0.655 and 1.339 µg/ml which confirmed that proteins were present in supernatants (Fig 4.2). Some TYI-S-33 proteins appeared to be still present in *Giardia* supernatant; however, *Giardia*-specific protein bands were also present on SDS-PAGE gel (Fig 4.2). Moreover, no bands were observed for non-supplemented DMEM which confirmed that trophozoites were incubated in a protein-free medium (Fig 4.2).

To obtain exploitable supernatant for the study of *Giardia* assemblage A and B secretion profiles, *Giardia* trophozoites had to be incubated in different conditions than their standard *in vitro* culture. *Giardia*-specific proteins were observed (Fig 4.2); however, these proteins may have been the products of death



Giardia isolate	[protein] (µg/ml)
GS	0.655
WB	1.115

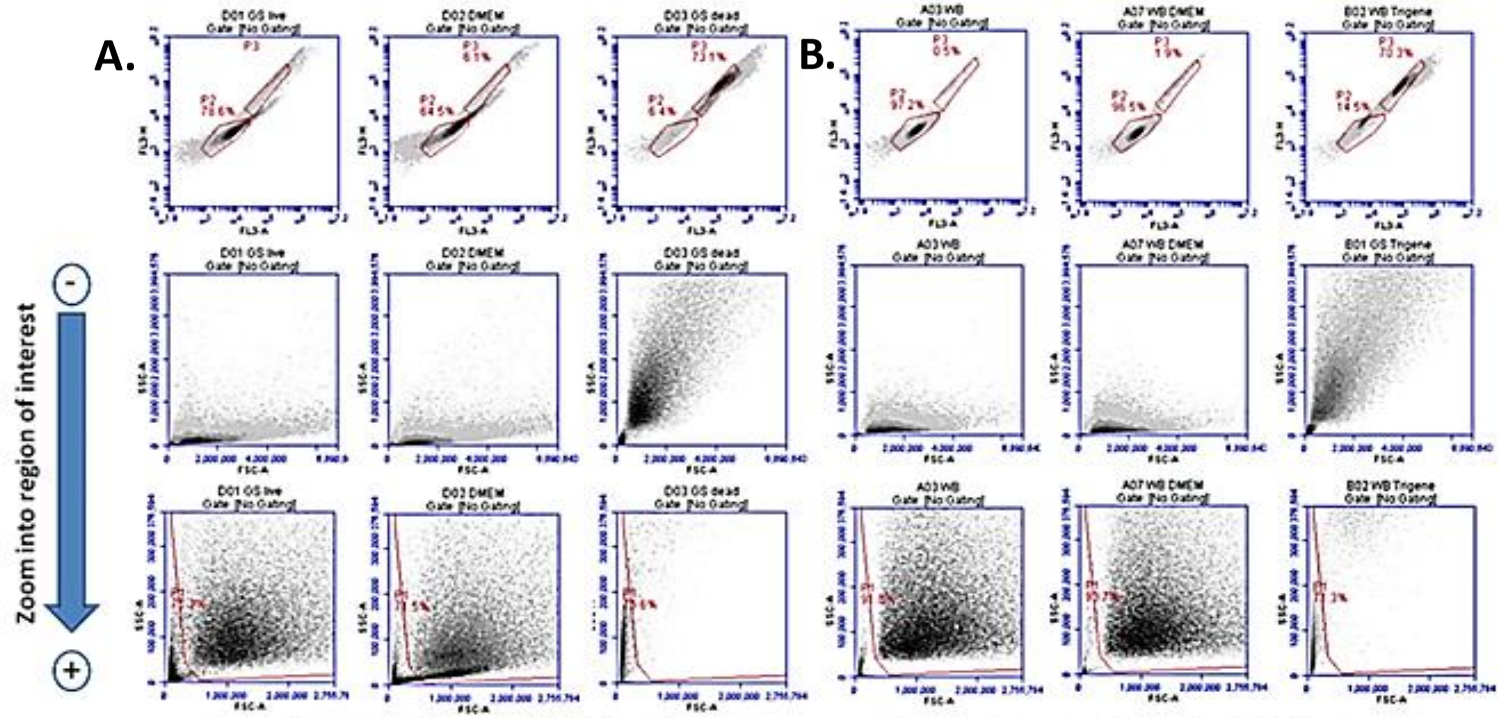
Giardia isolate	[protein] (µg/ml)
GS	1.189
WB	1.339

Fig 4. 2: *Giardia* supernatant contains less TYI-S-33 proteins after 45 minutes incubation in non-supplemented DMEM. Parasites were chilled on ice for 15 min, washed 3 times in pre-warmed PBS, centrifuged 10 min at 3,000 rpm between each wash; and then were incubated in pre-warmed non-supplemented DMEM for 45 min. at 37°C. After 45 min incubation, parasites were chilled on ice for 5 min and centrifuged 10 min at 3,000 rpm. Supernatants were collected and SDS-PAGE on 12% gels, using a SYPRO staining, and BCA assay were performed. Protocol was repeated to confirm the results**. A.** First *Giardia* supernatant protein profile and protein concentration after incubation in DMEM**. B.** Repeat of incubation in DMEM, protein profile and concentration. Data represents two independent experiments.
MW: Molecular weight; 2: TYI-S-33, 1:500 dilution; 3: GS supernatant; 4: WB supernatant; 5: non-supplemented DMEM; 6: TYI-S-33, 1:100 dilution.

cell caused by the incubation in DMEM. A viability flow cytometry assay was performed on parasites to analyse the percentage of dead and living trophozoites during incubation in DMEM (Fig 4.3). *Giardia* trophozoites collected from standard growth condition were used as a control for living population. To compare *Giardia* trophozoites population incubated in DMEM with a dead population, trophozoites collected from *Giardia* culture were killed using 2 % Trigene, a disinfectant cleaner for laboratories applications, and used as a dead population control. Both controls of living and dead populations were represented by two gates: P2 and P3 respectively (Fig 4.3). When incubated with DMEM, both *Giardia* assemblage A and B trophozoites population had a similar profile to trophozoites under standard conditions of growth; the percentage of living cells within the populations incubated in DMEM were of 97.44 % for assemblage B (GS isolate) and 97.62 % for assemblage A (WB isolate) (Fig 4.3). When 2 % trigene was added, the flow cytometry profile was completely different from the normal condition of growth with a diffused granularity compared to the living cells which had a low granularity (Fig 4.3). The percentage of parasites still alive in population resupsended in 2 % Trigene was less than 1 %. This confirmed that parasites were still alive after 45 minutes incubation in DMEM. Moreover, it showed that proteins observed on SDS-PAGE were more likely to be secreted by living parasites rather than by-products of dead cells (Fig 4.2).

4.2.2/ Pilot mass spectrometry study on *Giardia* assemblage A and B secretion profile

SDS-PAGE and flow cytometry analyses asserted that *Giardia* secretes proteins (Fig 4.2 and 4.3). To determine the secretion profile for both assemblages, replicates for both assemblage A and B were prepared to run a mass spectrometry (MS) analysis. The MS analysis was performed by Prof Johnathan Wastling and Dr Dong Xia at the Institute of Infection and Global Health, from the University of Liverpool. *Giardia* trophozoites were incubated with DMEM for 45 minutes after three washes in PBS (as stated in Chapter II, section 2.1.1.2., Fig 2.2). At first, one replicate was prepared for each assemblage, BCA protein assay performed to estimate the amount of protein present in supernatant (Table 4.1). Only one



Samples	Count	% of This Plot
GS	40,532	98.93%
GS + DMEM	42,051	97.44%
GS + 2%Trigene	40,937	99.74%

Samples	Count	% of This Plot
WB	47,872	99.28%
WB + DMEM	47,294	97.62%
WB + 2%Trigene	44,693	99.77%

Fig 4. 3: *Giardia* trophozoites are alive after incubation in non-supplemented DMEM. Parasites were chilled on ice for 15 min, washed 3 times in pre-warmed PBS, centrifuged 10 min at 3,000 rpm between each wash; and then were incubated in pre-warmed non-supplemented DMEM for 45 min. at 37°C. After 45 min incubation, parasites were chilled on ice for 5 min and centrifuged 10 min at 3,000 rpm. Pellets were collected and respuspended in PBS. Trophozoites collected from culture and respuspended in either PBS or 2% trigene were used as live and death control respectively. Proportion of living/dead trophozoites by flow cytometry. 5 µl of propidium iodide (PI) were added in each sample to stain DNA liberated in the milieu after cell death. Flow cytometry was performed using the BD Accuri™ C6 flow cytometer, with a blue laser (λ= 488 nm) and an optical filter 585/40. Gate P2 and P3 represent alive and dead trophozoites respectively. Data were analysed using the BD Accuri C-flow software (version 1.0.227.4). Data represents two independent experiments**.** A. Flow cytometry analysis for GS isolate. B. Flow cytometry analysis for WB isolate.

Table 4. 1: *Giardia* supernatant protein concentration before mass spectrometry analysis. Parasites were chilled on ice for 15 min, washed 3 times in pre-warmed PBS, centrifuged 10 min at 3,000 rpm between each wash; and then were incubated in pre-warmed non-supplemented DMEM for 45 min. at 37°C. After 45 min incubation, parasites were chilled on ice for 5 min and centrifuged 10 min at 3,000 rpm. BCA protein assays were performed before running MS analysis (either Orbitrap or QeXactive) on samples.

Giardia	[protein] (µg/ml)	
Replicate	Orbitrap	QeXactive
GS 1	0.365	-
WB 1	0.725	-
GS 2	-	0.922
WB 2	-	0.738
GS 3	-	0.672
WB 3	-	0.798

Table 4. 2: Total amount of proteins obtained for *Giardia* replicates after Orbitrap and QeXactive MS analyses. After preparation of supernatants and pellets following protocol stated in chapter 2 (section 2.1.1.2.). Pellet samples were first resuspended in 25 mM Ambic. For all samples, proteins were denatured with 1% Rapigest™ in 25 mM Ambic, followed by 3 x freeze-thaw cycles and 2 x 10 min sonication in water bath; then incubated at 80°C for 10 min, reduced with 3 mM dithiothreitol at 60°C for 10 min. Proteins were alkylated with 9 mM iodoacetamide at room temperature for 30 min in the dark, then proteins were digested by adding a 50:1 protein:trypsin ratio at 37°C overnight. MS analysis was either performed via Orbitrap MS or QeXactive MS. Bioinformatics and statistical analyses were performed using Progenesis LC-MS software (v. 4.1), Mascot software and the online database: GiardiaDB. **A.** Total amount of proteins identified for both MS techniques. **B.** Total amount of proteins in *Giardia* pellet and supernatant via the Orbitrap MS**. C.** Total amount of proteins identified in *Giardia* pellet and supernatant via QeXactive MS.

A.

Giardia isolate	Total amount of proteins identified	
	Orbitrap	QeXactive
GS	442	910
WB	660	1,025

B.

Giardia isolate	Total amount of proteins identified for Orbtirap MS	
	Pellet	Supernatant
GS	357	222
WB	650	146

C.

Giardia isolate	Total amount of proteins identified for QeXactive MS	
	Pellet	Supernatant
GS	878	677
WB	942	804

replicate per assemblage were sent to obtain a preliminary profile. Both pellet and supernatant were collected and run through an Orbitrap MS. Orbitrap is part of the linear ion trap (LTQ) MS family. A total amount of 442 proteins and 660 proteins were identified for assemblage B (GS) and assemblage A (WB) respectively (Table 4.2 A.). For assemblage B (GS isolate), 357 proteins were identified in the pellet and 222 were identified in the supernatant; as for assemblage A (WB isolate), 650 proteins were identified in the pellet and 146 in the supernatant (Table 4.2 B.). For both assemblages, most of the proteins identified in the supernatant were also identified in the pellet; 137 and 136 proteins for GS and WB isolate respectively (Fig 4.4). 85 proteins were isolated and identified in GS supernatant and only 10 were identified in WB supernatant.

Although this dataset showed an interesting preliminary secretion profile for both assemblages (data not shown), only one set of replicates was used and the assay needed to be reproduced. Two more replicates per assemblages were prepared and sent for analysis (Table 4.1). MS techniques and apparatus are constantly improving and changing. By the time the two new replicates were sent for analysis, a new mass spectrometer was developed and used at the Institute of Infection and Global Health: the QeXactive MS. The QeXactive MS is a hybrid mass spectrometer combining a quadrupole mass filter and an Orbitrap analyser; it is a very sensitive LC-MS/MS technique which allows the identification of more proteins than the Orbitrap. By using QeXactive MS, the number of proteins identified was increased by approximately two fold compared to the Orbitrap dataset. Indeed, 910 proteins and 1025 proteins were identified in GS and WB isolates respectively (Table 4.2 A.). In GS isolate, 878 proteins were identified in pellet and 677 in supernatant; in WB isolate, 942 and 804 proteins were identified in pellet and supernatant respectively (Table 4.2 C.). A lot more proteins were identified in both pellet and supernatant for both assemblages with QeXactive MS; 645 proteins for GS isolate and 721 proteins for WB isolate (Fig 4.5). Only 32 proteins were isolated only in GS supernatant; which is approximately two fold less than with the Orbitrap (Fig 4.4). Interestingly, 83 proteins were identified only in WB supernatant with QeXactive (Fig 4.5). All these results indicated that QeXactive MS was a very sensitive technique compared to the Orbitrap MS. For both dataset, proteins

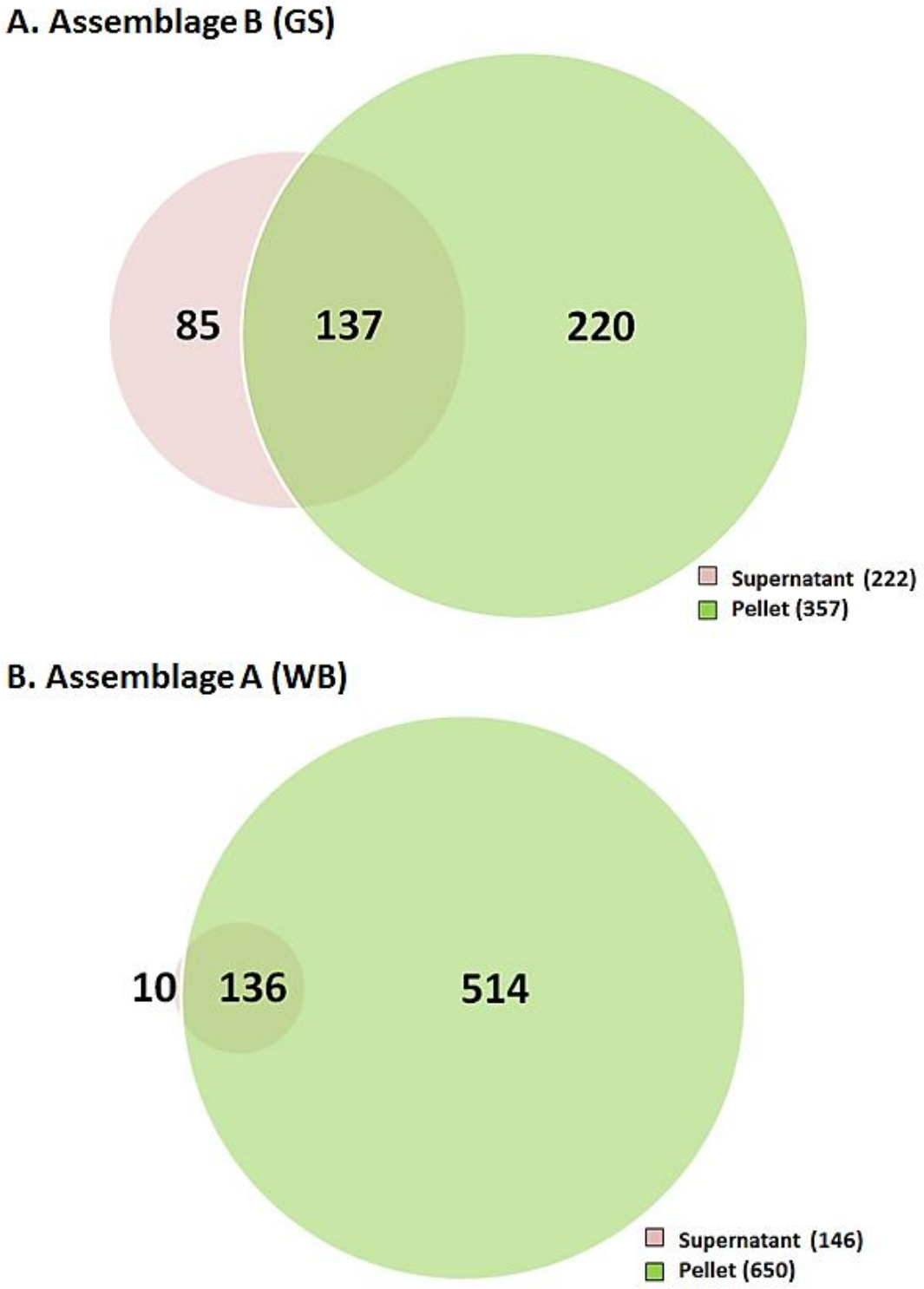


Fig 4. 4: Common proteins between pellet and supernatant identified via Orbitrap MS. **A.** 137 proteins were identified in both supernatant and pellet for GS isolate, 85 and 220 were identified only in supernatant and pellet respectively. **B.** 136 proteins were identified in both supernatant and pellet for WB isolate, 10 and 514 were identified only in supernatant and pellet respectively.

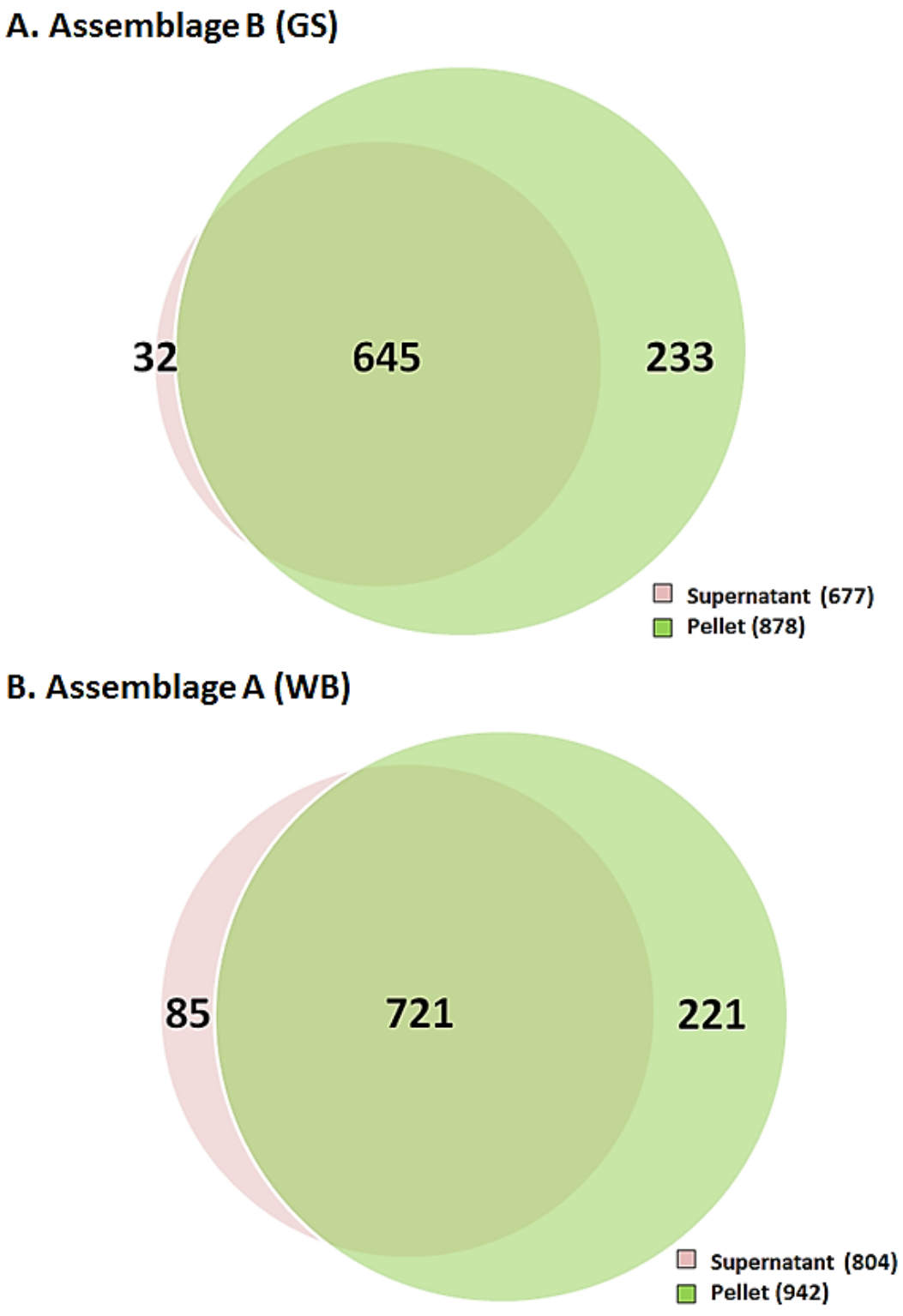


Fig 4. 5: Common proteins between pellet and supernatant identified via QeXactive MS. A. 645 proteins were identified in both supernatant and pellet for GS isolate, 32 and 233 were identified only in supernatant and pellet respectively. **B.** 721 proteins were identified in both supernatant and pellet for WB isolate, 83 and 221 were identified only in supernatant and pellet respectively.

present in both pellet and supernatant were identified, and a supernatant : pellet ratio was calculated to determine which proteins were more likely to be secreted or released by the trophozoites. Proteins were then ranked from the highest ratio to the lowest which showed a high variability in ranking between the datasets for both assemblages (Data not shown). For both Orbitrap and QeXactive datasets, proteins only present in supernatant were also ranked from the most abundant to the less abundant. Surprisingly, in these datasets, not a single protein was found in common in the two dataset, and this for each *Giardia* assemblage (Data not shown). Some proteins that were only identified in Orbitrap supernatant dataset were found to be identified in both pellet and superntant via QeXactive MS. This was explained by the high sensitiity of the QeXactive MS. Due to this high variability between dataset and to verify the correlation between the two datasets and the two replicates within QeXactive dataset, a Spearman correlation test was also performed to compare the different protein expression profiles (Fig 4.6). For both assemblages, Qexactive datasets appeared to be highly correlated indicating that data from both replicates were exploitable together (Fig 4.6). However, the Orbitrap dataset for the supernatant samples appeared to be highly variable and poorly correlated to the pellet replicate Orbitrap dataset. Therefore, for both assemblages, the two datasets were also poorly correlated which demonstrated that the datasets were not exploitable together and that the protocol had to be optimised.

4.3/ Optimisation of protocol for the preparation of *Giardia* supernatant prior to mass spectrometry analysis

Due to the low correlation between the previous datasets and the high variability in protein rankings, three new replicates per *Giardia* assemblage had to be prepared and sent to the Institute of Infection and Global Health. To ensure that the best would be obtained, each replicate was passed through both the Orbitrap MS and QeXactive MS. However, prior to run the new MS analysis, the preparation of supernatant had to be optimised, and a new step was added to concentrate the proteins present in supernatants.

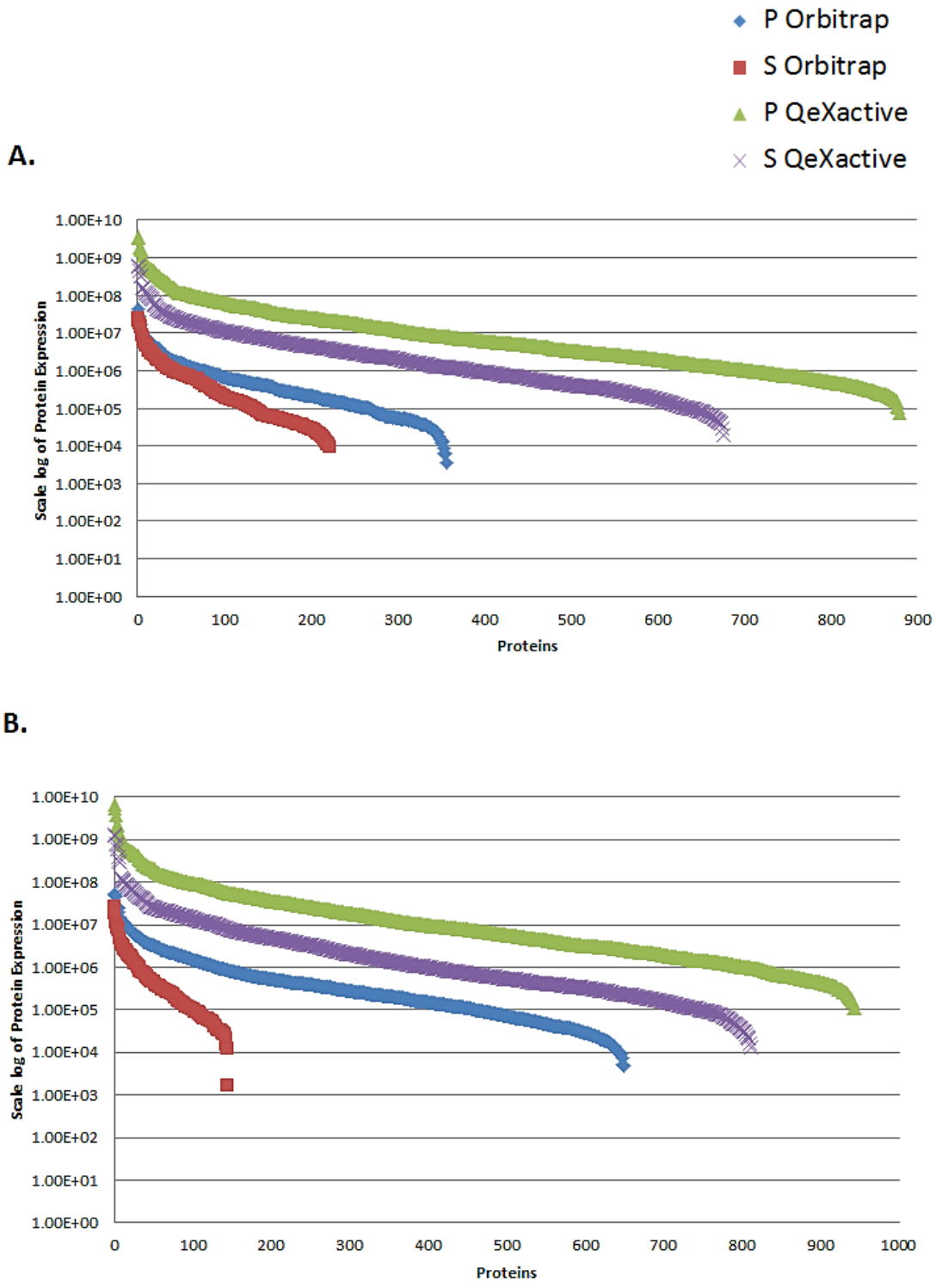


Fig 4. 6: Protein expression profile for *Giardia* assemblage A and B. Both GS and WB pellet (P) and supernatant (S) replicates were analysed via Orbitrap and QeXactive MS. Orbitrap supernatant datasets show a variation in protein expression profile compared to the pellet Orbitrap datasets and also the QeXactive datasets for both assemblage. The correlation of these two datasets was tested via Spearman correlation tests (Appendix VII). **A.** Protein expression profile for GS isolate pellet and supernatant replicates. **B.** Protein expression profile for WB isolate pellet and supernatant replicates. Data represents one independent experiment for Orbitrap dataset and two independent experiments for QeXactive dataset.

Two different approaches (i) Trichloroacetic acid (TCA) precipitation and (ii) Vivaspin columns with a 3,000 molecular weight cutt off (MWCO) were chosen and compared. Old samples stored at - 20°C were used to test which techniques was the best option for this study, which explained the low protein concentrations prior to precipitate proteins (Table 4.3). BCA protein assays were also performed after the protein concentration or precipitation in order to verify the final protein concentration in *Giardia* supernatants (Table 4.3). TCA precipitation was shown to give very low protein concentrations which suggested that this was not the best method to use in this study. Vivaspin columns were then the approach taken to optimise the protein concentration within *Giardia* supernatants. An initial protocol of (i) an initial centrifugation at 12,000 rcf for 30 minutes to remove DMEM; (ii) two centrifugations with 25 mM Ammnonium bicarbonate (Ambic) at 12,000 rcf for 30 min to remove DMEM; (iii) one hour incubation with 25 mM Ambic and 1 % Rapigest™ which is a reagent used to solubilise proteins making them more susceptible to enzymatic cleavage, at room temperature; (iv) and a final centrifugation at 3,000 rcf for 2 minutes to recover proteins (Fig 4.7). However, protein concentrations were lower than the initial concentrations, with a variation in concentration ranging from 1.5- to 5-fold (Table 4.3). Alternatively, a similar Vivaspin protocol was also performed but with an extra centrifugation at 12,000 rcf for 30 minutes, and incubation with 25 mM Ambic only. Final protein concentrations were also lower than the initial concentrations; but only by 1.65-fold (Table 4.3).

The fact that protein concentrations were lower after protein precipitation suggested that proteins may be degrading during incubation at room temperature. An SDS-PAGE electrophoresis was performed on samples concentrated with and without 1 % Rapigest™ to verify protein profiles within samples (Fig 4.8). BSA was used as a control of the technique as proteins with high molecular weights are present in high quantity. The SDS-PAGE showed that proteins were not degraded after one hour incubation at room temperature (Fig 4.8). It also showed that incubated with 25 mM Ambic only had a better protein profile as they appeared more concentrated on the gel (Fig 4.7). This indicated that the use of 25 mM Ambic without 1 % Rapigest™ allowed a better concentration of the proteins present in

Table 4. 3: Vivaspin columns give supernatant samples more concentrated in proteins than TCA precipitation. Parasites were chilled on ice for 15 min, washed 3 times in pre-warmed PBS, centrifuged 10 min at 3,000 rpm between each wash; and then were incubated in pre-warmed non-supplemented DMEM for 45 min. at 37°C. After 45 min incubation, parasites were chilled on ice for 5 min and centrifuged 10 min at 3,000 rpm. BCA protein assays were performed before and after protein concentration.

	[protein] (µg/ml)					
	TCA Precipitation		Vivaspin column with 1% Rapigest		Vivaspin column without 1% Rapigest	
Giardia isolate	Before TCA	After TCA	Before protein concentration	After protein concentration	Before protein concentration	After protein concentration
WB	0.38	0.061	0.49	0.266	0.495	0.3
GS	0.094	0.061	1.04	0.206	0.695	0.42

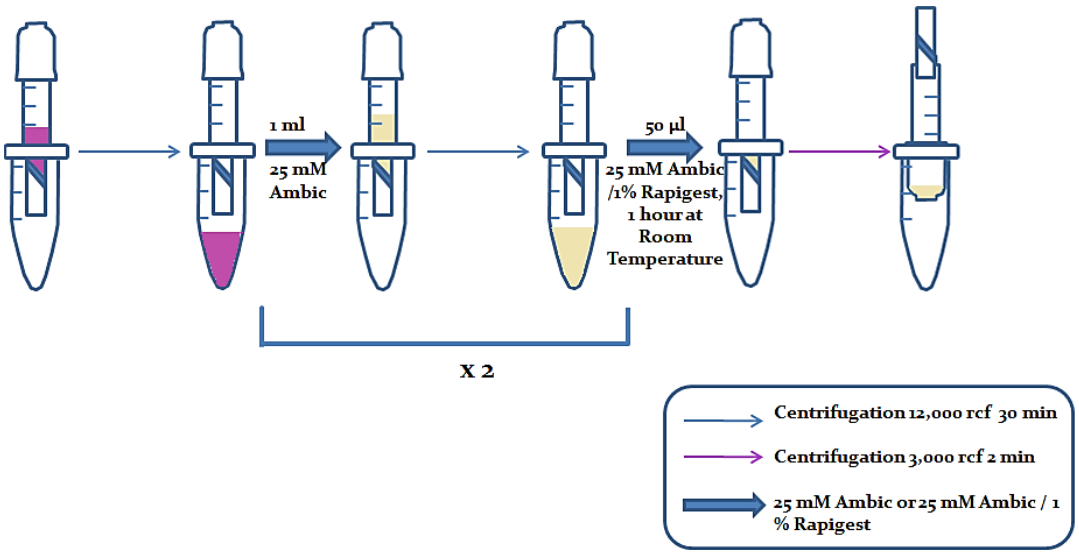


Fig 4. 7: Initial Vivaspin column protocol with 25mM Ambic / 1 % Rapigest™. Parasites were chilled on ice for 15 min, washed 3 times in pre-warmed PBS, centrifuged 10 min at 3,000 rpm between each wash; and then were incubated in pre-warmed non-supplemented DMEM for 45 min. at 37°C. After 45 min incubation, parasites were chilled on ice for 5 min and centrifuged 10 min at 3,000 rpm. Supernatants were collected and protein concentration within samples was assessed by BCA assay. Proteins were concentrated in Vivaspin columns. First, samples were spun at 12,000 rcf for 30 min. 25 mM Ambic was added and samples were centrifuged for 30 min at 12,000 rcf twice. Samples were incubated at room temperature with 50 µl 25mM Ambic/1 % Rapigest™ solution for one hour. A final centrifugation was done for 2 min at 3,000 rcf. BCA and SDS-PAGE were performed to determine protein concentration and profile.

the samples as suggested by the BCA protein assay performed previously (Table 4.3). According to these results, a final protocol of (i) an initial centrifugation at 12,000 rcf for 30 minutes to remove DMEM; (ii) 3 centrifugations at 12,000 rcf for 30 minutes; (iii) one hour incubation with 25 mM Ambic at room temperature; and (iv) a protein recovery centrifugation at 3,000 rcf for 2 minutes was set (Fig 4.9).

Once this protocol was established, *Giardia* trophozoites were incubated in non-supplemented DMEM as stated in Chapter II (section 2.1.1.2., Fig 2.2). Three replicates were prepared with at least two passages between supernatant preparations; pellet samples were stored at - 20°C until shipment to Liverpool for MS analysis. Protein concentrations in *Giardia* supernatant were assessed via BCA protein assay (Fig 4.10 B.). Proteins present in supernatants were concentrated via the final Vivaspin column protocol (Fig 4.9) and a final BCA protein assay was performed on concentrated samples (Fig 4.10 B.). An SDS-PAGE electrophoresis was performed on samples after protein concentration to verify their protein profiles (Fig 4.10 A.). According to the BCA protein assays and the SDS-PAGE gel, only two replicates were exploitable for MS analysis GS replicate 1 and WB replicate 2; therefore two more replicates per assemblage had to be prepared (Fig 4.10).

The fact that DMEM containing phenol-red was used for the incubation with *Giardia* trophozoites may have an effect on the supernatant concentrations. Indeed, phenol red may skew the BCA protein assay results. Moreover, some phenol red was still visible after protein concentration via Vivaspin columns which suggested that some DMEM was still present in samples and may interfere with the proteins recovery. To counteract that effect, phenol red-free DMEM was used to prepare the last replicates. Because no phenol red was present in these samples and to reduce the time at room temperature for the samples, two centrifugations in 25 mM Ambic at 12,000 rcf for 30 min were done instead of three previously set prior to the protein incubation at room temperature (Fig 4.9). As stated before, a BCA protein assay was performed before and after the protein concentration in Vivaspin columns, and a SDS-PAGE electrophoresis was also performed to verify the protein profile (Fig 4.11). The SDS-PAGE gel showed a good profile for GS replicates with a lot of bands; However, WB isolate showed less bands which appeared weaker than the one observed for GS replicates, particularly for the replicate WB 2

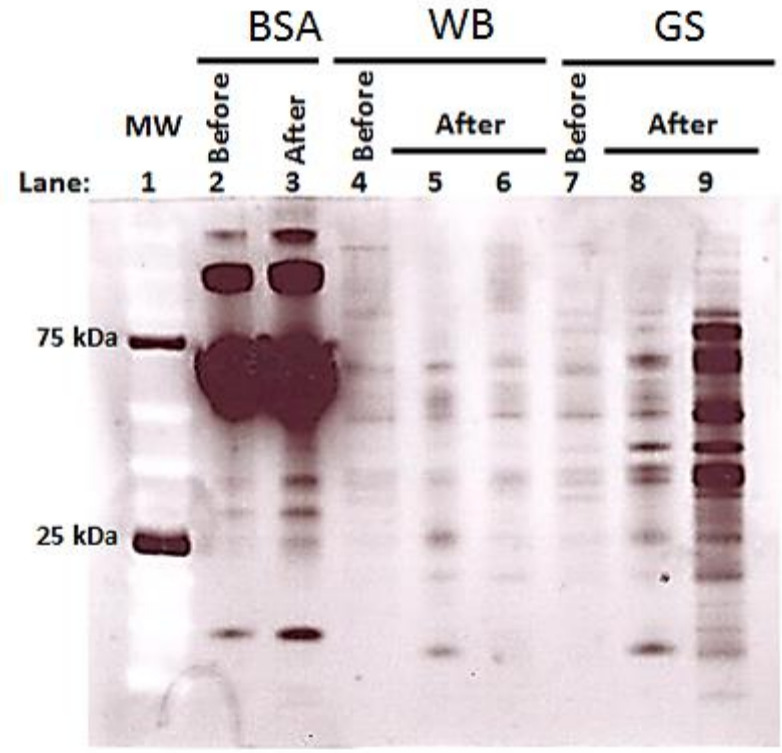


Fig 4. 8: Incubation with 25 mM Ambic without 1 % Rapigest™ gives a better protein profile for *Giardia* supernatant. Parasites were chilled on ice for 15 min, washed 3 times in pre-warmed PBS, centrifuged 10 min at 3,000 rpm between each wash; and then were incubated in pre-warmed non-supplemented DMEM for 45 min. at 37°C. After 45 min incubation, parasites were chilled on ice for 5 min and centrifuged 10 min at 3,000 rpm. Supernatants were collected and protein concentration within samples was assessed by BCA assay. Proteins were concentrated using Vivaspin columns. Lane 5 and 8: 2 x 30 min centrifugation with 25 mM Ambic, protein concentration with 1 % Rapigest™/25 mM Ambic. Lane 6 and 9: 3 x 30 min centrifugation with 25 mM Ambic, protein concentration with 25 mM Ambic. Before = before protein concentration via Vivaspin columns, After = after protein concentration using Vivaspin columns.

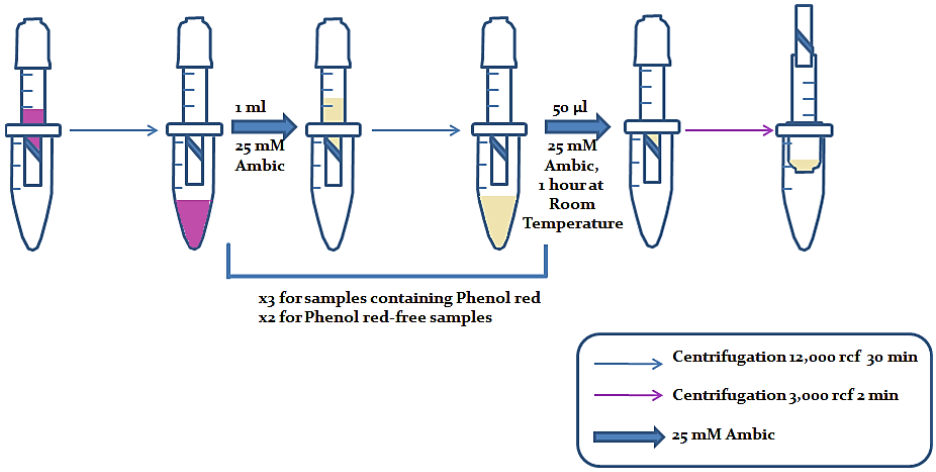
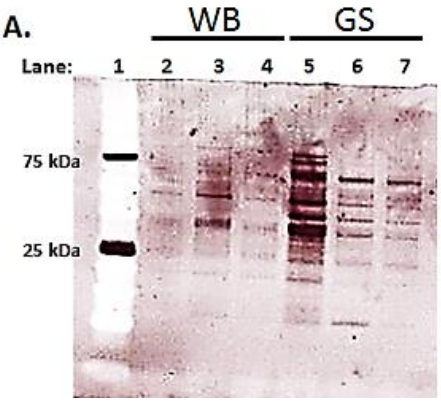


Fig 4. 9: Final protocol for the protein recovery of *Giardia* supernatant samples on Vivaspin columns. Parasites were chilled on ice for 15 min, washed 3 times in pre-warmed PBS, centrifuged 10 min at 3,000 rpm between each wash; and then were incubated in pre-warmed non-supplemented DMEM for 45 min. at 37°C. After 45 min incubation, parasites were chilled on ice for 5 min and centrifuged 10 min at 3,000 rpm. Supernatants were collected and protein concentration within samples was assessed by BCA assay. Supernatants were transferred in Vivaspin columns with a 3,000 MWCO and centrifuged at 12,000 rcf for 30 min. Proteins were washed up to 3 times with 25 mM Ambic (depending on presence of phenol red within DMEM) and centrifuged at 12,000 rcf for 30 min. Then, 50 µl of 25 mM Ambic were added and samples were left at room temperature for one hour; final spin at 3,000 rcf for 2 min to recover proteins. BCA assay and SDS-PAGE were performed to assess protein level and protein degradation prior to proteomic assay.

A.



B.

Giardia Replicate	[protein] (µg/ml) Before protein concentration	After protein concentration
WB 1	0.176	0.06
WB 2	0.256	0.41
WB 3	0.02	0.11
GS 1	0.176	0.66
GS 2	0.576	0.16
GS 3	0.456	0.21

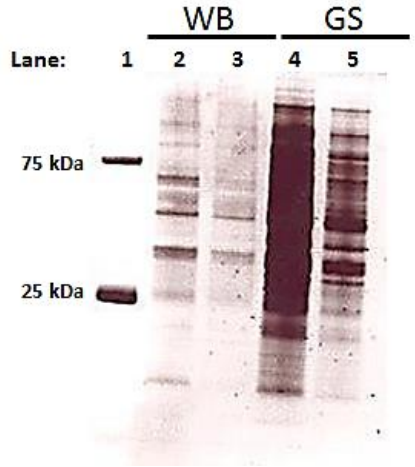
Fig 4. 10: Only WB 2 and GS 1 replicates are exploitable for mass spectrometry. Parasites were chilled on ice for 15 min, washed 3 times in pre-warmed PBS, centrifuged 10 min at 3,000 rpm between each wash; and then were incubated in pre-warmed non-supplemented DMEM for 45 min. at 37°C. After 45 min incubation, parasites were chilled on ice for 5 min and centrifuged 10 min at 3,000 rpm. Supernatants were collected and protein concentration within samples was assessed by BCA assay. Proteins were concentrated via Vivaspin columns. Supernatants were spun at 12,000 rcf for 30 min. 25 mM Ambic was added and samples were centrifuged for 30 min at 12,000 rcf, 3 times. Samples were incubated at room temperature with 50 µl 25mM Ambic solution for one hour. A final centrifugation was done for 2 min at 3,000 rcf. **A.** SDS-PAGE on 4-12% gel was performed to see the protein profile after protein concentration for the 3 replicates for each *Giardia* isolate. Lane 1: MW, 2: WB 1, 3: WB2, 4: WB 3, 5: GS 1, 6: GS 2, 7: GS. 3 **B.** BCA assays were performed before and after protein concentration on Vivaspin columns.

(Fig 4.11 A.). The BCA protein assays showed a better protein recovery than previously (Fig 4.11 B.). Even if the protein concentrations were lower than before the protein recovery, according to the SDS-PAGE and BCA assay, enough proteins were present in the replicates for MS analysis. For each *Giardia* assemblages, three replicates were finally prepared and were sent to the Institute of Infection and Global Health to be analysed via Orbitrap and QeXactive analyses (Table 4.4).

The same MS protocol as for the first MS analysis was followed in this analysis (as stated in Chapter II, section 2.4.2). For both MS techniques, peptides isolated were matched to *Giardia* proteins via the online database GiardiaDB. A total of 1,151 and 1,542 proteins were identified for assemblage B (GS isolate) via Orbitrap MS and QeXactive MS analyses respectively; which represent an increase in protein identified of approximately 1.5- to 2.5-fold compared to the first datasets (Table 4.5 A.). For assemblage A (WB isolate), a total of 1,219 and 1,641 proteins were identified via Orbitrap and QeXactive MS analysis respectively; which represent an increase in proteins of approximately 2-fold compared to the previous datasets (Table 4.5 A.). For assemblage B, 1,100 proteins were identified in pellet samples and 689 in the supernatant samples via Orbitrap MS analysis which represents a 3-fold increase in proteins compared to the first Orbitrap dataset (Table 4.2 B., 4.5 B.). The new QeXactive analysis showed a lesser increase, 1.5-fold, with 1,515 proteins identified in pellet samples and 973 in supernatants (Table 4.5 C.). For assemblage A, the new Obitrap MS dataset showed an increase of approximately 2- and 3-fold for pellet and supernatant samples respectively, with 1,184 proteins identified in pellets and 461 in supernatants (Table 4.5 B.). The new QeXactive dataset showed a similar increase in protein identification to the new Orbitrap, 1,617 proteins were identified in pellet samples and 514 in supernatant samples (Table 4.5 C.). Interestingly, these new datasets showed that less proteins were present in assemblage A supernatant than in assemblage B supernatant; which was also the case within the previous MS analysis (Table 4.2 and 4.5). A lot more proteins were identified via both MS technique which suggested that the new replicates were of better quality than the first prepared.

Even if these datasets appeared to be exploitable and prior to any secretion profile analysis, they were compared to the first MS datasets by Spearman

A.



B.

Giardia Replicate	[protein] (µg/ml) Before protein concentration	After protein concentration
WB 1	0.885	0.5375
WB 2	0.9325	1.035
GS 1	1.035	0.5375
GS 2	0.8825	0.835

Fig 4. 11: Replicates, incubated in phenol red-free non-supplemented DMEM, are good for MS analysis. Parasites were chilled on ice for 15 min, washed 3 times in pre-warmed PBS, centrifuged 10 min at 3,000 rpm between each wash; and then were incubated in pre-warmed non-supplemented, phenol red-free DMEM for 45 min. at 37°C. After 45 min incubation, parasites were chilled on ice for 5 min and centrifuged 10 min at 3,000 rpm. Supernatants were collected and protein concentration within samples was assessed by BCA assay. Proteins were concentrated in Vivaspin columns. First, samples were spun at 12,000 rcf for 30 min. 25 mM Ambic was added and samples were centrifuged for 30 min at 12,000 rcf, twice. Samples were incubated at room temperature with 50 µl 25mM Ambic solution for one hour. A final centrifugation was done for 2 min at 3,000 rcf. **A**. SDS-PAGE on 4-12% gel was performed to see the protein profile after protein concentration for the 2 new replicates for each *Giardia* isolate. Lane 1: MW, 2: WB 1, 3: WB2, 4: GS 1, 5: GS 2. **B**. BCA assays were performed before and after protein concentration on Vivaspin columns.

Table 4. 4: *Giardia* supernatant protein concentration before mass spectrometry analysis. GS 1 and WB 1 represent samples incubated in DMEM containing red phenol (GS 1 and WB2 respectively on Fig 4.7). For both GS and WB, replicates 2 and 3 represent samples incubated in red phenol-free DMEM (Replicate 1 and 2 respectively on Fig 4.8).

Giardia Replicate	[protein] (µg/ml)
GS 1	0.660
WB 1	0.410
GS 2	0.538
WB 2	0.538
GS 3	0.835
WB 3	1.035

Table 4. 5: Total number of proteins obtained for *Giardia* replicates after Orbitrap MS and QeXactive MS analyses. After preparation of supernatants and pellets following protocol stated in chapter 2 (section 2.1.1.2.). Pellet samples were first resuspended in 25 mM Ambic. For all samples, proteins were denatured with 1% Rapigest™ in 25 mM Ambic, followed by 3 x freeze-thaw cycles and 2 x 10 min sonication in water bath; then incubated at 80°C for 10 min, reduced with 3 mM dithiothreitol at 60°C for 10 min. Proteins were alkylated with 9 mM iodoacetamide at room temperature for 30 min in the dark, then proteins were digested by adding a 50:1 protein:trypsin ratio at 37°C overnight. MS analysis was either performed via Orbitrap MS or QeXactive MS. Bioinformatics and statistical analyses were performed using Progenesis LC-MS software (v. 4.1), Mascot software and the online database GiardiaDB. **A.** Total amount of proteins identified for both MS techniques. **B.** Total amount of proteins in *Giardia* pellet and supernatant via the Orbitrap MS. **C.** Total amount of proteins identified in *Giardia* pellet and supernatant via QeXactive MS. Data represents three independent experiments.

A.

Giardia isolate	Total number of proteins identified	
	Orbitrap	QeXactive
GS	1,151	1,542
WB	1,219	1,641

B.

Giardia isolate	Total number of proteins identified for Orbtirap MS	
	Pellet	Supernatant
GS	1,100	689
WB	1,184	461

C.

Giardia isolate	Total number of proteins identified for QeXactive MS	
	Pellet	Supernatant
GS	1,515	973
WB	1,617	514

correlation test. This test showed that the two datasets were poorly correlated and could not be exploited together to investigate *Giardia* trophozoites secretion profile. As for the previous dataset, the different datasets were compared via a Spearman correlation test (Fig 4.12). The previous and new datasets were shown to be poorly correlated. The new orbitrap datasets showed a high correlation within assemblages and between them. The QeXactive datasets were also highly correlated (Fig 4.12). This suggests that the protein datasets obtained from the new *Giardia* replicates were exploitable for the study of *Giardia* protein secretion profile but not the first sets of data which were excluded from the analysis. The new Orbitrap and QeXactive datasets were however analysed separately. The fact that they had a higher correlation between techniques than the first datasets but high enough may have skew the analysis if the datasets were merged. For both dataset, proteins most likely to be secreted were identified and then data were merged and compared to verify analysis (Chapter V).

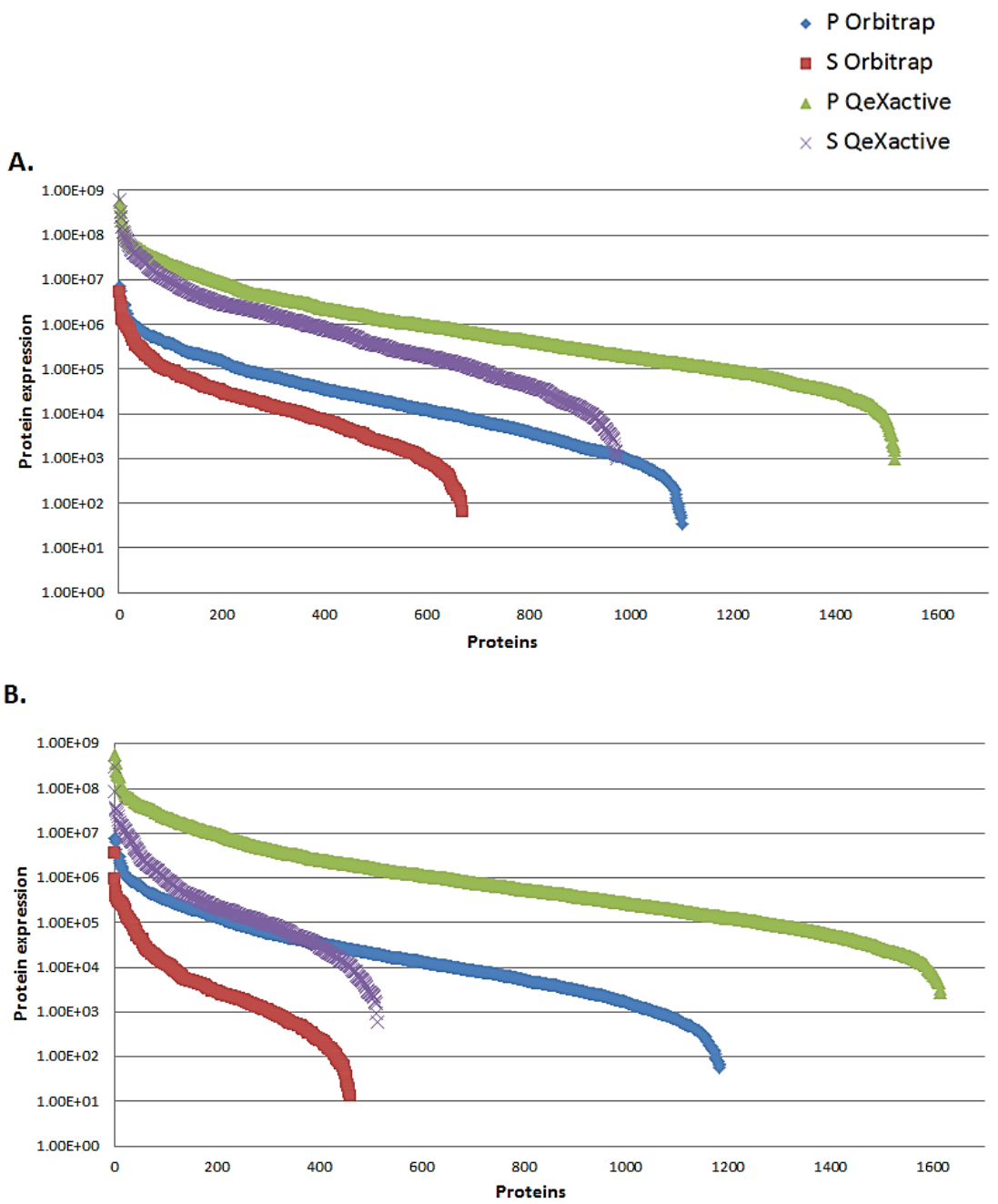


Fig 4. 12: Protein expression profile for *Giardia* assemblage A and B. Both GS and WB pellet (P) and supernatant (S) replicates were analysed via Orbitrap and QeXactive MS. Supernatant protein expression profile are similar to each other within each assemblage, so are pellet protein expression profiles. The correlation of these four new datasets was tested via Spearman correlation tests (Appendix VIII). **A.** Protein expression profile for GS isolate pellet and supernatant replicates. **B.** Protein expression profile for WB isolate pellet and supernatant replicates. Data represents three biological replicates.

Chapter V

Quantitative identification of the most abundant *Giardia* trophozoite-secreted proteins.

Glossary

Gene ontology: is the representation of a given gene and its product. The gene ontology covers three domains:

- Cellular component which is the protein location in the cell or in the extracellular environment
- Molecular function which is the elemental activities of the gene product a molecular level, e.g.: catalysis
- Biological process which represents the sets of event with a beginning and an end that are pertinent to the functioning of integrated living units such as cell, tissues, organs and organisms.

Lineage-specific protein: In this study, proteins were considered lineage-specific when no orthologs were identified in either assemblage A (WB and DH isolates) or B (GS isolate).

Orbitrap MS: Ion trap analyser trapping ions in an orbital motion around the spindle. It is part of the LTQ MS family.

QeXactive MS: Hybrid MS which combines a quadrupole mass filter and an Orbitrap analyser.

Ratioed proteins: proteins identified in both pellet and supernatant for which a supernatant: pellet ratio was obtained.

Signal peptide: Short peptide, 5 to 30 AAs long, present at the N-terminus of most of the newly synthetized proteins that are destined towards secretory pathways (Blobel and Dobberstein, 1975).

Transmembrane domain: any three-dimensional protein structure that is thermodynamically stable in a membrane. Most TM are single α helixes of usually 20 AAs (Sharpe et al., 2010).

5.1/ Introduction

Proteins are the most abundant of the cellular macromolecules. They are involved in many cellular functions such as enzymatic reaction, transport of molecules across membranes and the conservation of cellular cytoskeletal structures (Lodish et al., 2000). The proteome, which represents the cell's repertoire of proteins, is the final product of genome expression and comprises all the proteins present in the cell at a particular time. The copy numbers of individual proteins can highly vary within a cell; from less than 20,000 per cell to 100 million copies. Most of the abundant proteins are found in every eukaryotic cell type suggesting a specific housekeeping function for these abundant proteins in all eukaryotic cells (Brown, 2002).

Proteomics attempts to study biological processes by the systematic analysis of the proteins expressed in cells, tissues, organisms or sub-cellular fractions under a given set of conditions at a specific time point (Steuart, 2010, Aebersold and Goodlett, 2001). Tools for proteomics have been available since 1970; however, its true potential has only been shown within the last decade. In particular, the advances made in genomics and computational biology have permitted to develop and explore in depth proteomes (Steuart, 2010, Aebersold and Mann, 2003). The different techniques used in proteomics concentrate on separating proteins and then on identifying them. Thanks to its ability to detect and identify individual proteins from whole cells, proteomics has allowed the investigation of pathogenic organisms such as protozoan pathogens (Wastling et al., 2009). Indeed, proteomics can be used to identify key proteins within organisms providing insight for taxonomy, cell biology, pathology, host-pathogen interactions, secretion profile and towards the identification of potential new drugs/targets.

Most of the proteomic studies so far reported using *Giardia* were undertaken in trophozoites undergoing encystation (Wampfler et al., 2014, Faso et al., 2013, Lingdan et al., 2012). Only a few studies have focused on proteins secreted by *Giardia* and their role in the host-pathogen interaction (Table 5.1) (Cotton et al., 2014, Ringqvist et al., 2008, Roxstrom-Lindquist et al., 2006). These studies were focused on parasite interaction with intestinal cell lines. No studies

have yet attempted to quantify proteins which are the product of steady state secretion by healthy, growing *Giardia* trophozoites and which were hypothesized as the primary mediators of giardiasis pathology in this analysis.

The aims of this study were:

I. To identify, to the limit of existing technology, the proteins expressed by populations of healthy, growing human infective *Giardia* trophozoites

II. To further identify which of these proteins are released by healthy growing populations of human infective *Giardia* trophozoites

III. To provide quantitation or the relative abundance and strain specific expression of trophozoite proteins from two human infective assemblages.

IV. To provide quantitation of the relative abundance of retained and released proteins for both assemblages.

V. To evaluate the specific enrichment of released proteins in order to describe which are secreted by trophozoites of each assemblage.

VI. To compare the secreted proteins of each assemblage to identify secreted proteins which are conserved in both assemblages and secreted proteins which are assemblage-specific.

Briefly, the genome reference strain for each of the human infective *Giardia* assemblages was cultured, WB for assemblage A and GS for assemblage B. *Giardia* trophozoites were taken from cultures under standard conditions and harvested during the midlog phase of their growth curves. Trophozoites were washed 3x in PBS and then incubated in non-supplemented DMEM for 45 minutes at 37°C. After incubation, trophozoites and supernatant were separated by centrifugation; and both trophozoite pellet and supernatant were harvested. Supernatants and pellets were sent to the Institute of Infection and Global Health at the University of Liverpool for mass spectrometry analysis. Proteins contained in supernatant were concentrated in Vivaspin columns with 25 mM Ambic. Proteins were digested and passed through the Orbitrap MS and the QeXactive MS. Once samples had run through both Orbitrap MS and Qexactive MS, and peptides matched to their proteins on GiardiaDB.org, abundances were normalised against the reference run

Table 5. 1: List of *Giardia* factors found in supernatants proposed to have a role in the host-pathogen interactions.

Virulence factor	Function	References
Cathepsin B-like protease	Degradation of intestinal IL-8 leading to attenuation of host pro-inflammatory response	Cotton et al., 2014
Enolase	Undetermined	Ringqvist et al., 2008
ADI	Reduction of intestinal arginine levels and NO	Ringqvist et al., 2008
	Alteration of host innate defence	Ringqvist et al., 2008; Roxstrom-lindquist et al., 2006; Rodriguez-Fuentes et al., 2006
OCT	Reduction of intestinal arginine levels and NO production by intestinal cells	Ringqvist et al., 2008

by Progenesis LC-MS which highlights differences in protein expression between control and treated samples with support from statistical analysis calculated by Progenesis LC-MS software (Jackson et al., 2014). For a more reliable abundance not influenced by the size of proteins, an "intensity based absolute quantification" (iBAQ) number which represents the number of theoretically observable peptides in a protein based on an *in silico* digestion was determined (Schwanhausser et al., 2011). The average normalised abundance was then divided by the iBAQ values giving the Expression-iBAQ and datasets were then further analysed bioinformatically.

5.2/ *Giardia* trophozoite proteome

5.2.1/ QeXactive MS is more sensitive than Orbitrap MS

The results obtained from GS pellet (P), GS supernatant (SP), WB P, WB SP datasets using two distinct proteomics systems were compared. For the Orbitrap MS system 1,100 pellet and 690 supernatant proteins were identified in GS and a total of 1,184 pellet and 461 supernatant proteins in WB. Furthermore, 639 GS proteins were identified in both pellet and supernatant, 51 only in supernatant and 461 only in pellet giving a total of 1,151 GS proteins identified (Fig 5.1 A.). Similarly 426 WB proteins were present in both pellet and supernatant, 35 only in supernatant and 758 proteins only in pellet for a total of 1,219 WB proteins identified via Orbitrap MS (Fig 5.1 B.).

For QeXactive MS, 1,515 GS pellet proteins and 973 GS supernatant proteins were identified compared with 1,617 WB pellet and 514 WB supernatant proteins (Fig 5.2). Thus, both strains gave comparable and consistent results by both techniques with the sensitivity of detection being roughly 1.5 times greater for QeXactive MS and a range of detection which spanned 5 logs. Almost all of the proteins identified by use of the Orbitrap MS were also identified by the QeXactive MS, and in total the two techniques identified 1,587 GS proteins and 1,690 WB proteins (Fig 5.3). This represents over a quarter of the open reading frames (ORFs) predicted by the respective genomes in this single life-cycle stage under this steady state set of conditions (Jerlstrom-Hultqvist et al., 2010).

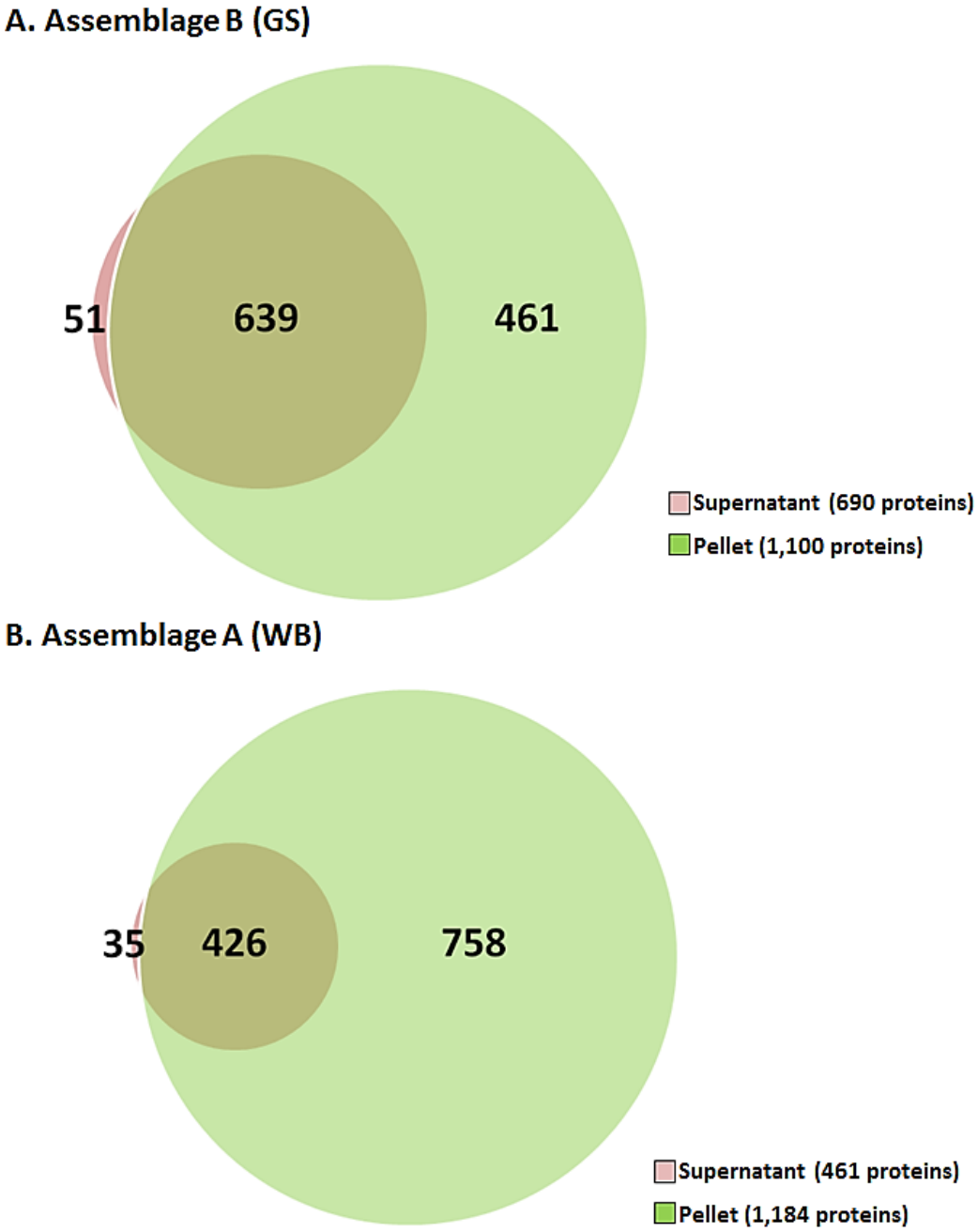


Fig 5. 1: Pellet and supernatant common proteins identified via Orbitrap MS. A. 639 proteins were identified in both supernatant and pellet for GS isolate, 51 and 461 were identified only in supernatant and pellet respectively. **B.** 426 proteins were identified in both pellet and supernatant for WB isolate, 35 and 758 proteins were present only in supernatant and pellet respectively. Data represents three biological replicates

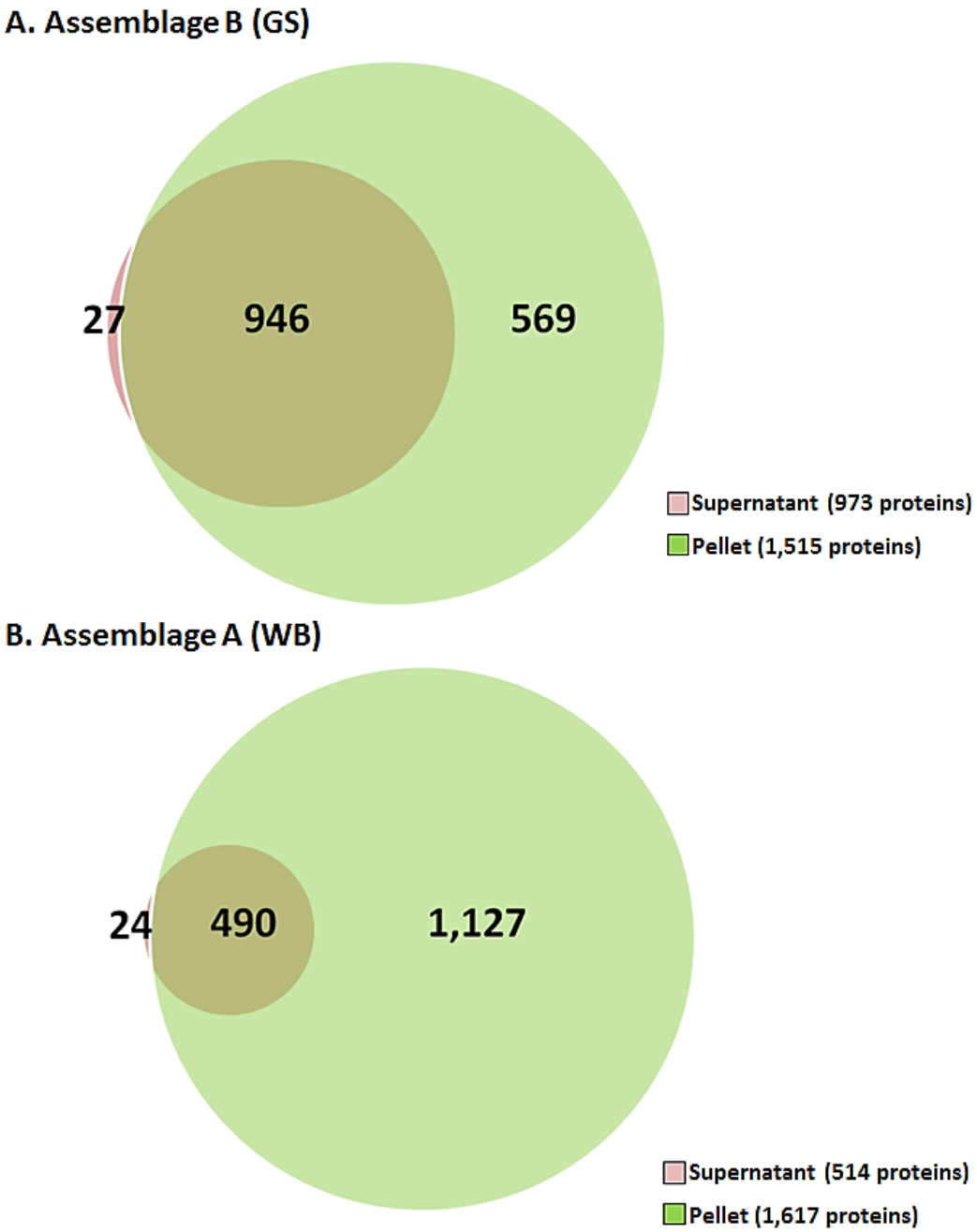


Fig 5. 2: Pellet and supernatant common proteins identified via QeXactive MS. A. 946 proteins were identified in both supernatant and pellet for GS isolate, 27 and 569 were identified only in supernatant and pellet respectively. **B.** 490 proteins were identified in both pellet and supernatant for WB isolate, 24 and 1,127 proteins were present only in supernatant and pellet respectively. Data represents three biological replicates.

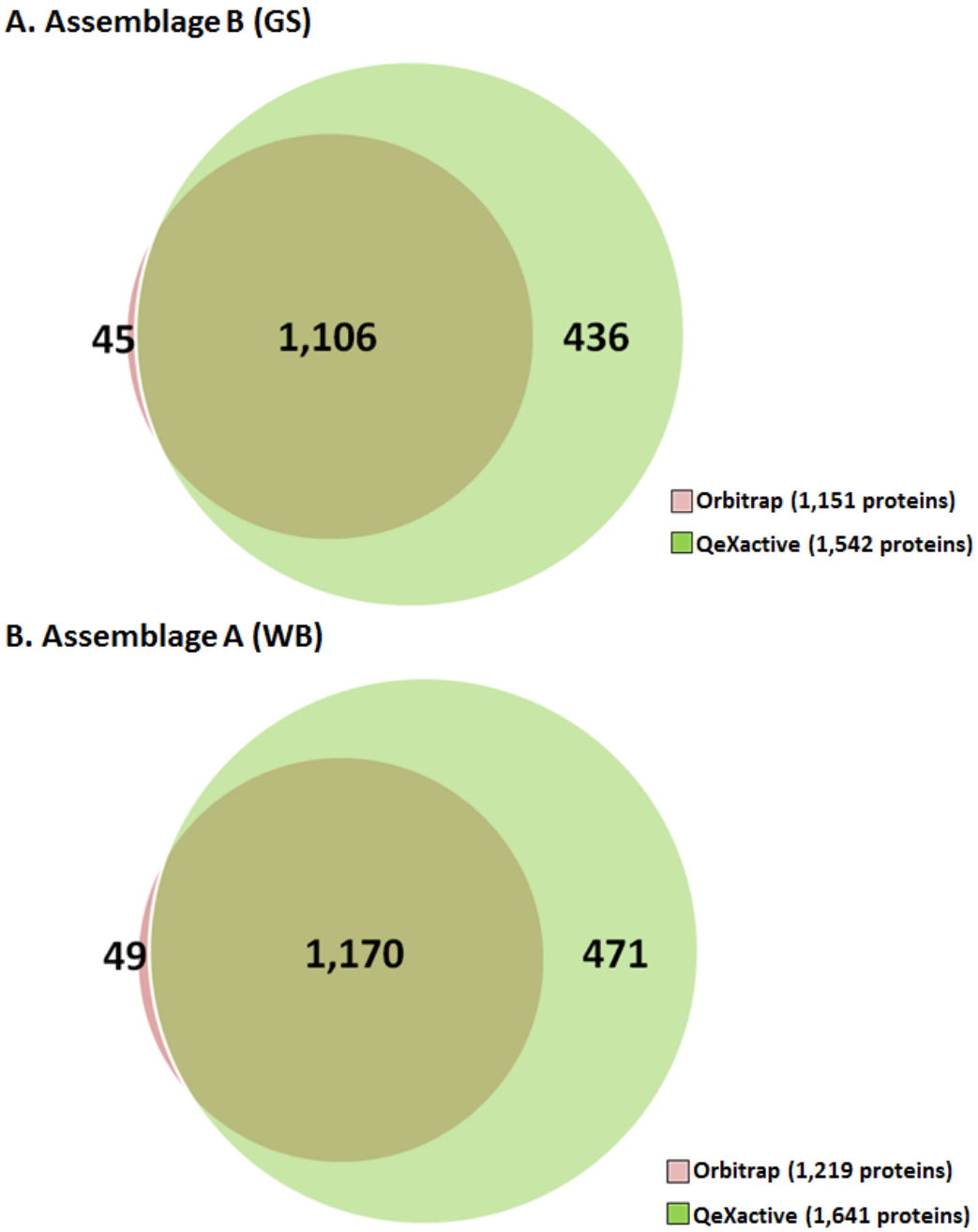


Fig 5. 3: Orbitrap and QeXactive common proteins for each *Giardia* isolate. A. 1,587 proteins were identified in total via both MS techniques for GS isolate. 1,106 proteins were present in both datasets, 42 and 439 were identified only via Orbitrap Ms and QeXactive respectively. **B.** A total of 1,690 proteins were identified for WB isolate via both MS techniques. 1,170 proteins were present in both dataset, 44 and 471 proteins were identified only in Orbitrap MS dataset and QeXactive Ms datasets respectively.

In total 1,542 GS proteins and 1,641 WB proteins were identified by QeXactive MS (Fig 5.3). Of these, 946 GS proteins were present in both pellet and supernatant, 45 GS proteins in supernatant only and 436 GS proteins in pellet only. By comparison 490 WB proteins identified in supernatant and pellet, 24 WB proteins in supernatant only and1, 127 WB proteins in pellet only. So, although both assemblages showed similar numbers of proteins identified overall, a greater proportion of proteins were identified in the supernatant of GS than of WB.

All these results suggest that QeXactive MS is a technique more sensitive than Orbitrap MS. Therefore, for the rest of the study, the data were presented according to their QeXactive supernatant abundance for a simplified and clear organisation of the results.

5.2.2/ Assemblage-specific expression of trophozoite proteins

Results showed that roughly 61 % of GS and WB proteins identified were functionally or structurally annotated in the genomic databases. Between 36 and 37% of proteins were annotated as hypothetical and only 2-3 % of annotated proteins were found to be specific to each strain (Fig 5.4 A., 5.5 A.). Most proteins identified were annotated proteins with syntenic orthologs in both assemblages which represent the single best hits in reciprocal blasts of the two genomic databases and occupy equivalent genomic locations. For the four datasets, proteins were matched to their orthologs in *Giardia* assemblage A (WB and DH isolates) and B (GS isolate) using the online database GiardiaDB (as stated in Chapter II, section 2.4, Fig 2.4 A and 2.5). Their signal peptide (SignalP) and transmembrane domain (TM) and MW were also obtained from GiardiaDB. The orthologs of roughly three quarters of the proteins were found in the proteomes of both strains.

For annotated proteins, over 85% of orthologs were identified in both assemblages. For hypothetical proteins this fell significantly to approximately 75%. 1 % of GS and 3% of WB hypothetical proteins were identified only for their own assemblage. Less than 1% of proteins identified in GS datasets recognised orthologs in sub-assemblage A2 (DH isolate) genome but not in the WB isolate (Fig 5.4 B., 5. 5 B.).

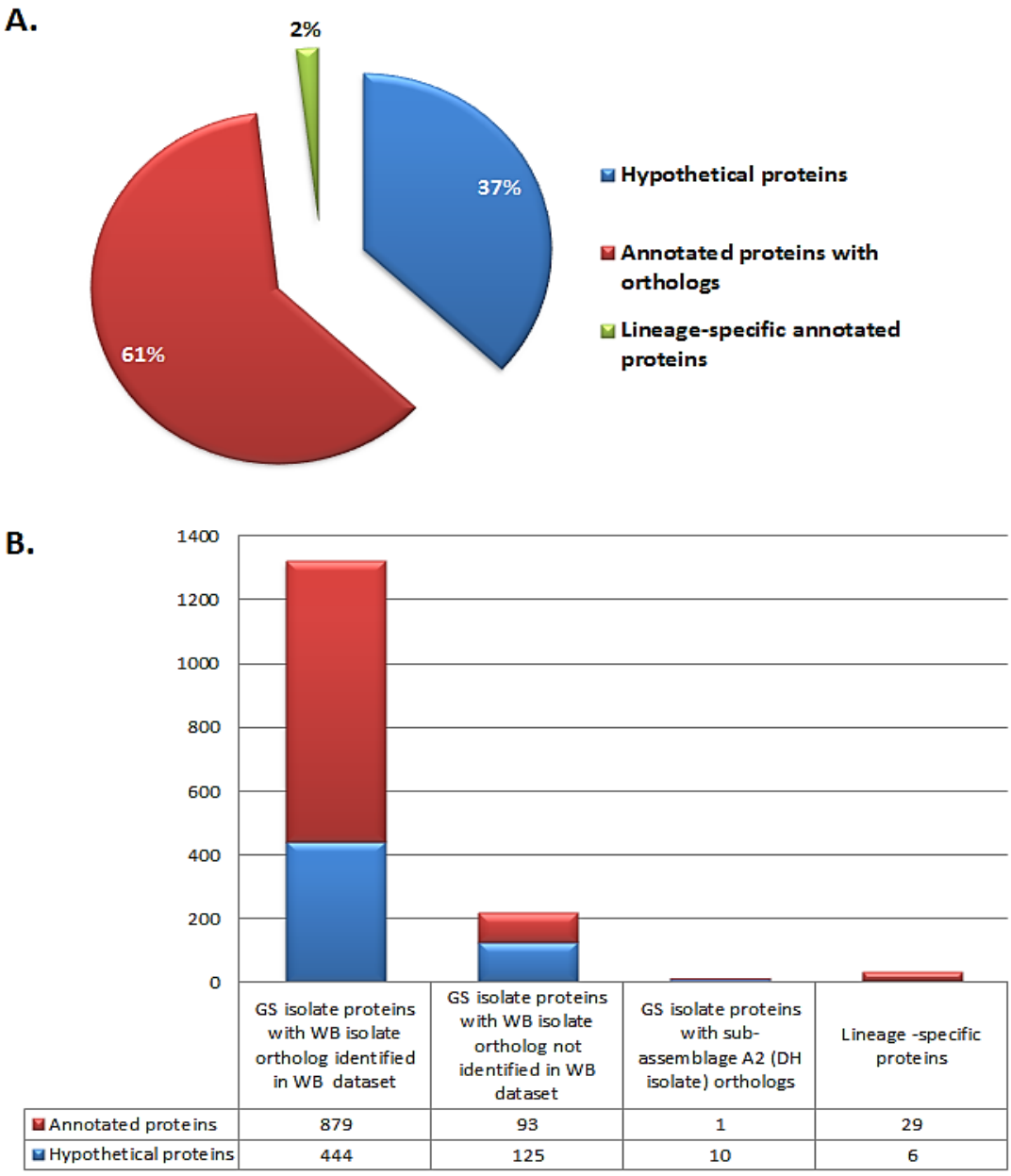


Fig 5. 4: Proportion of GS isolate hypothetical, lineage-specific and annotated proteins shared with assemblage A (WB and DH isolates). Protein orthologs were defined using gene synteny information on GiardiaDB.org. A. 1,587 proteins have been identified in total in assemblage B dataset via both MS techniques, 585 proteins (37 %) are uncharacterised hypothetical proteins (including some lineage-specific proteins), 974 (61 %) are annotated proteins that are orthologous to assemblage A, and 29 (2 %) are annotated assemblage B-specific proteins. **B.** 1,323 proteins are also identified in WB isolate dataset as orthologs, 218 proteins have known orthologs which were either not included in the proteomics dataset cut-off or were not identified, 11 proteins have assemblage A2 orthologs known, and 35 are assemblage B-specific proteins. Data represent three biological replicates.

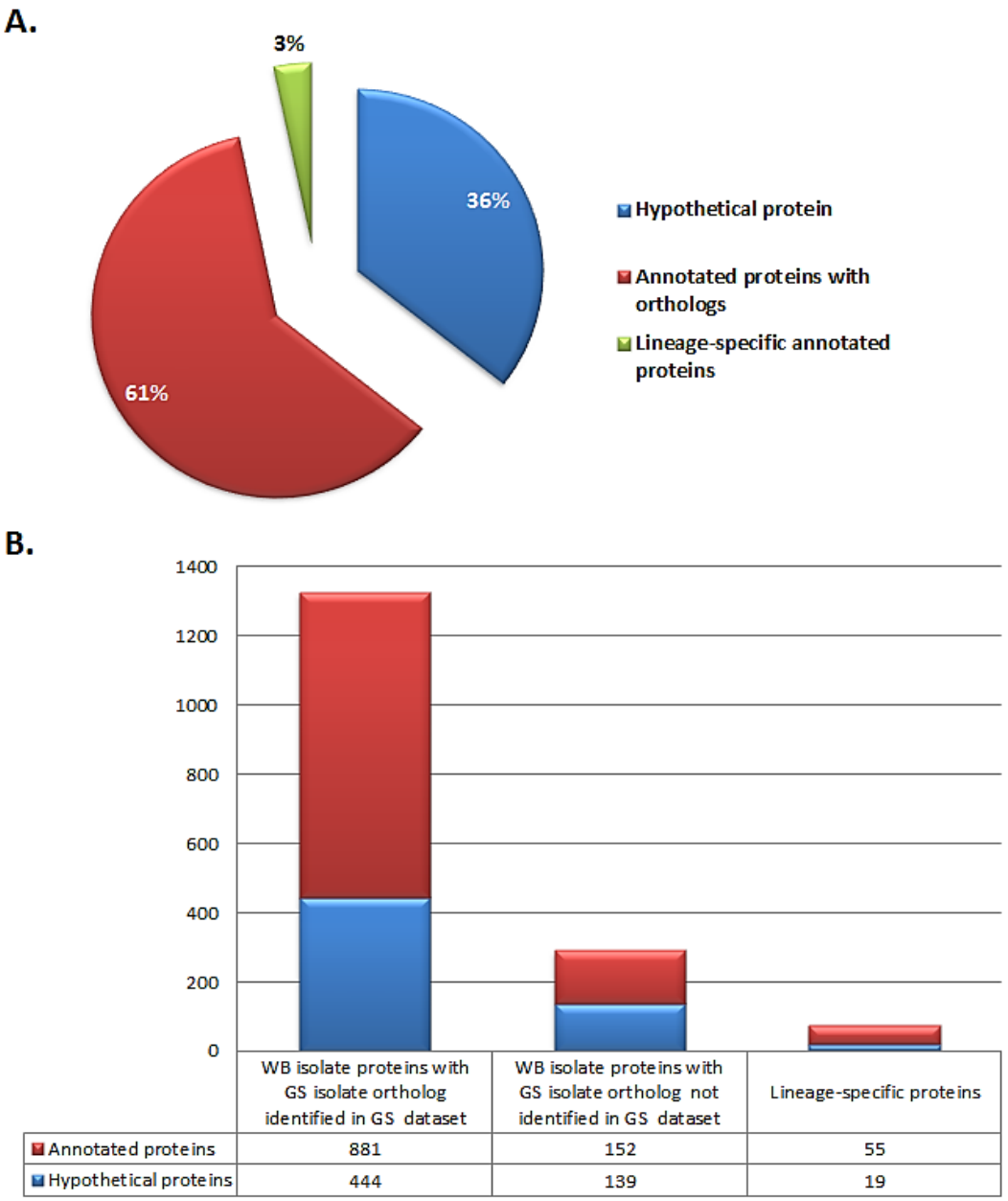


Fig 5. 5: Proportion of WB isolate hypothetical, lineage-specific and annotated proteins shared with assemblage B (GS isolate). Protein orthologs were defined using gene synteny information on GiardiaDB.org. **A**. 1,690 proteins have been identified in total in assemblage A dataset via both MS techniques, 602 proteins (36 %) are uncharacterised hypothetical proteins (including some lineage-specific proteins,) 1,033 (61 %) are annotated proteins that are orthologous to assemblage A, and 55 (2 %) are annotated assemblage A-specific proteins. **B**. 1,326 proteins have also been identified in GS isolate dataset as orthologs, 290 proteins have known orthologs which were not included in the proteomics dataset cut-off or were not identified, and 74 proteins are assemblage A-specific proteins. Data represents three biological replicates.

Comparison between the WB and GS proteomes identified VSPs, high cysteine membrane proteins (HCMPs), NEK kinases, and hypothetical proteins as the major classes of lineage-specific proteins (Tables 5.2 and 5.3). A few of the GS isolate proteins were also shown to have an ortholog in DH isolate (sub-assemblage A2) but not in WB isolate (Table 5.2).

5.3/ Proteins secreted by healthy trophozoites during logarithmic growth.

Others have observed high levels of abundant cytoplasmic house-keeping proteins released into media by cultured trophozoites (Skarin et al., 2011, Ringqvist et al., 2008), whether this represents an inherent leakiness, or subpatent levels of cell death in cultures is not clear but for this study it was reasoned that although the levels of such proteins could be high, the levels could never be enriched in the supernatants compared with the levels present in the cells. Thus, by comparing normalised concentrations for each protein found in the supernatant with that found in the cell, it was reasoned that the highest level of enrichment would be of those that were truly secreted and those expressed on the external face of the parasite surface and then clipped or shed. It was realised, though, that very soluble highly abundant cytosolic proteins would be present at higher levels than insoluble cytoskeletal proteins or proteins associated with internal membrane systems, that quantitation was likely to be less robust for proteins expressed at low concentrations which might limit the identification of proteins secreted at very low levels and that some biases might exist to make some proteins more readily detectable than others.

To evaluate supernatant enrichment, proteins identified in the SP datasets were gathered and compared to their concentration in the pellet to provide a ratio using the following formula: $\frac{SP\ Expression}{P\ Expression}$. These proteins were then ranked from highest to lowest ratio and an arbitrary cut-off invoked such that the top 50 were considered as the most likely to be secreted. Proteins identified only in SP were also included in the analysis as most likely to be secreted. All the proteins selected as "of interest" were ranked according to their SP expression from most to least abundant

Table 5. 2: List of the *Giardia* GS isolate proteins ortholog to sub-assemblage A2 (DH isolate) proteins and lineage-specific proteins identified via Orbitrap and QeXactive MS. Protein sequences were compared to their coding sequence and matched to their orthologs in Assemblage A (WB and DH isolates) using *Giardia* database: GiardiaDB.org. Annotated proteins are highlighted in red and hypothetical proteins in blue. Proteins were ranked according to QeXactive SP abundance, from most to least abundant.

Assemblage B		Sub-assemblage A2 (DH isolate) syntenic orthologs		Orbitrap		QeXactive	
Protein accession number	Protein Description	Protein accession number	Protein Description	Pellet Expression-iBAQ	Supernatant Expression-iBAQ	Pellet Expression-iBAQ	Supernatant Expression-iBAQ
GL50581_2786	VSP	-	-	1.01E+06	1.04E+06	7.13E+07	7.41E+07
GL50581_2954	VSP	-	-	1.05E+05	2.29E+05	1.18E+07	2.25E+07
GL50581_3367	Axoneme-associated protein GASP-180	-	-	6.40E+05	1.31E+05	4.14E+07	1.44E+07
GL50581_1791	VSP	-	-	7.80E+04	1.46E+05	7.09E+06	7.72E+06
GL50581_2417	Protein disulfide isomerase PDI4	-	-	1.73E+05	5.82E+04	1.41E+07	7.41E+06
GL50581_2495	Alpha-2 giardin	-	-	3.93E+05	1.02E+05	8.41E+06	5.90E+06
GL50581_4146	Alpha-14 giardin	-	-	8.50E+04	3.63E+04	5.42E+06	5.21E+06
GL50581_418	Coiled-coil protein	-	-	4.01E+05	1.08E+05	1.29E+07	4.20E+06
GL50581_3619	Cathepsin B precursor	-	-	-	-	9.52E+05	2.11E+06
GL50581_1933	Hypothetical protein	DHA2_153131	Hypothetical protein	2.13E+04	1.19E+04	3.01E+06	1.68E+06
GL50581_1	VSP	-	-	1.15E+04	2.76E+04	2.36E+05	1.46E+06
GL50581_968	Protein C21orf2	-	-	2.65E+04	1.08E+04	1.50E+06	1.31E+06
GL50581_3401	VSP	-	-	4.40E+03	1.10E+04	4.93E+05	1.22E+06
GL50581_1971	Kinase, STE STE20	-	-	1.53E+05	6.86E+03	7.67E+06	1.04E+06
GL50581_4162	Hypothetical protein	DHA2_152684	Hypothetical protein	6.28E+03	8.70E+03	4.97E+05	8.75E+05
GL50581_1059	Alpha-19 giardin	-	-	3.32E+04	1.29E+04	1.87E+06	6.99E+05
GL50581_3192	Hypothetical protein	-	-	2.94E+04	4.83E+03	1.19E+06	6.61E+05
GL50581_572	VSP	-	-	1.48E+03	4.04E+03	6.50E+04	5.39E+05
GL50581_3818	VSP	-	-	-	1.06E+04	7.85E+04	5.06E+05
GL50581_1902	VSP	DHA2_150066	VSP	1.11E+05	1.62E+05	1.58E+05	4.27E+05
GL50581_3457	High cysteine protein	-	-	-	1.49E+04	3.43E+04	3.48E+05
GL50581_85	High cysteine membrane protein Group 4	-	-	-	-	1.13E+05	2.69E+05

End of Table 5. 2.

GL50581_902	Hypothetical protein	DHA2_152876	DNA-dependent RNA polymerase beta' subunit/160 kD subunit	1.63E+04	3.47E+03	1.12E+06	2.52E+05
GL50581_3990	Hypothetical protein	DHA2_154376	Putative SUN domain protein	1.28E+04	3.89E+02	1.49E+06	1.53E+05
GL50581_4507	Kinase, AGC AKT	-	-	4.65E+03	-	1.04E+06	6.83E+04
GL50581_4510	Hypothetical protein	DHA2_150600	Hypothetical protein	1.92E+03	-	2.80E+05	8.12E+03
GL50581_2669	Protein 21.1	-	-	1.64E+03	-	1.26E+04	2.85E+03
GL50581_303	CAMP-specific 3',5'-cyclic phosphodiesterase 4B	-	-	-	-	1.03E+05	-
GL50581_159	Cathepsin B precursor	-	-	1.38E+04	-	-	-
GL50581_3336	Copine	-	-	1.34E+04	-	-	-
GL50581_322	High cysteine membrane protein Group 4	-	-	-	3.68E+03	-	-
GL50581_3037	High cysteine protein	-	-	2.27E+04	7.21E+03	8.62E+05	-
GL50581_2285	Hypothetical protein	-	-	5.76E+03	-	-	-
GL50581_2613	**Hypothetical protein**	-	-	-	-	**8.02E+05**	-
GL50581_249	Hypothetical protein	-	-	-	-	1.89E+04	-
GL50581_2719	Hypothetical protein	-	-	3.17E+03	-	6.12E+05	-
GL50581_3585	Hypothetical protein	-	-	-	-	8.60E+04	-
GL50581_3209	Kinase, NEK	-	-	5.38E+02	-	3.48E+04	-
GL50581_3224	Kinase, NEK	-	-	-	1.61E+03	-	-
GL50581_1495	RNA-dependent helicase p68, putative	-	-	1.09E+03	-	-	-
GL50581_2487	VSP	-	-	-	4.37E+03	2.29E+05	-
GL50581_261	Hypothetical protein	DHA2_150373	Hypothetical protein	4.29E+03	-	4.29E+05	-
GL50581_2869	Hypothetical protein	DHA2_152889	Hypothetical protein	1.71E+03	-	7.60E+04	-
GL50581_731	Hypothetical protein	DHA2_153916	Hypothetical protein	-	-	3.29E+04	-
GL50581_643	Hypothetical protein	DHA2_150523	Hypothetical protein	-	-	2.61E+04	-
GL50581_3981	Hypothetical protein	DHA2_154381	Hypothetical protein	-	-	1.41E+05	-

Table 5. 3: List of the *Giardia* WB isolate lineage-specific proteins identified via Orbitrap and QeXactive MS. Protein sequences were compared to their coding sequence and matched to their orthologs in Assemblage B (GS isolate) using *Giardia* database: GiardiaDB.org. Annotated proteins are highlighted in red and hypothetical proteins in blue. Proteins were ranked according to QeXactive SP abundance, from most to least abundant.

Assemblage A		Orbitrap		QeXactive	
Protein accession number	Protein Description	Pellet Expression-iBAQ	Supernatant Expression-iBAQ	Pellet Expression-iBAQ	Supernatant Expression-iBAQ
GL50803_33279	VSP	6.18E+06	3.61E+06	3.68E+08	2.98E+08
GL50803_104250	hypothetical protein	1.93E+06	2.84E+05	9.67E+07	1.59E+07
GL50803_137740	VSP-3	2.57E+05	2.61E+05	1.47E+07	1.18E+07
GL50803_112312	Elongation factor 1-alpha	-	4.11E+04	-	2.22E+06
GL50803_114815	Tenascin precursor	1.69E+04	1.17E+05	1.69E+06	1.76E+06
GL50803_103713	Protein disulfide isomerase PDI4	2.89E+05	1.38E+04	2.44E+07	1.68E+06
GL50803_137613	VSP-188 with INR	4.51E+04	2.35E+04	2.41E+06	1.36E+06
GL50803_103454	High cysteine membrane protein Group 1	1.98E+04	1.99E+04	2.11E+06	1.33E+06
GL50803_102813	Protein 21.1	4.28E+04	1.33E+04	3.16E+06	9.89E+05
GL50803_137716	Axoneme-associated protein GASP-180	7.13E+05	7.41E+03	5.01E+07	6.28E+05
GL50803_102841	hypothetical protein	8.66E+03	-	1.08E+06	5.66E+05
GL50803_114787	Alpha-7.3 giardin	5.05E+04	1.10E+04	4.70E+06	3.64E+05
GL50803_15097	Alpha-14 giardin	8.70E+04	2.10E+03	8.81E+06	3.56E+05
GL50803_101074	VSP-88 with INR	4.59E+03	1.78E+03	4.04E+05	1.40E+05
GL50803_113677	Coiled-coil protein	6.33E+04	1.27E+03	4.64E+06	1.38E+05
GL50803_112473	hypothetical protein	-	-	-	1.35E+05
GL50803_16891	Protein C21orf2	2.16E+04	5.53E+02	1.24E+06	1.29E+05
GL50803_10238	hypothetical protein	3.91E+02	-	-	1.21E+05
GL50803_114777	hypothetical protein	9.27E+03	-	3.92E+05	9.41E+04
GL50803_8235	Protein 21.1	1.63E+04	6.28E+01	1.43E+06	8.87E+04
GL50803_114626	High cysteine membrane protein EGF-like	3.81E+02	4.43E+01	2.07E+05	3.31E+04
GL50803_4026	Alpha-19 giardin	1.81E+04	-	1.08E+06	2.10E+04

Table 5.3, cont.

GL50803_41472	VSP-49	-	-	-	2.08E+04
GL50803_135003	Histone H4	3.89E+05	-	3.43E+07	2.02E+04
GL50803_111936	VSP-162.1	-	-	-	1.86E+04
GL50803_137618	VSP-8	1.75E+03	7.47E+02	9.58E+04	1.68E+04
GL50803_114674	Hypothetical protein	-	7.63E+02	2.98E+04	1.40E+04
GL50803_137617	VSP-77	8.33E+02	1.54E+03	2.07E+04	1.09E+04
GL50803_14586	VSP with INR	-	8.03E+01	1.95E+04	7.81E+03
GL50803_103373	Alpha-7.1 giardin	1.63E+04	-	2.51E+06	6.81E+03
GL50803_32778	Protein 21.1	2.25E+04	-	1.80E+06	5.70E+03
GL50803_112633	High cysteine membrane protein EGF-like	-	-	-	5.60E+03
GL50803_112208	VSP-98.1 with INR	-	-	-	5.41E+03
GL50803_101307	Kinase, NEK	7.07E+04	-	2.21E+06	-
GL50803_101534	Kinase, NEK	2.53E+04	-	2.57E+06	-
GL50803_102542	Kinase, NEK	2.28E+03	-	-	-
GL50803_103540	hypothetical protein	-	-	4.81E+04	-
GL50803_107659	hypothetical protein	-	-	2.07E+05	-
GL50803_111933	VSP-162.2	3.32E+03	1.93E+01	-	-
GL50803_112018	hypothetical protein	-	-	2.47E+04	-
GL50803_112126	High cysteine membrane protein Group 3	-	-	9.69E+05	-
GL50803_112518	Kinase, NEK	4.67E+03	-	6.61E+05	-
GL50803_112673	High cysteine membrane protein VSP-like	1.06E+03	-	-	-
GL50803_112681	NSF	5.20E+04	1.43E+02	2.55E+06	-
GL50803_112682	hypothetical protein	1.95E+04	-	-	-
GL50803_112828	High cysteine membrane protein Group 4	-	-	2.19E+04	-
GL50803_112831	Cysteine protease	-	-	1.15E+04	-
GL50803_113094	Kinase, NEK	4.91E+02	-	3.95E+04	-
GL50803_113130	hypothetical protein	-	-	3.13E+05	-
GL50803_113213	High cysteine membrane protein EGF-like	1.84E+03	-	-	-
GL50803_11364	Kinase, AGC AKT	1.39E+04	-	1.67E+06	-
GL50803_113722	hypothetical protein	1.16E+03	-	1.53E+05	-
GL50803_114442	hypothetical protein	-	-	7.14E+05	-
GL50803_114623	hypothetical protein	3.61E+03	-	2.91E+05	-
GL50803_114636	hypothetical protein	-	-	3.15E+04	-

End of Table 5.3.

GL50803_114671	Protein 21.1	8.73E+02	-	2.17E+05	-
GL50803_115797	VSP-54 with INR	-	2.16E+01	-	-
GL50803_137708	VSP-48.2	-	2.49E+02	-	-
GL50803_137737	Kinase, NEK	3.39E+04	-	6.20E+05	-
GL50803_14679	Protein 21.1	5.62E+03	-	5.44E+05	-
GL50803_15004	NADPH oxidoreductase, putative	-	2.43E+03	-	-
GL50803_15965	Protein 21.1	3.31E+03	-	3.72E+05	-
GL50803_17273	hypothetical protein	-	-	7.75E+04	-
GL50803_22165	Kinase, STE STE20	8.66E+04	1.49E+02	5.31E+06	-
GL50803_27717	High cysteine membrane protein Group 3	1.53E+04	-	1.18E+06	-
GL50803_29078	hypothetical protein	-	-	7.88E+04	-
GL50803_32999	hypothetical protein	4.85E+04	-	2.71E+06	-
GL50803_40224	MDR protein-like protein	4.19E+03	-	2.77E+05	-
GL50803_6812	hypothetical protein	3.60E+04	-	2.64E+06	-
GL50803_8245	Glucosamine-6-phosphate deami0se	-	-	1.80E+04	-
GL50803_88245	Protein 21.1	1.18E+03	-	8.55E+04	-
GL50803_9327	Kinase, NEK	1.40E+03	-	6.84E+04	-
GL50803_94927	Kinase, NEK	6.91E+02	-	6.90E+04	-
GL50803_9870	Kinase, NEK	1.92E+03	-	9.88E+04	-

to obtain a quantitative secretion profile for each isolate and this was performed for each MS technique. Orbitrap and QeXactive secretion profile datasets were compared within isolate. Proteins were considered as most likely secreted when identified as such by QeXactive MS and confirmed by Orbitrap MS. The different secretion profiles were also compared between isolates to look for commonalities and determine which proteins may be "conserved" rather than "strain-specific" secreted factors.

The results showed a set of 15 proteins that were identified in both isolates by both techniques (Table 5.4). One protein was identified as annotated in GS isolate and hypothetical in WB isolate (GL50581_4180 Tenascin, GL50803_113038 Hypothetical protein). Apart from this protein, 11 proteins were annotated and identified with the same ortholog. These annotated proteins were mostly represented by Tenascins and Cathepsin B precursors. One VSP was also identified as secreted by both isolates, and an extracellular nuclease. Three proteins were hypothetical proteins. The five most abundant, secreted proteins appeared to be the same in both isolates: a hypothetical protein, followed by two Tenascins, a Cathepsin B precursor and another Tenascin (Table 5.4). Below the fifth most abundant conserved proteins, proteins appeared to have a greater variation in their SP abundances.

Within GS dataset, 31 proteins appeared to be most likely secreted; including the 15 conserved secreted proteins (Table 5.5). 23 proteins had an ortholog in WB isolate, 2 were ortholog to sub-assemblage A2 DH isolate, and 6 proteins were lineage-specific. 23 of these proteins were annotated and 8 hitherto hypothetical proteins. Most of the proteins secreted by GS isolate appeared to be VSPs and Tenascin precursors. Within WB dataset, 44 proteins appeared to be most likely secreted by trophozoites, including the 15 conserved secreted proteins (Table 5.6). 34 were orthologous to GS isolate proteins and 10 were lineage-specific. 36 proteins were annotated and 7 were hitherto hypothetical proteins. More *Giardia* protein families were represented in WB isolate secretion profile than in GS isolate.

Interestingly, the five most abundant conserved secreted proteins were also present in the top 10 secreted proteins amongst other VSPs, Tenascins, and Cathepsin B; and this regardless of the MS technique or the isolate. All these

Table 5. 4: *Giardia* trophozoite conserved secreted factors identified by Orbitrap MS and QeXactive MS. 15 proteins identified as most likely to be secreted by both GS and WB isolates. One protein was shown to be a Tenascin protein in GS isolate and a hypothetical protein in WB (highlighted in grey). Apart from this protein, 11 are annotated proteins (shown in red) and 3 are hypothetical proteins (shown in blue). Proteins are ranked according to GS QeXactive SP expression from most to least abundant. Protein ranking represents the proteins rank within this table, from most to least abundant.

Assemblage B	Assemblage A		Assemblage B (GS isolate)												Assemblage A (WB isolate)											
			Orbitrap			QeXactive									Orbitrap			QeXactive								
Protein Accesion Number	Protein Accesion Number	Protein Description	P Expression-iBAQ	SP Expression-iBAQ	ratio	P Expression-iBAQ	SP Expression-iBAQ	SP/P Ratio	Protein ranking	GO function	GO location	SignalP[a]	MW[b] (Da)	TM[c]	P Expression-iBAQ	SP Expression-iBAQ	ratio	P Expression-iBAQ	SP Expression-iBAQ	SP/P Ratio	Protein ranking	GO function	GO location	SignalP[a]	MW[b] (Da)	TM[c]
GL50581_4133	GL50803_5810	hypothetical protein	8.02E+05	1.25E+06	1.557457	5.71E+07	1.18E+08	2.063091	1	oxidation-reduction process	Cytoplasm	-	14486	-	1.32E+06	5.03E+05	0.379746	8.79E+07	3.76E+07	0.427606	1	oxidation-reduction process	Cytoplasm	-	14500	-
GL50581_1982	GL50803_95162	Tenascin-like	5.28E+05	7.63E+05	1.446031	2.97E+07	4.77E+07	1.607024	2	protein binding	Cytoplasm	0.99	61197	0	1.95E+05	1.61E+05	0.826049	1.62E+07	1.28E+07	0.79473	3	protein binding	Cytoplasm	0.98	61787	0
GL50581_4057	GL50803_10330	Tenascin precursor	2.57E+05	5.28E+05	2.056165	4.66E+06	2.02E+07	4.342293	3	protein binding	Cytoplasm	0.99	27781	0	2.04E+05	2.47E+05	1.212092	7.37E+06	1.70E+07	2.304388	2	protein binding	Cytoplasm	0.98	28090	0
GL50581_438	GL50803_16468	Cathepsin B precursor	1.82E+05	3.56E+05	1.956461	9.63E+06	1.79E+07	1.861309	4	proteolysis	Cytoplasm	0.78	34066	0	2.28E+04	2.23E+05	9.797395	1.34E+06	1.11E+07	8.321123	4	proteolysis	Cytoplasm	0.86	33951	0
GL50581_4316	GL50803_8687	Tenascin precursor	1.02E+05	2.44E+05	2.396286	6.22E+06	1.10E+07	1.770526	5	protein binding	Cytoplasm	0.98	53521	0	1.41E+03	2.27E+05	160.408	3.77E+05	7.17E+06	19.04955	5	protein binding	Cytoplasm	0.99	53969	0
GL50581_2767	GL50803_5258	hypothetical protein	8.82E+04	9.75E+04	1.10593	3.93E+06	1.09E+07	2.780918	6	-	Undetermined	-	11335	-	-	2.16E+04	-	3.42E+06	1.99E+06	0.582717	9	-	Undetermined	-	11782	-
GL50581_3607	GL50803_8742	Extracellular nuclease, putative	2.18E+04	7.62E+04	3.499864	1.03E+06	4.57E+06	4.436193	7	nuclease activity	Cytoplasm	1	30603	0	-	6.28E+03	-	-	4.06E+05	-	13	nuclease activity	Cytoplasm	0.99	30781	0
GL50581_3575	GL50803_16477	Tenascin-37	1.63E+04	6.48E+04	3.979612	8.35E+05	4.03E+06	4.830956	8	protein binding	Cytoplasm	0.99	62017	0	4.67E+03	4.52E+03	0.968303	2.46E+05	2.96E+05	1.201909	14	protein binding	Cytoplasm	0.95	61934	0
GL50581_2036	GL50803_15564	Cathepsin B precursor	-	2.51E+04	-	1.10E+06	3.95E+06	3.589608	9	proteolysis	Cytoplasm	1	33027	0	8.92E+03	1.57E+04	1.759113	3.58E+05	9.63E+05	2.69212	10	proteolysis	Cytoplasm	-	36985	-
GL50581_3484	GL50803_2661	Cyclin-dependent Kinases regulatory subunit	1.82E+04	3.28E+04	1.798703	1.14E+06	3.20E+06	2.803062	10	cell cycle	Cytoplasm	-	10358	-	-	6.93E+03	-	1.17E+06	4.18E+05	0.358384	12	cell cycle	Cytoplasm	-	10358	-
GL50581_4180	GL50803_113038	Tenascin (GS isolate) / hypothetical protein (WB isolate)	1.81E+04	6.19E+04	3.416486	1.20E+06	3.14E+06	2.620931	11	protein binding	Cytoplasm	1	60692	0	-	4.06E+02	-	-	3.56E+06	-	6	calcium ion binding / protein binding	Undetermined	1	60841	0
GL50581_727	GL50803_7715	High cysteine membrane protein Group 1	-	4.58E+04	-	-	2.54E+06	-	12	protein binding	Plasma membrane	0.99	73115	1	-	1.14E+03	-	-	6.61E+04	-	15	-	Plasma membrane	0.79	72830	1
GL50581_352	GL50803_16522	hypothetical protein	-	8.89E+03	-	1.15E+06	2.21E+06	1.928833	13	-	Undetermined	-	16921	-	3.92E+04	2.47E+04	0.629109	4.85E+06	2.53E+06	0.521377	7	-	Undetermined	-	16666	-
GL50581_2622	GL50803_12063	Variant-specific surface protein	5.18E+03	2.55E+04	4.918394	3.62E+05	1.94E+06	5.354665	14	-	Plasma membrane	1	43681	0	6.22E+03	5.42E+03	0.871456	3.45E+05	4.46E+05	1.293367	11	-	Plasma membrane	1	43782	0
GL50581_2318	GL50803_17516	Cathepsin B precursor	-	4.30E+03	-	-	7.81E+05	-	15	cysteine-type peptidase activity	Cytoplasm	1	34583	0	1.32E+04	4.96E+04	3.760729	3.36E+05	2.04E+06	6.066772	8	proteolysis	Cytoplasm	1	34681	0

[a] Signal peptide

[b] Molecular Weight

[c] Transmembrane Domain

Table 5. 5: List of the 31 proteins identified via Orbitrap and QeXactive MS as proteins most likely secreted by *Giardia* GS isolate trophozoites.
23 proteins are annotated (shown in red) and 8 are hypothetical proteins (shown in blue). 6 proteins are lineage-specific, and 2 are orthologs to proteins in sub-assemblage A2, DH isolate. The 15 proteins identified as conserved between the two isolates are highlighted in grey. Only proteins identified via both techniques were considered as secreted and shown in this table. The two secretion profiles obtained for GS isolate via Orbitrap or QeXactive are shown in Appendix IX A and B respectively. Proteins are ranked according to QeXactive SP expression from most to least abundant.

Assemblage B (GS isolate)													Assemblage A syntenic orthologs	
		Orbitrap			QeXactive									
Protein Accesion Number	Protein Description	P Expression-iBAQ	SP Expression-iBAQ	SP/P Ratio	P Expression-iBAQ	SP Expression-iBAQ	SP/P Ratio	GO function	GO location	SignalP[a]	MW[b] (Da)	TM[c]	Protein Accesion Number	Description
GL50581_4133	Hypothetical protein	8.02E+05	1.25E+06	1.5574566	5.71E+07	1.18E+08	2.06309	oxidation-reduction process	Cytoplasm	-	14486	-	GL50803_5810	hypothetical protein
GL50581_1982	Tenascin-like protein	5.28E+05	7.63E+05	1.4460306	2.97E+07	4.77E+07	1.60702	protein binding	Cytoplasm	0.99	61197	0	GL50803_95162	Tenascin-like
GL50581_2954	VSP	1.05E+05	2.29E+05	2.1795107	1.18E+07	2.25E+07	1.90956	-	Plasma membrane	-	17746	-	-	-
GL50581_4057	Tenascin precursor	2.57E+05	5.28E+05	2.0561652	4.66E+06	2.02E+07	4.34229	protein binding	Cytoplasm	0.99	27781	0	GL50803_10330	Tenascin precursor
GL50581_3909	Hypothetical protein	5.29E+04	1.07E+05	2.0237563	8.91E+06	1.85E+07	2.07789	-	Undetermined	-	11343	-	GL50803_1875	Hypothetical protein
GL50581_438	Cathepsin B precursor	1.82E+05	3.56E+05	1.9564608	9.63E+06	1.79E+07	1.86131	proteolysis	Cytoplasm	0.78	34066	0	GL50803_16468	Cathepsin B precursor
GL50581_1838	Hypothetical protein	8.98E+04	9.86E+04	1.0988986	9.86E+06	1.68E+07	1.70234	-	Undetermined	-	10368	-	GL50803_27918	Hypothetical protein
GL50581_4316	Tenascin precursor	1.02E+05	2.44E+05	2.3962861	6.22E+06	1.10E+07	1.77053	protein binding	Cytoplasm	0.98	53521	0	GL50803_8687	Tenascin precursor
GL50581_2767	Hypothetical protein	8.82E+04	9.75E+04	1.1059298	3.93E+06	1.09E+07	2.78092	-	Undetermined	-	11335	-	GL50803_5258	Hypothetical protein
GL50581_1626	A-type flavoprotein lateral transfer candidate	9.22E+04	1.06E+05	1.1541891	4.35E+06	8.70E+06	2.00174	oxidoreductase activity / hydrolase activity	Cytoplasm	-	46451	-	GL50803_10358	A-type flavoprotein lateral transfer candidate
GL50581_3607	Extracellular nuclease, putative	2.18E+04	7.62E+04	3.4998644	1.03E+06	4.57E+06	4.43619	nuclease activity	Cytoplasm	1	30603	0	GL50803_8742	Extracellular nuclease, putative
GL50581_3575	Tenascin-37	1.63E+04	6.48E+04	3.9796118	8.35E+05	4.03E+06	4.83096	protein binding	Cytoplasm	0.99	62017	0	GL50803_16477	Tenascin-37
GL50581_2036	Cathepsin B precursor	-	2.51E+04	-	1.10E+06	3.95E+06	3.58961	proteolysis	Cytoplasm	1	33027	0	GL50803_15564	Cathepsin B precursor
GL50581_3484	Cyclin-dependent kinases regulatory subunit	1.82E+04	3.28E+04	1.7987035	1.14E+06	3.20E+06	2.80306	cell cycle	Cytoplasm	-	10358	-	GL50803_2661	Cyclin-dependent Kinases regulatory subunit
GL50581_4180	Tenascin precursor	1.81E+04	6.19E+04	3.4164861	1.20E+06	3.14E+06	2.62093	protein binding	Cytoplasm	1	60692	0	GL50803_113038	hypothetical protein
GL50581_727	High cysteine membrane protein Group 1	-	4.58E+04	-	-	2.54E+06	-	protein binding	Plasma membrane	0.99	73115	1	GL50803_7715	High cysteine membrane protein Group 1
GL50581_352	Hypothetical protein	-	8.89E+03	-	1.15E+06	2.21E+06	1.92883	-	Undetermined	-	16921	-	GL50803_16522	hypothetical protein
GL50581_2622	Variant-specific surface protein	5.18E+03	2.55E+04	4.9183944	3.62E+05	1.94E+06	5.35467	-	Plasma membrane	1	43681	0	GL50803_12063	Variant-specific surface protein
GL50581_1	VSP	1.15E+04	2.76E+04	2.3954753	2.36E+05	1.46E+06	6.16205	-	Plasma membrane	1	41559	1	-	-
GL50581_3056	FixW protein, putative	-	9.86E+03	-	1.92E+05	1.24E+06	6.46307	oxidation-reduction process	Cytoplasm	0.06	23091	0	GL50803_23888	FixW protein, putative
GL50581_3401	VSP	4.40E+03	1.10E+04	2.501522	4.93E+05	1.22E+06	2.48425	-	Plasma membrane	1	38355	0	-	-
GL50581_4162	Hypothetical protein	6.28E+03	8.70E+03	1.3853992	4.97E+05	8.75E+05	1.76017	protein folding	Cytoplasm	-	13497	-	DHA2_152684	Hypothetical protein
GL50581_2318	Cathepsin B precursor	-	4.30E+03	-	-	7.81E+05	-	cysteine-type peptidase activity	Cytoplasm	1	34583	0	GL50803_17516	Cathepsin B precursor
GL50581_572	VSP	1.48E+03	4.04E+03	2.7354264	6.50E+04	5.39E+05	8.2832	protein binding	Plasma membrane	-	67208	-	-	-
GL50581_3818	VSP	-	1.06E+04	-	7.85E+04	5.06E+05	6.44979	protein binding	Plasma membrane	-	61825	-	-	-
GL50581_1902	VSP	1.11E+05	1.62E+05	1.4581109	1.58E+05	4.27E+05	2.70029	protein phosphorylation cell adhesion /	Plasma membrane	-	96279	-	DHA2_150066	VSP
GL50581_3457	High cysteine protein	-	1.17E+04	-	3.43E+04	3.48E+05	10.1259	metalloendopeptidase activity	Plasma membrane	-	74175	-	-	-
GL50581_3748	Hypothetical protein	9.08E+02	9.78E+02	1.0777412	7.75E+04	3.28E+05	4.23873	-	Undetermined	-	35528	-	GL50803_10527	Hypothetical protein
GL50581_2090	Hypothetical protein	-	2.18E+03	-	-	2.69E+05	-	-	Undetermined	-	33903	-	GL50803_8423	Hypothetical protein
GL50581_1340	Phosphoglycolate phosphatase	-	2.52E+03	-	-	1.46E+05	-	metabolic process	Cytoplasm	-	25403	-	GL50803_14841	Phosphoglycolate phosphatase
GL50581_3210	Sentrin	-	1.08E+03	-	-	3.67E+04	-	protein binding	Undetermined	-	11495	-	GL50803_7760	Sentrin

[a] Signal peptide
[b] Molecular Weight
[c] Transmembrane Domain

Table 5. 6: List of the 44 proteins identified via Orbitrap and QeXactive MS as proteins most likely secreted by *Giardia* WB isolate trophozoites. 37 proteins are annotated (shown in red) and 7 are hypothetical proteins (shown in blue). 10 proteins are lineage-specific. The 15 proteins identified as conserved between the two isolates are highlighted in grey. Only proteins identified via both techniques were considered as secreted and shown in this table. The two secretion profiles obtained for WB isolate via Orbitrap or QeXactive are shown in Appendix IX C and D respectively. Proteins are ranked according to QeXactive SP expression from most to least abundant.

Assemblage A (WB isolate)												Assemblage B syntenic orthologs	
		Orbitrap			QeXactive								
Protein Accesion Number	Protein Description	P Expression-iBAQ	SP Expression-iBAQ	SP/P Ratio	P Expression-iBAQ	SP Expression-iBAQ	SP/P Ratio	GO function	GO location	SignalP[a]	MW[b] (Da)	TM[c]	Accession	Description
GL50803_33279	VSP	6.18E+06	3.61E+06	0.58459	3.68E+08	2.98E+08	0.80993	protein binding	Plasma membrane	0.98	75981	1	-	-
GL50803_5810	Hypothetical protein	1.32E+06	5.03E+05	0.37975	8.79E+07	3.76E+07	0.42761	oxidation-reduction process	Cytoplasm	-	14500	-	GL50581_4133	Hypothetical protein
GL50803_17163	Peptidyl-prolyl cis-trans isomerase B precursor	9.04E+05	2.91E+05	0.32235	6.89E+07	3.03E+07	0.43974	protein folding	Cytoplasm	-	18044	-	GL50581_1019	Peptidyl-prolyl cis-trans isomerase B precursor
GL50803_10330	Tenascin precursor	2.04E+05	2.47E+05	1.21209	7.37E+06	1.70E+07	2.30439	protein binding	Cytoplasm	0.98	28090	0	GL50581_4057	Tenascin precursor
GL50803_16779	Cathepsin B precursor	1.03E+05	3.58E+05	3.48645	6.73E+06	1.56E+07	2.31252	proteolysis	Cytoplasm	1	32761	0	GL50581_78	Cathepsin B precursor
GL50803_16322	Neurogenic locus Notch protein precursor	4.46E+05	2.97E+05	0.66533	2.12E+07	1.46E+07	0.69085	calcium ion binding / protein binding	Cytoplasm	0.99	63070	0	GL50581_31	Neurogenic locus Notch protein precursor
GL50803_95162	Tenascin-like	1.95E+05	1.61E+05	0.82605	1.62E+07	1.28E+07	0.79473	protein binding	Cytoplasm	0.98	61787	0	GL50581_1982	Tenascin-like
GL50803_137740	VSP-3	2.57E+05	2.61E+05	1.01435	1.47E+07	1.18E+07	0.80039	protein binding	Plasma membrane	-	46480	-	-	-
GL50803_16468	Cathepsin B precursor	2.28E+04	2.23E+05	9.79739	1.34E+06	1.11E+07	8.32112	proteolysis	Cytoplasm	0.86	33951	0	GL50581_438	Cathepsin B precursor
GL50803_13272	hypothetical protein	2.71E+05	9.06E+04	0.33379	1.83E+07	9.33E+06	0.50938	-	Undetermined	-	40412	-	GL50581_3472	Inosine-uridine nucleoside N-ribohydrolase
GL50803_8687	Tenascin precursor	1.41E+03	2.27E+05	160.408	3.77E+05	7.17E+06	19.0496	protein binding	Cytoplasm	0.99	53969	0	GL50581_4316	Tenascin precursor
GL50803_14019	Cathepsin B precursor	4.40E+05	2.53E+05	0.57417	1.55E+07	6.56E+06	0.42377	proteolysis	Cytoplasm	1	33056	0	GSB_14019	Cathepsin B-like cysteine proteinase
GL50803_480	Translation initiation inhibitor	8.24E+04	5.15E+04	0.62574	9.47E+06	4.08E+06	0.43038	deaminase activity	Cytoplasm	-	12893	-	GL50581_4017	Translation initiation inhibitor
GL50803_113038	Hypothetical protein	-	4.06E+02	-	-	3.56E+06	-	calcium ion binding / protein binding	Undetermined	1	60841	0	GL50581_4180	Tenascin precursor
GL50803_8227	ATP-binding cassette protein 5	2.80E+04	8.18E+04	2.92188	2.93E+06	2.89E+06	0.98546	ATPase activity	Cytoplasm	-	118852	-	GL50581_3399	ATP-binding cassette protein 5
GL50803_16522	Hypothetical protein	3.92E+04	2.47E+04	0.62911	4.85E+06	2.53E+06	0.52138	-	Undetermined	-	16666	-	GL50581_352	Hypothetical protein
GL50803_112312	Elongation factor 1-alpha	-	4.11E+04	-	-	2.22E+06	-	translational elongation	Cytoplasm	-	49105	-	-	-
GL50803_17516	Cathepsin B precursor	1.32E+04	4.96E+04	3.76073	3.36E+05	2.04E+06	6.06677	proteolysis	Cytoplasm	1	34681	0	GL50581_2318	Cathepsin B precursor
GL50803_5258	Hypothetical protein	-	2.16E+04	-	3.42E+06	1.99E+06	0.58272	-	Undetermined	-	11782	-	GL50581_2767	Hypothetical protein
GL50803_114815	Tenascin precursor	1.69E+04	1.17E+05	6.90495	1.69E+06	1.76E+06	1.04278	calcium ion binding / protein binding	Cytoplasm	0.89	63018	0	-	-
GL50803_137613	VSP with INR	4.51E+04	2.35E+04	0.52085	2.41E+06	1.36E+06	0.56346	-	Plasma membrane	0.99	40536	1	-	-

End of Table 5.6

GL50803_103454	High cysteine membrane protein Group 1	1.98E+04	1.99E+04	1.00358	2.11E+06	1.33E+06	0.62925	-	Plasma membrane	1	79409	1	GSB_153702	Variant-specific surface protein
GL50803_16833	Tenascin-like	1.18E+04	1.92E+04	1.6281	1.16E+06	1.08E+06	0.93724	protein binding	Cytoplasm	0.99	64667	0	GL50581_1227	Tenascin-like protein
GL50803_15564	Cathepsin B precursor	8.92E+03	1.57E+04	1.75911	3.58E+05	9.63E+05	2.69212	proteolysis	Cytoplasm	-	36985	-	GL50581_2036	Cathepsin B precursor
GL50803_10217	Cathepsin B precursor	-	3.97E+04	-	-	7.56E+05	-	proteolysis	Cytoplasm	1	33478	0	GSB_155190	Cathepsin B-like cysteine proteinase
GL50803_15250	High cysteine membrane protein Group 6	5.10E+03	2.26E+04	4.44197	1.69E+05	5.09E+05	3.0102	protein binding	Plasma membrane	1	66839	1	GSB_154625	Variant-specific surface protein
GL50803_12063	Variant-specific surface protein	6.22E+03	5.42E+03	0.87146	3.45E+05	4.46E+05	1.29337	-	Plasma membrane	1	43782	0	GL50581_2622	Variant-specific surface protein
GL50803_2661	Cyclin-dependent Kinases regulatory subunit	-	6.93E+03	-	1.17E+06	4.18E+05	0.35838	cell cycle	Cytoplasm	-	10358	-	GL50581_3484	Cyclin-dependent Kinases regulatory subunit
GL50803_8742	Extracellular nuclease, putative	-	6.28E+03	-	-	4.06E+05	-	nuclease activity	Cytoplasm	0.99	30781	0	GL50581_3607	Extracellular nuclease, putative
GL50803_16477	Tenascin-37	4.67E+03	4.52E+03	0.9683	2.46E+05	2.96E+05	1.20191	protein binding	Cytoplasm	0.95	61934	0	GL50581_3575	Tenascin-37
GL50803_17476	CXC-rich protein	5.50E+03	6.12E+03	1.11393	4.80E+05	2.83E+05	0.58945	protein binding	Cytoplasm	1	224253	1	GL50581_4509	CXC-rich protein
GL50803_14573	Tenascin-X	9.85E+02	5.10E+03	5.17242	3.05E+05	2.79E+05	0.91504	protein binding		1	53351	0	GL50581_1475	Tenascin-X
GL50803_14546	Phosphoglycolate phosphatase	-	1.98E+03	-	-	2.45E+05	-	phosphoglycolate phosphatase activity	Cytoplasm	-	25277	-	GL50581_1485	Phosphoglycolate phosphatase
GL50803_112135	High cysteine membrane protein VSP-like	-	1.65E+03	-	-	1.40E+05	-	protein binding	Plasma membrane	1	120767	1	GSB_152985	Variant-specific surface protein
GL50803_101074	VSP with INR	4.59E+03	1.78E+03	0.38754	4.04E+05	1.40E+05	0.34619	-	Plasma membrane	1	76764	1	-	-
GL50803_10238	Hypothetical protein	3.91E+02	8.70E+03	22.2276	-	1.21E+05	-	carbohydrate metabolic process	Cytoplasm	1	113230	1	-	-
GL50803_5180	Peptide methionine sulfoxide reductase msrB	-	5.35E+02	-	1.24E+05	8.80E+04	0.70952	protein repair / response to oxidative stress/peptide-methionine (R or S)-S-oxide reductase activity	Cytoplasm	-	17407	-	GL50581_3084	Peptide methionine sulfoxide reductase msrB
GL50803_7715	High cysteine membrane protein Group 1	-	1.14E+03	-	-	6.61E+04	-	-	Plasma membrane	0.79	72830	1	GL50581_727	High cysteine membrane protein Group 1
GL50803_16936	High cysteine membrane protein EGF-like	-	7.38E+02	-	-	4.82E+04	-	protein binding	Plasma membrane	0.96	162611	1	GL50581_4085	High cysteine membrane protein EGF-like
GL50803_114674	Hypothetical protein	-	7.63E+02	-	2.98E+04	1.40E+04	0.46856	protein binding	Undetermined	0.99	73024	1	-	-
GL50803_137617	VSP -77	8.33E+02	1.54E+03	1.85102	2.07E+04	1.09E+04	0.52804	protein binding	Plasma membrane	1	75503	1	-	-
GL50803_116477	VSP-149	-	1.36E+02	-	-	7.98E+03	-	protein binding	Plasma membrane	1	76114	1	GSB_153909	Variant-specific surface protein
GL50803_14586	VSP with INR	-	8.03E+01	-	1.95E+04	7.81E+03	0.40055	protein binding	Plasma membrane	1	75531	1	-	-
GL50803_113319	High cysteine membrane protein TMK-like	-	1.34E+01	-	-	3.95E+03	-	cell adhesion	Plasma membrane	0.99	265709	1	GSB_152415 / GSB_152416	Variant-specific surface protein

[a] Signal peptide
[b] Molecular Weight
[c] Transmembrane Domain

results suggest that *Giardia* GS and WB isolates secrete mostly Tenascins, VSPs, Cathepsin B, HCMPs and hypothetical proteins. This also suggests the potential key role of these proteins in the early host-pathogen interactions; and particularly for the Cathepsin B, three Tenascins and the hypothetical proteins that are closely related.

5.4/ Discussion

Giardia trophozoites have been shown to release a Cathepsin B-like protease, Enolase, ADI and OCT upon interaction with host intestinal epithelia (Cotton et al., 2014, Ringqvist et al., 2008, Roxstrom-Lindquist et al., 2006). However, which proteins were actively secreted by the parasite has not previously been explored. In this study, *Giardia* trophozoites were shown to secrete primarily Cathepsin B proteins, High cysteine membrane proteins (including Tenascins and VSPs) and Hypothetical proteins. In each case the encoding loci are genetically polymorphic between isolates and a number of lineage specific features are present which could be exploited for strain specific identification and which could be directly associated with virulence.

5.4.1/ *Giardia* Cathepsins B family

In this study, the *Giardia* Cathepsin B (*G*CATB) family was shown to contain secreted and non-secreted trophozoite expressed proteins. Fourteen *G*CATBs were identified within GS and WB proteomes and secretomes, their coding genes were retrieved from GiardiaDB and aligned on ClustalW to assess their sequence similarity scores. A phylogenetic tree was also built by maximum likelihood approach using MEGA software (v 6.06) to determine their relationship (Table 5.7, Fig 5.6). Ten of the proteins were shown to be orthologs; four pairs of orthologs were part of the conserved secreted proteins. The last pair of orthologs and two GS-specific *G*CATBs were identified in the proteomes as non-secreted (Fig 5.6).

The two non-secreted GS lineage-specific *G*CATBs appeared to be near identical duplicates of the same gene; however, whether they are tandem repeat as might be expected is not clear as they do not seem to be located on the same contig of the genome assembly and their chromosome locations are unassigned. So,

this may bear further consideration; particularly as it represents a substantial and recent cladistic feature which could be used to discriminate the assemblage B lineage. Four pairs of *G*CATB orthologs were conserved secreted factors and comprised closely related orthologs (Fig 5.6). Two *G*CATBs were secreted by WB isolate only; one of these was closely related to one of the conserved *G*CATBs orthologous pairs and may represent another example of gene duplication and divergence. The other WB secreted *G*CATB appeared to be most orthologous to the two near identical non-secreted GS *G*CATBs previously mentionned.

Cathepsins B are Cysteine proteases that are involved in several *Giardia* pathogenicity processes such as encystation, excystation, attachment to the host intestinal epithelia and the degradation of mucin, intestinal IL-8 and degradation of intracellular junction (Table 5.1) (Cotton et al., 2014, DuBois et al., 2008, DuBois et al., 2006, Rodriguez-Fuentes et al., 2006, Ward et al., 1997, Paget and James, 1994). Cysteine proteases tend to differ not in substrate specificity but may in terms of activity, compartmentalisation, stability, environmental resistance, conditional optima, partner proteins and so forth. That several *G*CATBs appear to be secreted by growing trophozoites from both *Giardia* isolates suggests that different *G*CATBs may be responsible for different functions and perhaps even different pathogenic effects. Moreover, that ortholog proteins remain closely related suggests preservation of function between isolates. Interestingly, two *G*CATBs shared between the two isolates were only identified in WB secretome. The fact that these two proteins appear not to be secreted by GS isolate trophozoites suggests a potential mechanism by which the two isolates might vary in terms of pathogenicity, and also that some environmental changes may be needed to trigger the secretion of the non-secreted *G*CATB orthologs in GS isolate. Finally, the genetic heterogeneity of these proteins between assemblages is generally high compares with other proteins (50 to 80 % similarities, Table 5.7). As *G*CATBs are likely to be the primary mediators of pathology, it is to be expected that intra-assemblage variation in sequence will be evident and that this may vary with the virulence characteristics of the strains. Further testing of more and less virulent clinical isolates is warranted to validate these *G*CATBs as markers for upstream use in the characterisation and discrimination of *Giardia* strains.

Table 5.7: Sequence similarities of the genes coding for the Cathepsin B proteins identified in GS and WB isolate proteomes via Orbitrap and QeXactive MS. Proteins identified as secreted in both isolate are represented in blue, proteins identified as secreted in WB isolate only are represented in orange, and proteins present in proteomes but not as secreted factors are shown in black. Orthologs are shown in bold. All the GL50581_ accession numbers represent GS isolate proteins and all the GL50803_ represent WB isolate proteins. Similarity scores were assessed via ClustalW (Appendix X).

Giardia Isolate	Accession number	Length (bp)	*Giardia* Isolate target	Name	Length (bp)	Score (%)
GS	GL50581_2036	921	**WB**	**GL50803_15564**	**999**	**78.18**
			GS	GL50581_2318	930	60.8
			WB	GL50803_17516	927	56.03
			GS	GL50581_438	918	54.79
			WB	GL50803_16468	918	57.73
			GS	GL50581_2946	891	61.39
			WB	GL50803_16160	1083	61.45
			GS	GL50581_78	897	64.88
			WB	GL50803_16779	897	63.1
			WB	GL50803_14019	903	63.9
			WB	GL50803_10217	912	61.18
			GS	GL50581_159	912	58.22
			GS	GL50581_3619	912	58.66
WB	GL50803_15564	999	GS	GL50581_2318	930	55.81
			WB	GL50803_17516	927	58.36
			GS	GL50581_438	918	56.64
			WB	GL50803_16468	918	57.3
			GS	GL50581_2946	891	59.6
			WB	GL50803_16160	1083	61.36
			GS	GL50581_78	897	64.88
			WB	GL50803_16779	897	64.44
			WB	GL50803_14019	903	66
			WB	GL50803_10217	912	60.42
			GS	GL50581_159	912	61.18
			GS	GL50581_3619	912	61.07
GS	GL50581_2318	930	**WB**	**GL50803_17516**	**927**	**77.24**
			GS	GL50581_438	918	53.27
			WB	GL50803_16468	918	53.49
			GS	GL50581_2946	891	58.25
			WB	GL50803_16160	1083	56.99
			GS	GL50581_78	897	58.86
			WB	GL50803_16779	897	54.74
			WB	GL50803_14019	903	59.14
			WB	GL50803_10217	912	61.29
			GS	GL50581_159	912	60.86
			GS	GL50581_3619	912	60.64
WB	GL50803_17516	927	GS	GL50581_438	918	55.66
			WB	GL50803_16468	918	55.12
			GS	GL50581_2946	891	56.68
			WB	GL50803_16160	1083	56.85
			GS	GL50581_78	897	59.09
			WB	GL50803_16779	897	58.19
			WB	GL50803_14019	903	58.91
			WB	GL50803_10217	912	59.54
			GS	GL50581_159	912	60.53
			GS	GL50581_3619	912	60.86

End of Table 5.7:

GS	GL50581_438	918	**WB**	**GL50803_16468**	**918**	**79.74**
			GS	GL50581_2946	891	54.88
			WB	GL50803_16160	1083	57.3
			GS	GL50581_78	897	60.42
			WB	GL50803_16779	897	60.09
			WB	GL50803_14019	903	59.69
			WB	GL50803_10217	912	56.14
			GS	GL50581_159	912	58.33
			GS	GL50581_3619	912	57.79
WB	GL50803_16468	918	GS	GL50581_2946	891	58.81
			WB	GL50803_16160	1083	57.95
			GS	GL50581_78	897	58.08
			WB	GL50803_16779	897	57.19
			WB	GL50803_14019	903	58.8
			WB	GL50803_10217	912	55.7
			GS	GL50581_159	912	57.46
			GS	GL50581_3619	912	57.68
GS	GL50581_2946	891	**WB**	**GL50803_16160**	**1083**	**83.73**
			GS	GL50581_78	897	64.31
			WB	GL50803_16779	897	63.86
			WB	GL50803_14019	903	62.85
			WB	GL50803_10217	912	56.9
			GS	GL50581_159	912	58.92
			GS	GL50581_3619	912	58.47
WB	GL50803_16160	1083	GS	GL50581_78	897	62.32
			WB	GL50803_16779	897	65.44
			WB	GL50803_14019	903	65.45
			WB	GL50803_10217	912	57.35
			GS	GL50581_159	912	59.98
			GS	GL50581_3619	912	58.55
GS	GL50581_78	897	**WB**	**GL50803_16779**	**897**	**84.84**
			WB	GL50803_14019	903	81.16
			WB	GL50803_10217	912	63.1
			GS	GL50581_159	912	62.21
			GS	GL50581_3619	912	61.87
WB	GL50803_16779	897	WB	GL50803_14019	903	84.28
			WB	GL50803_10217	912	64.44
			GS	GL50581_159	912	61.65
			GS	GL50581_3619	912	62.65
WB	GL50803_14019	903	WB	GL50803_10217	912	62.46
			GS	GL50581_159	912	61.13
			GS	GL50581_3619	912	61.13
WB	GL50803_10217	912	GS	GL50581_159	912	76.86
			GS	GL50581_3619	912	76.75
GS	GL50581_159	912	GS	GL50581_3619	912	99.45

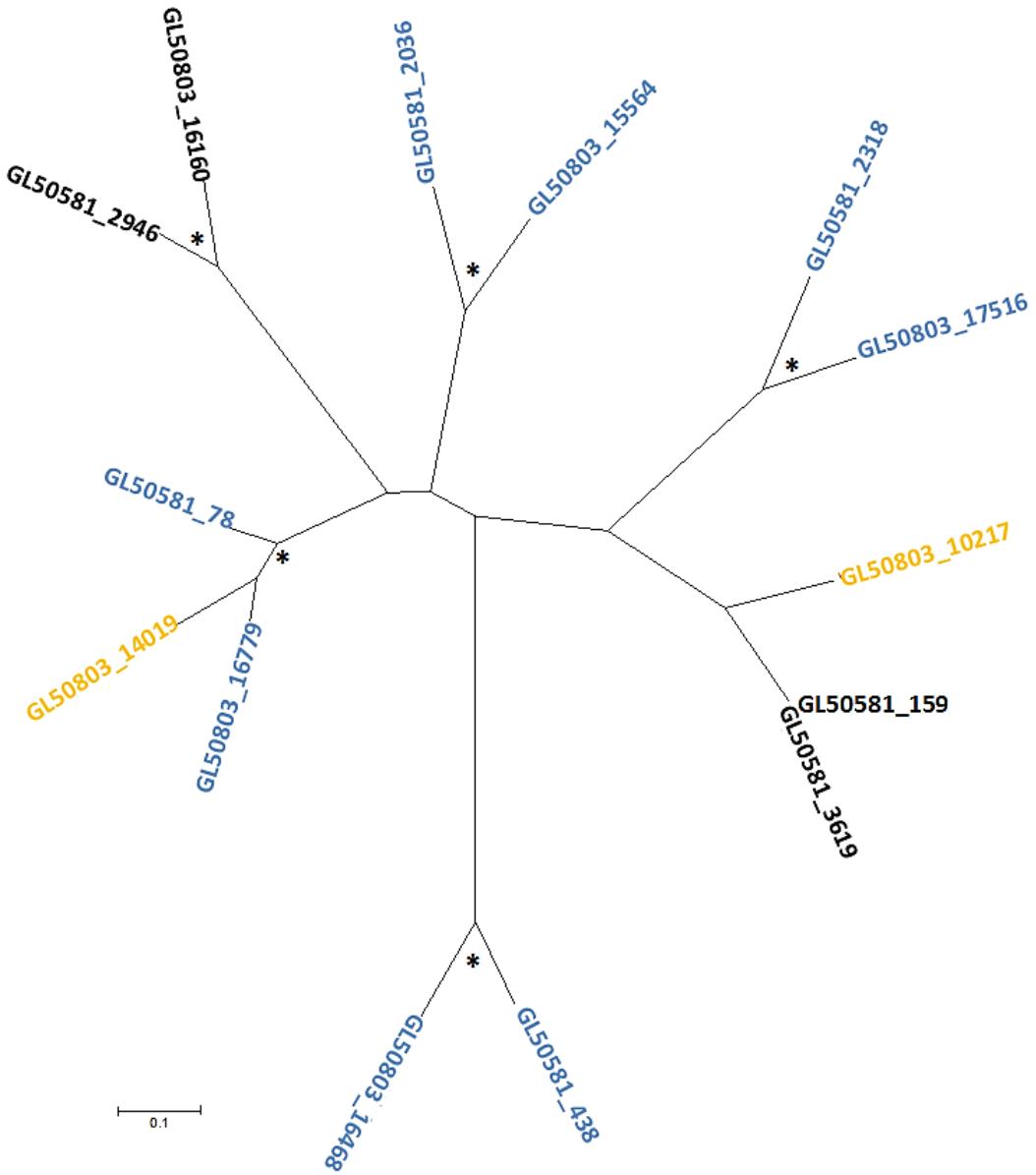


Fig 5. 6: Phylogenetic tree of the Cathepsins B identified in GS and WB isolate proteomes. Gene sequences were obtained from GiardiaDB, aligned using ClustalW. The tree was realised by a Maximum likelihood approach using MEGA software (v 6.06) (Appendix X). All the Cathepsin B proteins identified in the four datasets were included in the analysis. Proteins identified as secreted in both isolate are highlighted in blue, proteins identified as secreted only in WB isolate dataset are highlighted in orange, and proteins identified in proteomes but not as secreted factors are highlighted in black. Ortholog proteins are shown by *****.
All the GL50581_ accession numbers represent GS isolate proteins and all the GL50803_ represent WB isolate proteins.

5.4.2/ High cysteine membrane protein superfamily

Tenascins, VSPs and HCMPs together form the largest group of proteins enriched in the *Giardia* SPs. These three protein families have been shown to be rich in cysteine residues and to be either present at the surface of the parasite or as extracellular matrix (Bedore et al., 2014, Ankarklev et al., 2010); genes coding for these proteins were compared for similarities (Table 5.8). Interestingly, when aligned and analysed phylogenetically, a clear pattern emerges with VSPs segregated form Tenascins and HCMPS spanning the two groups (Fig 5.7).

VSPs are well characterised surface glycoproteins with transmembrane domains TM, which are expressed one at a time by *Giardia* trophozoites through RNAi regulated mechanism. They are quintessential virulence factors responsible for antigenic variations as well as cell signalling, immune evasion, and potentially in the protection against luminal proteases, oxygen and free-radicals (Table 5.1) (Adam et al., 2010, Touz et al., 2005, Nash, 2002, Singer et al., 2001). VSPs are hypervariable by nature and thus, it is to be expected that they do not form orthologous pairs. This was the case for most of the VSPs observed in this study; intriguingly though, a few proteins annotated VSPs were conserved between isolates. This suggests that they are not actually VSPs and would not be subject to "one at a time" control expression but are actually misannotated HCMPs which may have a conserved function in both GS and WB isolates.

HCMPs are an enigmatic group of proteins with few associated functional studies. They may protect trophozoites against proteolysis (Table 5.1) (Davids et al., 2006, Nash, 2002). In other eukaryotes, and in keeping with the reductive potential of cysteine itself, HCMPs have also been shown to protect against oxidative damage which suggests that may have the same function in *Giardia* trophozoites (Requejo et al., 2010). In *Giardia*, it appears that one lineage of HCMPs has given rise to the VSPs whilst another has given rise to a group with high homology to mammalian Tenascins.

Tenascins are characterised by the presence of epidermal growth factor (EGF-) repeats and are able to act as weak ligands for EGF receptors. Mammalian Tenascins are extracellular matrix proteins which modulate cell adhesion and

migration (Chiquet-Ehrismann and Chiquet, 2003). They appear to have evolved from a group of proteins specific to vertebrates, presumably co-evolving with the EGF receptor (Chiquet-Ehrismann and Chiquet, 2003). Thus, the presence of homologous proteins in *Giardia* evolving independently from HCMPs is a clear example of the kind of convergent evolution best described as molecular MIMICRY (Oldstone, 1998). The role of *Giardia*-secreted Tenascins has yet to be established, but they are defined by the presence of EGF repeats and it is to be anticipated that they will act similarly as weak ligands of the EGF receptor. The fact that several types of Tenascins are secreted by *Giardia* trophozoites suggests that Tenascins may have one or more functions in the host-pathogen interactions during *Giardia* infection. While VSPs are hypervariable and do not have orthologs, Tenascins and HCMPs showed low similarities between the orthologs that could be identified suggesting that all these proteins were highly variable (Table 5.8). The phylogenetic tree indicated that the conserved secreted proteins were more closely related to their orthologs than the other proteins though they remain more variable than most other orthologs (Fig 5.7). Interestingly, at least one syntenic set of Tenascin orthologs showed more divergence from its ortholog than to other Tenascins.

It is likely that rather like the extended VSG family of *Trypanosoma brucei*, the HCMP superfamily of *Giardia* contains multifunctional proteins with a range of related contingency functions from immune evasion to resisting environmental pressure to virulence and pathogenicity. The degree of variation between orthologs in this group is the highest of any family and it is to be anticipated that they may be good markers for strain discrimination and even source tracking. As with the Cathepsin B family, the genes identified in this study warrant systematic investigation as markers against a diverse range of clinically virulent and avirulent *Giardia* strains.

5.4.3/ Secreted formerly hypothetical proteins (SFHPs)

One other annotation of proteins formed a substantial group of secreted proteins, those encoded by reading frames annotated as hypothetical. Although this study provides evidence that they are indeed expressed and indeed secreted, there is no functional data associated with them and their roles are not clear.

Table 5.8: Sequence similarities of the genes coding for the Tenascins, High cysteine membrane proteins and VSPs identified as secreted in GS and WB isolate proteomes. Proteins identified in both isolate are represented in blue, proteins identified in WB isolate only are represented in orange, and proteins identified in GS isolate only are shown in green. Orthologs are shown in bold. All the GL50581_ accession numbers represent GS isolate proteins and all the GL50803_ represent WB isolate proteins. Similarity scores were assessed via ClustalW (Appendix XI). **a**: Tenascin, **b**: High cysteine membrane protein, **c**: VSP.

	Name	Length (bp)		Name	Length (bp)	Score (%)
b	GL50581_727	2058	b	**GL50803_7715**	**2049**	**70.28**
			b	GL50581_3457	2205	41.69
			b	GL50803_15250	1872	40.49
			b	GL50803_112135	3435	46.7
			b	GL50803_16936	4566	48.25
			b	GL50803_113319	7590	52.58
			b	GL50803_103454	2211	44.8
			a	GL50581_4057	771	49.81
			a	GL50803_10330	774	50.26
			a	GL50581_4316	1503	43.91
			a	GL50803_8687	1515	44.55
			a	GL50581_3575	1749	43.22
			a	GL50803_16477	1749	44.14
			a	GL50581_1982	1719	44.1
			a	GL50803_95162	1719	43.05
			a	GL50803_114815	1779	41.77
			a	GL50803_14573	1500	45.6
			c	GL50581_2622	1245	48.35
			c	GL50803_12063	1245	47.95
			c	GL50581_1902	2772	44.07
			c	GL50581_3818	1836	41.83
			c	GL50581_3401	1149	50.91
			c	GL50581_572	2022	38.97
			c	GL50581_1	1224	46.24
			c	GL50581_2954	546	55.68
			c	GL50803_33279	2247	40.62
			c	GL50803_137740	1356	46.83
			c	GL50803_137613	1185	48.1
			c	GL50803_101074	2220	40.62
			c	GL50803_137617	2205	41.01
			c	GL50803_116477	2247	44.12
			c	GL50803_14586	2202	41.84
b	GL50803_7715	2049	b	GL50581_3457	2205	41.63
			b	GL50803_15250	1872	40.12
			b	GL50803_112135	3435	46.8
			b	GL50803_16936	4566	47.73
			b	GL50803_113319	7590	53.59
			b	GL50803_103454	2211	45.49
			a	GL50581_4057	771	49.03
			a	GL50803_10330	774	49.35
			a	GL50581_4316	1503	44.24
			a	GL50803_8687	1515	43.56
			a	GL50581_3575	1749	43.4
			a	GL50803_16477	1749	42.88
			a	GL50581_1982	1719	42.82
			a	GL50803_95162	1719	43.51
			a	GL50803_114815	1779	42.55
			a	GL50803_14573	1500	43.87
			c	GL50581_2622	1245	47.07
			c	GL50803_12063	1245	44.74
			c	GL50581_1902	2772	44.7
			c	GL50581_3818	1836	43.19
			c	GL50581_3401	1149	50.13
			c	GL50581_572	2022	39.17
			c	GL50581_1	1224	46.9
			c	GL50581_2954	546	55.13
			c	GL50803_33279	2247	43.48
			c	GL50803_137740	1356	48.08
			c	GL50803_137613	1185	49.87
			c	GL50803_101074	2220	44.12
			c	GL50803_137617	2205	42.85
			c	GL50803_116477	2247	42.75
			c	GL50803_14586	2202	42.31
b	GL50581_3457	2205	b	GL50803_15250	1872	43.59
			b	GL50803_112135	3435	48.57
			b	GL50803_16936	4566	48.84
			b	GL50803_113319	7590	76.83
			b	GL50803_103454	2211	43.08
			a	GL50581_4057	771	47.47
			a	GL50803_10330	774	48.97
			a	GL50581_4316	1503	45.24
			a	GL50803_8687	1515	43.37
			a	GL50581_3575	1749	42.08
			a	GL50803_16477	1749	44.25
			a	GL50581_1982	1719	46.6
			a	GL50803_95162	1719	46.07
			a	GL50803_114815	1779	47.95
			a	GL50803_14573	1500	44.8
			c	GL50581_2622	1245	50.12
			c	GL50803_12063	1245	47.31

Table 5.8, cont.

			c	GL50581_1902	2772	46.89
			c	GL50581_3818	1836	47.55
			c	GL50581_3401	1149	52.83
			c	GL50581_572	2022	45.94
			c	GL50581_1	1224	52.7
			c	GL50581_2954	546	58.06
			c	GL50803_33279	2247	45.35
			c	GL50803_137740	1356	53.76
			c	GL50803_137613	1185	51.05
			c	GL50803_101074	2220	43.17
			c	GL50803_137617	2205	44.26
			c	GL50803_116477	2247	44.26
			c	GL50803_14586	2202	40.96
b	GL50803_15250	1872	b	GL50803_112135	3435	48.08
			b	GL50803_16936	4566	49.09
			b	GL50803_113319	7590	49.63
			b	GL50803_103454	2211	43.75
			a	GL50581_4057	771	51.75
			a	GL50803_10330	774	49.87
			a	GL50581_4316	1503	41.85
			a	GL50803_8687	1515	42.38
			a	GL50581_3575	1749	41.17
			a	GL50803_16477	1749	41.57
			a	GL50581_1982	1719	41.13
			a	GL50803_95162	1719	41.24
			a	GL50803_114815	1779	40.58
			a	GL50803_14573	1500	42.93
			c	GL50581_2622	1245	45.38
			c	GL50803_12063	1245	44.74
			c	GL50581_1902	2772	44.71
			c	GL50581_3818	1836	41.18
			c	GL50581_3401	1149	48.22
			c	GL50581_572	2022	40.81
			c	GL50581_1	1224	47.79
			c	GL50581_2954	546	55.68
			c	GL50803_33279	2247	43.06
			c	GL50803_137740	1356	45.87
			c	GL50803_137613	1185	49.7
			c	GL50803_101074	2220	42.52
			c	GL50803_137617	2205	43.86
			c	GL50803_116477	2247	43.06
			c	GL50803_14586	2202	41.83
b	GL50803_112135	3435	b	GL50803_16936	4566	44.19
			b	GL50803_113319	7590	49.87
			b	GL50803_103454	2211	46.54
			a	GL50581_4057	771	49.68
			a	GL50803_10330	774	53.88
			a	GL50581_4316	1503	48.64
			a	GL50803_8687	1515	48.78
			a	GL50581_3575	1749	48.2
			a	GL50803_16477	1749	48.54
			a	GL50581_1982	1719	49.56
			a	GL50803_95162	1719	49.62
			a	GL50803_114815	1779	48
			a	GL50803_14573	1500	48.27
			c	GL50581_2622	1245	53.33
			c	GL50803_12063	1245	50.84
			c	GL50581_1902	2772	44.16
			c	GL50581_3818	1836	50.49
			c	GL50581_3401	1149	54.13
			c	GL50581_572	2022	51.63
			c	GL50581_1	1224	53.19
			c	GL50581_2954	546	58.42
			c	GL50803_33279	2247	49.04
			c	GL50803_137740	1356	53.24
			c	GL50803_137613	1185	54.6
			c	GL50803_101074	2220	48.83
			c	GL50803_137617	2205	49.57
			c	GL50803_116477	2247	47.26
			c	GL50803_14586	2202	48.37
b	GL50803_16936	4566	b	GL50803_113319	7590	43.23
			b	GL50803_103454	2211	48.35
			a	GL50581_4057	771	54.86
			a	GL50803_10330	774	54.13
			a	GL50581_4316	1503	51.1
			a	GL50803_8687	1515	53
			a	GL50581_3575	1749	51.4
			a	GL50803_16477	1749	50.37
			a	GL50581_1982	1719	50.49
			a	GL50803_95162	1719	52.12
			a	GL50803_114815	1779	50.93
			a	GL50803_14573	1500	51.47
			c	GL50581_2622	1245	53.33
			c	GL50803_12063	1245	49.96
			c	GL50581_1902	2772	46.07
			c	GL50581_3818	1836	50.22
			c	GL50581_3401	1149	51.35
			c	GL50581_572	2022	50.15
			c	GL50581_1	1224	51.39
			c	GL50581_2954	546	61.9

Table 5.8, cont.

				c	GL50803_33279	2247	49.13
				c	GL50803_137740	1356	53.39
				c	GL50803_137613	1185	53.84
				c	GL50803_101074	2220	47.21
				c	GL50803_137617	2205	48.71
				c	GL50803_116477	2247	47.35
				c	GL50803_14586	2202	47.91
	b	GL50803_113319	7590	b	GL50803_103454	2211	53.51
				a	GL50581_4057	771	50.58
				a	GL50803_10330	774	57.75
				a	GL50581_4316	1503	52.3
				a	GL50803_8687	1515	53.07
				a	GL50581_3575	1749	51.11
				a	GL50803_16477	1749	52.66
				a	GL50581_1982	1719	52.3
				a	GL50803_95162	1719	54.28
				a	GL50803_114815	1779	52.73
				a	GL50803_14573	1500	52.8
				c	GL50581_2622	1245	55.18
				c	GL50803_12063	1245	57.51
				c	GL50581_1902	2772	50.87
				c	GL50581_3818	1836	54.68
				c	GL50581_3401	1149	56.66
				c	GL50581_572	2022	54.2
				c	GL50581_1	1224	54.9
				c	GL50581_2954	546	61.54
				c	GL50803_33279	2247	54.65
				c	GL50803_137740	1356	59.37
				c	GL50803_137613	1185	56.2
				c	GL50803_101074	2220	54.5
				c	GL50803_137617	2205	53.06
				c	GL50803_116477	2247	52.43
				c	GL50803_14586	2202	50.23
	a	GL50581_4057	771	**a**	**GL50803_10330**	**774**	**74.45**
				a	GL50581_4316	1503	50.45
				a	GL50803_8687	1515	50.32
				a	GL50581_3575	1749	49.81
				a	GL50803_16477	1749	51.23
				a	GL50581_1982	1719	50.71
				a	GL50803_95162	1719	51.36
				a	GL50803_114815	1779	50.97
				a	GL50803_14573	1500	48.38
				b	GL50803_103454	2211	49.81
				c	GL50581_2622	1245	45.65
				c	GL50803_12063	1245	45.53
				c	GL50581_1902	2772	48.64
				c	GL50581_3818	1836	49.55
				c	GL50581_3401	1149	44.36
				c	GL50581_572	2022	48.77
				c	GL50581_1	1224	43.32
				c	GL50581_2954	546	44.32
				c	GL50803_33279	2247	46.17
				c	GL50803_137740	1356	43.97
				c	GL50803_137613	1185	45.27
				c	GL50803_101074	2220	51.1
				c	GL50803_137617	2205	48.9
				c	GL50803_116477	2247	48.38
				c	GL50803_14586	2202	49.55
	a	GL50803_10330	774	a	GL50581_4316	1503	48.45
				a	GL50803_8687	1515	48.45
				a	GL50581_3575	1749	49.87
				a	GL50803_16477	1749	50.13
				a	GL50581_1982	1719	51.29
				a	GL50803_95162	1719	53.1
				a	GL50803_114815	1779	48.84
				a	GL50803_14573	1500	51.42
				b	GL50803_103454	2211	50.13
				c	GL50581_2622	1245	46.51
				c	GL50803_12063	1245	46.25
				c	GL50581_1902	2772	52.71
				c	GL50581_3818	1836	48.58
				c	GL50581_3401	1149	44.44
				c	GL50581_572	2022	51.55
				c	GL50581_1	1224	44.96
				c	GL50581_2954	546	47.25
				c	GL50803_33279	2247	48.84
				c	GL50803_137740	1356	44.57
				c	GL50803_137613	1185	46.77
				c	GL50803_101074	2220	49.74
				c	GL50803_137617	2205	51.03
				c	GL50803_116477	2247	50.26
				c	GL50803_14586	2202	50.78
	a	GL50581_4316	1503	**a**	**GL50803_8687**	**1515**	**74.98**
				a	GL50581_3575	1749	47.64
				a	GL50803_16477	1749	47.57
				a	GL50581_1982	1719	47.24
				a	GL50803_95162	1719	49.04
				a	GL50803_114815	1779	45.84
				a	GL50803_14573	1500	57.73
				b	GL50803_103454	2211	44.71

Table 5.8, cont.

				c	GL50581_2622	1245	42.41
				c	GL50803_12063	1245	43.37
				c	GL50581_1902	2772	47.04
				c	GL50581_3818	1836	43.58
				c	GL50581_3401	1149	45.34
				c	GL50581_572	2022	42.65
				c	GL50581_1	1224	44.93
				c	GL50581_2954	546	51.65
				c	GL50803_33279	2247	42.91
				c	GL50803_137740	1356	41
				c	GL50803_137613	1185	42.7
				c	GL50803_101074	2220	45.11
				c	GL50803_137617	2205	42.38
				c	GL50803_116477	2247	47.11
				c	GL50803_14586	2202	44.58
	a	GL50803_8687	1515	a	GL50581_3575	1749	47.59
				a	GL50803_16477	1749	47.85
				a	GL50581_1982	1719	45.15
				a	GL50803_95162	1719	47.72
				a	GL50803_114815	1779	47.06
				a	GL50803_14573	1500	59.2
				b	GL50803_103454	2211	43.89
				c	GL50581_2622	1245	41.77
				c	GL50803_12063	1245	42.25
				c	GL50581_1902	2772	45.35
				c	GL50581_3818	1836	42.64
				c	GL50581_3401	1149	44.47
				c	GL50581_572	2022	45.61
				c	GL50581_1	1224	43.95
				c	GL50581_2954	546	48.72
				c	GL50803_33279	2247	44.82
				c	GL50803_137740	1356	40.56
				c	GL50803_137613	1185	45.49
				c	GL50803_101074	2220	43.76
				c	GL50803_137617	2205	45.87
				c	GL50803_116477	2247	45.28
				c	GL50803_14586	2202	45.08
	a	GL50581_3575	1749	**a**	**GL50803_16477**	**1749**	**76.73**
				a	GL50581_1982	1719	50.44
				a	GL50803_95162	1719	49.8
				a	GL50803_114815	1779	49.17
				a	GL50803_14573	1500	44.67
				b	GL50803_103454	2211	42.82
				c	GL50581_2622	1245	43.94
				c	GL50803_12063	1245	44.1
				c	GL50581_1902	2772	45.11
				c	GL50581_3818	1836	41.57
				c	GL50581_3401	1149	45.08
				c	GL50581_572	2022	41.97
				c	GL50581_1	1224	43.79
				c	GL50581_2954	546	58.61
				c	GL50803_33279	2247	41.8
				c	GL50803_137740	1356	42.63
				c	GL50803_137613	1185	46.5
				c	GL50803_101074	2220	43.11
				c	GL50803_137617	2205	44.88
				c	GL50803_116477	2247	42.77
				c	GL50803_14586	2202	43
	a	GL50803_16477	1749	a	GL50581_1982	1719	51.02
				a	GL50803_95162	1719	50.38
				a	GL50803_114815	1779	49.06
				a	GL50803_14573	1500	46.67
				b	GL50803_103454	2211	44.25
				c	GL50581_2622	1245	44.34
				c	GL50803_12063	1245	43.05
				c	GL50581_1902	2772	44.71
				c	GL50581_3818	1836	39.97
				c	GL50581_3401	1149	47.35
				c	GL50581_572	2022	43.4
				c	GL50581_1	1224	44.36
				c	GL50581_2954	546	53.3
				c	GL50803_33279	2247	44.6
				c	GL50803_137740	1356	46.17
				c	GL50803_137613	1185	48.35
				c	GL50803_101074	2220	43.45
				c	GL50803_137617	2205	45.63
				c	GL50803_116477	2247	43.57
				c	GL50803_14586	2202	43.28
	a	GL50581_1982	1719	**a**	**GL50803_95162**	**1719**	**76.27**
				a	GL50803_114815	1779	53.52
				a	GL50803_14573	1500	47.13
				b	GL50803_103454	2211	45.55
				c	GL50581_2622	1245	46.18
				c	GL50803_12063	1245	45.06
				c	GL50581_1902	2772	47.59
				c	GL50581_3818	1836	42.87
				c	GL50581_3401	1149	49
				c	GL50581_572	2022	44.39
				c	GL50581_1	1224	44.77
				c	GL50581_2954	546	54.4

Table 5.8, cont.

			c	GL50803_33279	2247	48.52
			c	GL50803_137740	1356	44.32
			c	GL50803_137613	1185	47.59
			c	GL50803_101074	2220	44.5
			c	GL50803_137617	2205	44.39
			c	GL50803_116477	2247	45.49
			c	GL50803_14586	2202	45.2
a	GL50803_95162	1719	a	GL50803_114815	1779	52.59
			a	GL50803_14573	1500	44.6
			b	GL50803_103454	2211	44.44
			c	GL50581_2622	1245	47.15
			c	GL50803_12063	1245	46.75
			c	GL50581_1902	2772	47.06
			c	GL50581_3818	1836	42.76
			c	GL50581_3401	1149	47.26
			c	GL50581_572	2022	44.62
			c	GL50581_1	1224	46.32
			c	GL50581_2954	546	51.1
			c	GL50803_33279	2247	47.06
			c	GL50803_137740	1356	46.61
			c	GL50803_137613	1185	45.91
			c	GL50803_101074	2220	45.9
			c	GL50803_137617	2205	45.03
			c	GL50803_116477	2247	45.43
			c	GL50803_14586	2202	46.13
a	GL50803_114815	1779	a	GL50803_14573	1500	47.87
			b	GL50803_103454	2211	42.21
			c	GL50581_2622	1245	46.91
			c	GL50803_12063	1245	44.1
			c	GL50581_1902	2772	48.96
			c	GL50581_3818	1836	41.93
			c	GL50581_3401	1149	49.43
			c	GL50581_572	2022	42.83
			c	GL50581_1	1224	46.41
			c	GL50581_2954	546	54.95
			c	GL50803_33279	2247	46.04
			c	GL50803_137740	1356	46.24
			c	GL50803_137613	1185	47.17
			c	GL50803_101074	2220	44.91
			c	GL50803_137617	2205	43.11
			c	GL50803_116477	2247	42.95
			c	GL50803_14586	2202	43.73
a	GL50803_14573	1500	c	GL50581_2622	1245	42.57
			c	GL50803_12063	1245	43.94
			c	GL50581_1902	2772	46.6
			c	GL50581_3818	1836	43.07
			c	GL50581_3401	1149	44.65
			c	GL50581_572	2022	44.13
			c	GL50581_1	1224	40.6
			c	GL50581_2954	546	52.38
			c	GL50803_33279	2247	46.47
			c	GL50803_137740	1356	41.81
			c	GL50803_137613	1185	43.97
			c	GL50803_101074	2220	46.33
			c	GL50803_137617	2205	44.27
			c	GL50803_116477	2247	45.33
			c	GL50803_14586	2202	43.87
			b	GL50803_103454	2211	44.33
c	GL50581_2622	1245	**c**	**GL50803_12063**	**1245**	**77.67**
			c	GL50581_1902	2772	49.96
			c	GL50581_3818	1836	46.83
			c	GL50581_3401	1149	41.95
			c	GL50581_572	2022	48.27
			c	GL50581_1	1224	42.32
			c	GL50581_2954	546	52.38
			c	GL50803_33279	2247	48.51
			c	GL50803_137740	1356	42.73
			c	GL50803_137613	1185	42.19
			c	GL50803_101074	2220	49.48
			c	GL50803_137617	2205	51.08
			c	GL50803_116477	2247	48.43
			c	GL50803_14586	2202	47.07
			b	GL50803_103454	2211	50.76
c	GL50803_12063	1245	c	GL50581_1902	2772	48.11
			c	GL50581_3818	1836	44.1
			c	GL50581_3401	1149	44.21
			c	GL50581_572	2022	48.76
			c	GL50581_1	1224	39.87
			c	GL50581_2954	546	52.93
			c	GL50803_33279	2247	49.08
			c	GL50803_137740	1356	40.8
			c	GL50803_137613	1185	42.03
			c	GL50803_101074	2220	47.71
			c	GL50803_137617	2205	47.71
			c	GL50803_116477	2247	48.43
			c	GL50803_14586	2202	50.84
			b	GL50803_103454	2211	47.71
c	GL50581_1902	2772	c	GL50581_3818	1836	47.44
			c	GL50581_3401	1149	53.52
			c	GL50581_572	2022	45.94

End of Table 5.8

			c	GL50581_1	1224	50.9
			c	GL50581_2954	546	73.26
			c	GL50803_33279	2247	46.42
			c	GL50803_137740	1356	51.99
			c	GL50803_137613	1185	48.1
			c	GL50803_101074	2220	43.06
			c	GL50803_137617	2205	45.22
			c	GL50803_116477	2247	45.04
			c	GL50803_14586	2202	43.69
			b	GL50803_103454	2211	44.55
c	GL50581_3818	1836	c	GL50581_3401	1149	51.87
			c	GL50581_572	2022	44.61
			c	GL50581_1	1224	50.9
			c	GL50581_2954	546	60.26
			c	GL50803_33279	2247	47.71
			c	GL50803_137740	1356	51.03
			c	GL50803_137613	1185	52.15
			c	GL50803_101074	2220	46.13
			c	GL50803_137617	2205	47.6
			c	GL50803_116477	2247	50.16
			c	GL50803_14586	2202	46.46
			b	GL50803_103454	2211	43.57
c	GL50581_3401	1149	c	GL50581_572	2022	49.26
			c	GL50581_1	1224	46.74
			c	GL50581_2954	546	54.76
			c	GL50803_33279	2247	52.74
			c	GL50803_137740	1356	45.6
			c	GL50803_137613	1185	45.26
			c	GL50803_101074	2220	51.78
			c	GL50803_137617	2205	54.4
			c	GL50803_116477	2247	55
			c	GL50803_14586	2202	52.31
			b	GL50803_103454	2211	50.13
c	GL50581_572	2022	c	GL50581_1	1224	47.96
			c	GL50581_2954	546	58.97
			c	GL50803_33279	2247	49.31
			c	GL50803_137740	1356	48.89
			c	GL50803_137613	1185	47.51
			c	GL50803_101074	2220	47.38
			c	GL50803_137617	2205	46.83
			c	GL50803_116477	2247	44.31
			c	GL50803_14586	2202	43.22
			b	GL50803_103454	2211	42.98
c	GL50581_1	1224	c	GL50581_2954	546	59.89
			c	GL50803_33279	2247	52.12
			c	GL50803_137740	1356	45.67
			c	GL50803_137613	1185	44.47
			c	GL50803_101074	2220	52.12
			c	GL50803_137617	2205	50.82
			c	GL50803_116477	2247	52.12
			c	GL50803_14586	2202	52.94
			b	GL50803_103454	2211	
c	GL50581_2954	546	c	GL50803_33279	2247	62.82
			c	GL50803_137740	1356	58.24
			c	GL50803_137613	1185	56.04
			c	GL50803_101074	2220	62.45
			c	GL50803_137617	2205	63.74
			c	GL50803_116477	2247	63
			c	GL50803_14586	2202	63.55
			b	GL50803_103454	2211	54.4
c	GL50803_33279	2247	c	GL50803_137740	1356	53.54
			c	GL50803_137613	1185	59.83
			c	GL50803_101074	2220	46.98
			c	GL50803_137617	2205	47.17
			c	GL50803_116477	2247	43.79
			c	GL50803_14586	2202	46
			b	GL50803_103454	2211	41.7
c	GL50803_137740	1356	c	GL50803_137613	1185	48.19
			c	GL50803_101074	2220	51.77
			c	GL50803_137617	2205	51.77
			c	GL50803_116477	2247	50.74
			c	GL50803_14586	2202	48.6
			b	GL50803_103454	2211	49.41
c	GL50803_137613	1185	c	GL50803_101074	2220	55.86
			c	GL50803_137617	2205	54.09
			c	GL50803_116477	2247	53.25
			c	GL50803_14586	2202	54.01
			b	GL50803_103454	2211	52.24
c	GL50803_101074	2220	c	GL50803_137617	2205	81.68
			c	GL50803_116477	2247	44.28
			c	GL50803_14586	2202	56.95
			b	GL50803_103454	2211	44.19
c	GL50803_137617	2205	c	GL50803_116477	2247	42.45
			c	GL50803_14586	2202	56.63
			b	GL50803_103454	2211	44.04
c	GL50803_116477	2247	c	GL50803_14586	2202	45.32
			b	GL50803_103454	2211	43.69
c	GL50803_14586	2202	b	GL50803_103454	2211	43.05

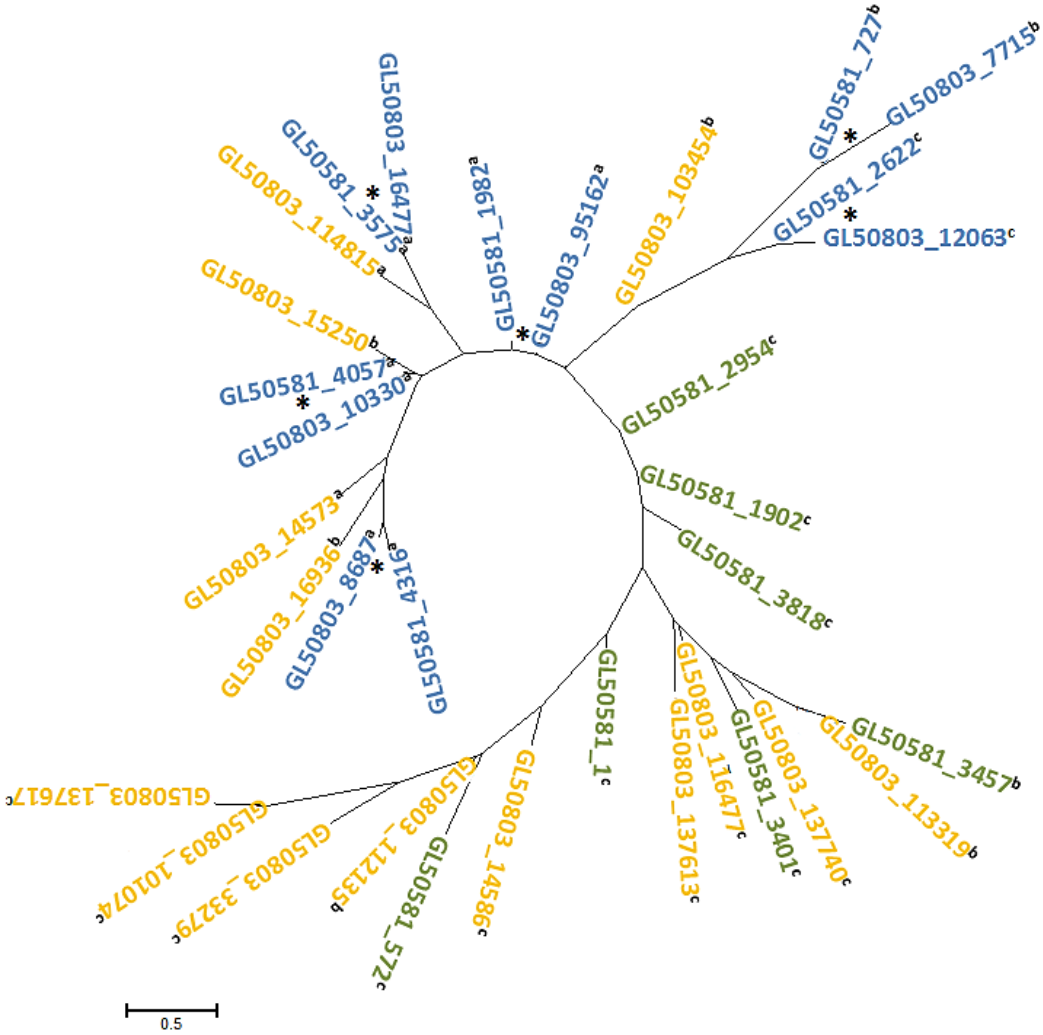


Fig 5. 7: Phylogenetic tree of the Tenascins, VSPs and High cysteine membrane proteins identified as secreted factors in GS and WB isolate proteomes. Gene sequences were obtained from GiardiaDB, aligned using ClustalW. The tree was realised by a Maximum likelihood approach using MEGA software (v 6.06) (Appendix XI). Only the proteins present in the different secretion profiles were included in the analysis. Proteins identified in both isolate are highlighted in blue, proteins identified in WB isolate dataset only are highlighted in orange, and proteins identified in GS isolate dataset are highlighted in green. Ortholog proteins are shown by *****. **a**: Tenascin, **b**: HCMP, **c**: VSP.
All the GL50581_ accession numbers represent GS isolate proteins and all the GL50803_ represent WB isolate proteins.

Interestingly, alignment and clustering of sequences suggest that all eight of these proteins show some degree of inter-relatedness. They fall into three sets of orthologous pairs with similarity from 77.86 % to 88.64 % (Table 5.9). One orthologous pair could be linked to the HCMP/Tenascin family and this may suggest that all the hypothetical secreted proteins are actually extended members of this group. Two lineage-specific proteins cluster with two of the three orthologous pairs implying that they may have arisen by duplication and divergence of this group (Fig 5.8). As with the *G*CATB and HCMP families, the SFHPs of *Giardia* show substantial heterogeneity between assemblages and are good candidate for intra-assemblage strain discrimination which may warrant further testing against a range of clinical isolates

A comparative genomic analysis between assemblage A, WB isolate, assemblage B, GS isolate, and assemblage E, P15 isolate, identified 5,012 protein coding genes of which approximately 90 % were shared between the three *Giardia* isolates (Jerlstrom-Hultqvist et al., 2010). This correlates with the fact that over 90 % of the proteins expressed were shared between GS and WB lineages. This suggests that a set of core genes may be common to all *Giardia* isolates and at the origin of most of the protein expression occurring in *Giardia* trophozoites. In their genome analysis, Jerlstrom-Hultqvist et al also identified 5, 31 and 38 protein coding genes as WB-, GS-, and P15- specific respectively. The proteins identified as lineage-specific in this study were compared to the WB- and GS-specific protein coding genes from Jerlstrom-Hulqvist and collaborator study to look for correlation. Only two GS hypothetical proteins, GL50581_3192 and GL50581_2613, were present in both studies as lineage-specific (Table 5.2). As previously stated, in this *Giardia* proteome analysis, proteins with no defined orthologs in either assemblage A (DH and WB isolates) or assemblage B (GS isolate) were considered lineage-specific; therefore, these proteins were also compared to the protein coding sequences shared between P15 and GS or P15 and WB (Table 5.2 and 5.3). None of the lineage-specific proteins identified in this study were in Jerlstrom-Hultqvist et al study. This suggests that these two GS lineage-specific may allow further testing as assemblage-specific markers and for discrimination of lineages within assemblage B.

Table 5.9: Sequence similarities of the genes coding for the hypothetical proteins identified as secreted in GS and WB isolate proteomes. Orthologs are highlighted in bold. Similarity scores were assessed via ClustalW (Appendix XII).

Name	Length (bp)	Name	Length (bp)	Score (%)
GL50581_2767	309	**GL50803_5258**	**321**	**84.47**
		GL50581_4133	396	55.34
		GL50803_5810	396	55.34
		GL50581_4180	1725	85.44
		GL50803_113038	1725	85.11
		GL50581_352	453	55.66
		GL50803_16522	453	53.72
GL50803_5258	321	GL50581_4133	396	51.09
		GL50803_5810	396	51.71
		GL50581_4180	1725	79.75
		GL50803_113038	1725	85.98
		GL50581_352	453	58.88
		GL50803_16522	453	53.89
GL50581_4133	396	**GL50803_5810**	**396**	**88.64**
		GL50581_4180	1725	81.06
		GL50803_113038	1725	82.58
		GL50581_352	453	48.23
		GL50803_16522	453	45.96
GL50803_5810	396	GL50581_4180	1725	79.04
		GL50803_113038	1725	84.85
		GL50581_352	453	54.8
		GL50803_16522	453	53.03
GL50581_4180	1725	**GL50803_113038**	**1725**	**77.86**
		GL50581_352	453	80.13
		GL50803_16522	453	77.26
GL50803_113038	1725	**GL50581_352**	**453**	**76.6**
		GL50803_16522	453	79.91
GL50581_352	453	GL50803_16522	453	78.15

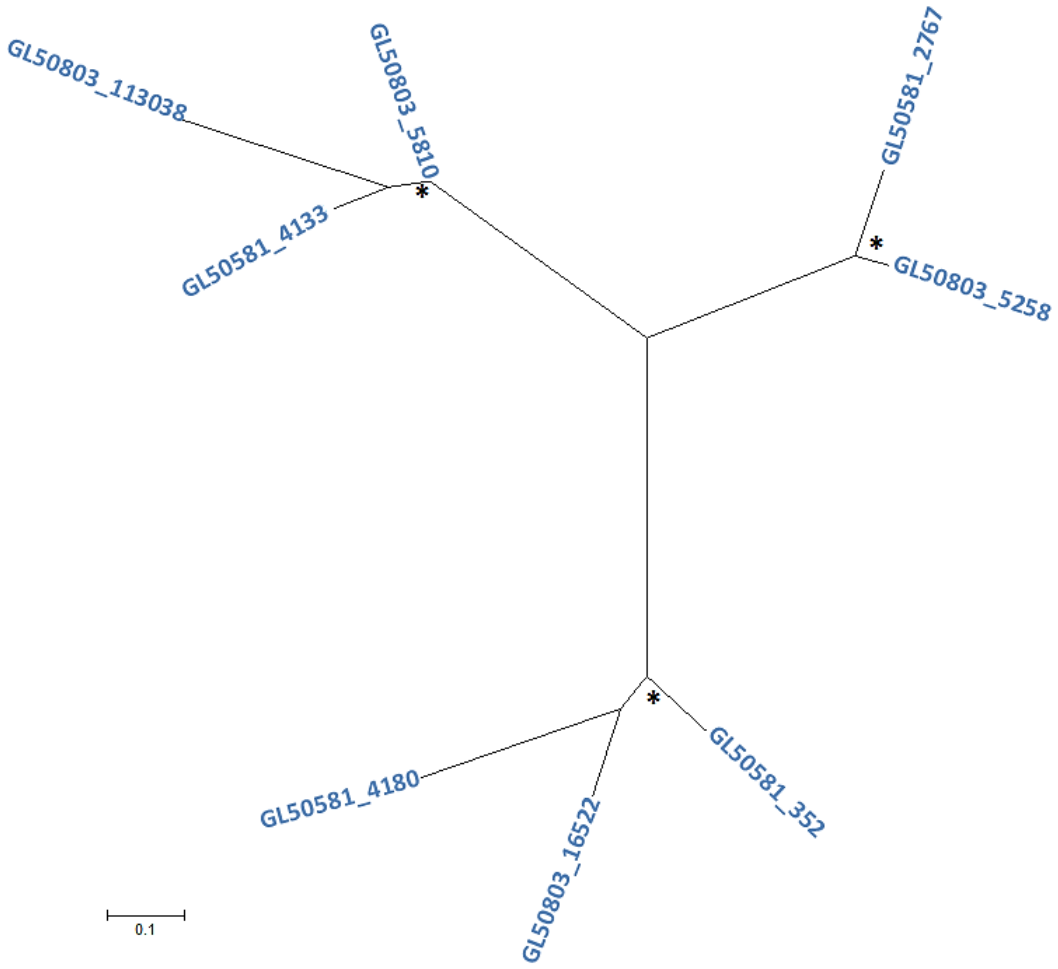


Fig 5. 8: Phylogenetic tree of the hypothetical proteins identified as secreted by GS and WB isolate trophozoites. Gene sequences were obtained from GiardiaDB, aligned using ClustalW. The tree was realised by a Maximum likelihood approach using MEGA software (v 6.06) (Appendix XII). Only the proteins identified in both GS and WB isolates via both MS techniques were included in this analysis. Ortholog proteins are shown by ***.**
All the GL50581_ accession numbers represent GS isolate proteins and all the GL50803_ represent WB isolate proteins.

5.4.4/ Other *Giardia*-secreted proteins

5.4.4.1/ Other conserved *Giardia*-secreted proteins

When WB and GS secretomes were compared, 15 proteins were shown to be conserved between the two secretomes, (Table 5.4). The five most abundant conserved secreted proteins were the same hypothetical protein, Cathepsin B precursor and three Tenascins in both isolates; which suggests that these proteins may play an important role in *Giardia* secretory pathway. An extracellular nuclease was also shown to be secreted by both *Giardia* isolates; this protein is also an ortholog to assemblage E P15 isolate suggesting a function conserved between all *Giardia* assemblages. In bacterial pathogens such as *V. cholera* and *S. aureus*, extracellular nucleases have been shown to be involved in nutrient acquisition, detachment from biofilms but also inhibition of biofilm initiation and evasion of neutrophil extracellular traps (Kiedrowski et al., 2014, Seper et al., 2011). In Trypanosomitidae, such as *Leishmania*, nucleases have been shown to be present at the surface of the plasma membrane and involved in the salvage of preformed purines from host sources (Gottlieb, 1989). Some studies have shown that the importance of extracellular nuclease such as DNase in cystic fibrosis. Extracellular DNases have been shown to reduce mucus thickness by cleaving nucleic acids present in it (Harms et al., 1998, Brandt et al., 1995) which suggests that *Giardia* extracellular nuclease may play a role in the degradation of mucin during infection. Interestingly, bacterial extracellular nucleases have also been shown to have a mucolytic activity during host-pathogen interactions (Sanders et al., 2007) which suggests that giardial extracellular nuclease may also have a mucolytic activity. Yet, its role in *Giardia*-host interactions requires further investigations.

Interestingly, a cyclin-dependent kinases regulatory subunit was shown to be released by both *Giardia* isolates. Cyclin-dependent kinases (CDKs) have been shown to be involved in the cell cycle progression and regulation; and, in eukaryotes, CDKs are also implicated in the control of gene transcription. Moreover, in complex cell cycles, they are acting as information processors integrating extracellular and intracellular signals to ensure the smooth coordination

of cell cycle events in the case of environmental change or mechanical failure (Morgan, 1997); which suggests that CDKs may communicate with some surface proteins. CDKs are regulated by CDKs regulatory subunits. *Giardia* CDKs regulatory subunit has a high signalP suggesting this protein may be released in the environment or present at the surface of the parasite. This protein may be acting as transmitter of any environmental changes within the host intestinal lumen to CDKs which will regulate the parasite cell cycle accordingly to this signal. This may allow trophozoites to adapt to any environmental changes or may even be at the origin of trophozoite encystation, as CDKs can regulate the gene transcription function. However, the involvement of CDKs regulatory subunit in these different *Giardia* processes has to be further investigated to confirm or refute these hypotheses.

Surprisingly, two proteins belonging to the phosphoglycolate phosphatase family have been shown to be secreted by GS and WB isolate trophozoites. These two proteins are both orthologs to a different phosphoglycolate phosphatase within each isolate (Table 5.5 and 5.6). In plants and blood red cells, this enzyme is responsible for the assimilation of oxygen and the affinity of Haemoglobin for oxygen respectively (Rose et al., 1986). In chemoautotrophic organisms, this enzyme participates in CO_2 assimilation (Shively et al., 1998). In *E. coli*, phosphoglycolate phosphatase has been shown to be involved in DNA repair after exposure to oxidative stress and also in the bacteria carbon metabolic process (Teresa Pellicer et al., 2003). Being part of the carbon metabolic process, this protein should be located within the cell cytoplasm; yet, in *Giardia*, proteomic analysis suggests that it may be secreted. However, the proteomic analysis matched only three and four peptides to these proteins indicating that the identification may not be robust and the ratio obtained was not reliable. This along with the possibility of a misannotation of these proteins suggests that the results should be interpreted with care.

5.4.4.2/ GS lineage-specific secreted proteins

During this study, differences in GS and WB secretion profiles were noticed. Indeed, some proteins appeared to be secreted in one isolate but only present as non-secreted factor in the other isolate. Three proteins were secreted by GS isolate

trophozoites: (i) A-type flavoprotein lateral transfer candidate, (ii) FixW protein, (iii) Sentrin.

(i) A-type flavoprotein lateral transfer candidate has been shown to have a high oxygen reductase activity during *Giardia* infection (Di Matteo et al., 2008); this protein may be released in the host intestinal environment to scavenge O_2 thus allowing *Giardia* trophozoites to survive in the small intestinal environment and promoting its pathogenicity. This protein was also expressed in both pellet and supernatant in WB isolate but to a very low abundance suggesting that GS isolate trophozoites may be more sensitive to oxidative stress early in the infection. Whether GS isolate trophozoites require A-type flavoprotein lateral transfer candidate throughout the infection or just at the beginning remains unclear and should be investigated. Moreover, whether WB isolate trophozoites only require this protein when some drastic changes occur within the host intestinal lumen occur also requires further investigation.

(ii) FixW protein has been shown to be a redox sensor in non-pathogenic bacteria such as *Rhizobium leguminosarum* (Martinez et al., 2004). Redox sensors can detect hypoxia, nictric oxide or carbaon monoxide changes in the bacteria environment; however, not a lot is known about FixW protein yet. Another Fix protein, FixL, in *M. tuberculosis* has been shown to be involved in environmental redox changes sensing and thus involved in *M. tuberculosis* pathogenicity (Bhat et al., 2012). However, the role of this bacterial protein in *Giardia* remains unclear. Moreover, FixW protein has a very low signalP suggesting that this protein is unlikely to be secreted (Table 5.5). The fact that this protein has been identified as secreted in GS isolate suggests a possible misannotation of this protein in the database. Protein annotations are based on sequence homology with other proteins that may be not closely related, thus leading to some wrong annotation of proteins.

(iii) The last protein identified as most likely secreted by GS isolate trophozoites is a Sentrin protein. Sentrin is a protein involved in the ubiquitination of proteins to render them resistant to degradation (Kamitani et al., 1998). This protein appears to be an evolutionary conserved protein as it has been identified in prokaryotic and eukaryotic organisms such as *S. cerevisiae*, *A. thaliana* and *Homo*

sapiens, which suggests that Sentrin may have a specialised function in cell metabolism. Most of the proteins ubiquitinated by Sentrin appear to be located in the cell nucleus, and the overexpression of this protein protects organisms against tumour necrosis factor-induced cell death (Kamitani et al., 1998). Due to its role in the ubiquitination of proteins, Sentrin was expected to be present in *Giardia* trophozoite cytosol and not in its secretome. Why this protein would be secreted in GS isolate remains unclear. What would be the advantages for *Giardia* GS isolate trophozoites to secrete such a protein are also undetermined. Similarly to phosphoglycolate phosphatase, the identification of this protein during the proteomic analysis was not as robust as for the other proteins as only two peptides were matched to Sentrin protein.

5.4.4.3/ WB lineage-specific secreted proteins

Six proteins were secreted by WB isolate trophozoites but not GS isolate trophozoites: (i) Peptidyl-prolyl cis-trans isomerase B (PPIB) precursor, (ii) Neurogenic locus notch protein precursor, (iii) Translation initiation inhibitor, (iv) ATP-binding cassette (ABC) protein 5, (v) Peptide methionine sulfoxide reductase (MsrB), (vi) CXC rich protein and (vii) Elongation factor 1-α (EF-1 α) (Table 5.6).

(i) PPIBs are conserved proteins that bind the immunosuppressive drug Cyclosporin A with high affinity (Price et al., 1991). They are mainly located in the ER but can be associated with secretory pathway and released in biological fluids. PPIBs have been shown to be secreted by parasites such as *T. gondii* and *N. caninum* to control the acute phase of both toxoplamosis and neosporosis (Bell et al., 2006). It has been suggested that these proteins may play a crucial role in the stimulation of interferon (IFN)-γ. In this study, PPIB was shown to be present in both P and SP with a SP : P ratio of approximately 0.3; which suggests that this protein may not be released by the parasite. Or if released, perhaps to a very low level when there are no host stimulus to enforce the parasite survival. Whether this protein is released and has the same function as *T. gondii* and *N. caninum* PPIB requires further investigation.

(ii) Neurogenic locus notch proteins are transmembrane proteins with repeated extracellular EGF- domains and intracellular notch domains. In mammals,

these proteins are part of an evolutionary conserved intercellular pathway that regulates interactions between physically adjacent cells (Vardar et al., 2003). In *Schistosoma mansoni*, Notch proteins may play major roles in pairing process, proliferation, differentiation of vitelline cell, expression of female-specific genes, embryogenesis and sexual reproduction (Morais et al., 2013, Loverde et al., 2007). In this study, Neurogenic locus notch protein precursor was identified in both P and SP with a SP : P ratio of approximately 0.6 suggesting that this protein is more abundant within the trophozoites. In *Giardia*, this transmembrane protein was shown to have a high signalP suggesting that if not secreted, this protein may be present at the cell surface of trophozoites. This protein may be labile and easily detach from the surface membrane which could explain why it has been identified in high abundance in the supernatant too. What its role in *Giardia* is remains unknown and should be further investigated.

(iii) Translation initiation inhibitors are proteins inhibiting the initiation of the mRNA translation into proteins; therefore they should be located into the cell cytosol (Malys and McCarthy, 2011). However, in this study, one translation initiation inhibitor was identified as most likely secreted by *Giardia* WB isolate trophozoites. This protein was abundantly present in both P and SP and with a low SP : P ratio of approximately 0.5 which suggests that this protein may not be secreted or released by the trophozoites. Why this protein is present in such abundance in the SP is not known, over 15 peptides were present and matched to this protein during the analysis. A misannotation in the *Giardia* database may be plausible. However, another explanation would be the over-representation of this protein in the supernatant due to its high solubility and stability suggesting the adoption of a new role in *Giardia* as previously shown for ADI, OCT and enolase (Ringqvist et al., 2008).

(iv) ABC proteins play a role in transducing the energy of ATP binding and hydrolysis to a wide variety of physiological processes in all eukaryotic and bacterial organisms. Most of them are associated with transport of substrates across biological membranes. In many bacteria, ABC transporters play roles in virulence and survival to various environments (Linton and Higgins, 1998). ABC proteins possess two TM domains and a hydrophobic tail where it hydrolyses ATP

which suggests that these proteins are more likely not being secreted (Dean et al., 2001). However, some of the ABC transporters have been shown to be vacuolar in plants (Jasinski et al., 2001). The fact that ABC protein 5 is more abundant in WB isolate supernatant than pellet suggests that this protein may be released in the external environment which is unexpected due to its two TM domains. Perhaps, this protein is abundantly present in vacuoles in *Giardia* WB isolate and is involved in some of the parasite resistance mechanisms as vacuolar ABC proteins observed in plants (Jasinski et al., 2001). As with VSPs and HCMPs, this protein may be shed into the supernatant and have an undetermined role in the host-parasite interactions. Not all the functions of the ABC transporter family have been understood yet, and require further investigations.

(v) MsrB catalyses the reduction of free and protein-bound methionine sulfoxides to corresponding methionines (Kim and Gladyshev, 2004). The reduction of protein methionine residues constitutes a mechanism for the scavenging of reactive oxygen species (ROS) which are responsible for a fundamental innate defence against pathogens in various host organisms (Kim and Gladyshev, 2004). MrsB is an antioxidant protein that protect organisms from the cytotoxic effect of ROS and therefore from cell death. MsrB has been shown to be crucial for the virulence of *Salmonella typhimurium* and for the immune evasion of *Schistosoma mansoni* (Denkel et al., 2011, Oke et al., 2009). This is a cytosolic protein which is unlikely to be secreted. Similarly to Sentrin, only three peptides were matched to this protein during the proteomic analysis indicating an unreliable ratio. This suggests that this data should also be interpreted with care.

(vi) CXC rich proteins are proteins rich in cysteine residues. *Giardia* has been shown to express various types of cysteine-rich proteins such as VSPs, Tenascins and HCMPs. Another CXC rich protein family was identified in *Giardia* cyst but also involved in the antigenic variation process in trophozoites (Davids et al., 2006, Adam et al., 1992). Similarly to Tenascins, VSPs and HCMPs, these proteins have a very high SignalP and only one TM domain suggesting that they may be labile surface protein. CXC rich protein may be involved in the parasite escape from the host immune system; however to what degree this protein is involved in this immune evasion has yet to be determined. Moreover, CXC rich protein exists in GS

isolate but was not identified as secreted in this study, what trigger the expression and secretion of this protein at the surface of the trophozoite remains unclear and should be investigated. Also, how different from VSPs are these proteins is unknown and should be studied.

(vii) EF-1α is responsible for the enzymatic delivery of the aminoacyl transfer RNAs (tRNAs) to the ribosome. This protein is normally located in the nucleus. However, a study suggested that *Giardia* WB trophozoites may release EF-1α into the culture medium during *in vitro* growth (Skarin et al., 2011), which correlates with the WB lineage-specific EF-1α identified as secreted in this study. Many organisms have been shown to express EF-1α in excess, suggesting that this protein may have some other functions (Condeelis, 1995). EF-1α was studied in the context of pathogenicity and virulence for various micro-organisms. In *H. pylori*, virulence factor CagA is translocated into the host cell and disrupts the epithelial cell polarity (Tan et al., 2009). The secreted *Leishmania* EF-1α appears to down-regulate host inflammatory cell signalling (Nandan and Reiner, 2005). In this analysis, EF-1α was shown to be of high abundance, high stability and high solubility suggesting that it might be over-represented into supernatants and be involved in *Giardia* pathogenicity. However, it remains to be seen what role has EF-1α as secreted factor in giardial infection. Whether this protein is also released by *Giardia* GS isolate later during infection or is only released by WB isolate trophozoites throughout the infection requires further investigation.

5.4.4.4/ Discrepancy in Orbitrap and QeXactive datasets

Although various proteins have been identified by both MS techniques, a few differences and inconsistencies were also observed within the isolate secretion profiles obtained by Orbitrap and QeXactive MS (Data shown in Appendix IX). For example, in GS secretion profiles, two different PDI5 were identified as the most abundant secreted proteins. PDIs are enzymes located in the ER which catalyses the formation and breakage of disulfide bonds between cysteine residues within proteins as they fold (Wilkinson and Gilbert, 2004). PDI5 were excluded from the analysis as two different PDI5 were identified in the two datasets, but whether or not these proteins can be secreted or released by the parasite should be verified,

especially considering their high signalP which suggests that PDIs may have the potential to be secreted. Apart from PDI, in both isolate datasets, some proteins present in the proteasome and some cytoskeletal proteins such as spindle pole protein, β-tubulin, and Axonemal p66.0 were suggested to be secreted by either the Orbitrap MS dataset or the QeXactive Ms dataset. This indicates that these two datasets are best analysed together to increase robustness, as it has been in this study.

5.5/ Summary

This study shows that both *Giardia* GS and WB isolate trophozoites have an enriched proteome and that most proteins expressed are shared between isolates. When the secretome analysis was performed, several proteins known to be cytosolic and with low signalP were identified as secreted which suggests that more proteins than expected may be misannotated within the databases used in this study, GiardiaDB and Uniprot, and that a correction of these databases should be done.

Although several proteins have been shown to potentially be secreted by trophozoites, only seven candidates appear to be more likely secreted: (i) Cathepsin B (ii) Tenascin, (iii) Hypotethical proteins, (iv) Extracellular nuclease; which were all conserved between isolates or lineage-specific. (v) A-type flavoprotein lateral transfer candidate identified in GS isolate, (vi) PPIB and (vii) EF-1α (lineage-specific), which were both identified in WB isolate (Table 5.10). Four other families were abundantly present in secretome but are known cell-surface proteins: VSPs, HCMPs, CXC-rich protein and Neurogenic locus notch protein precursor.

Table 5. 70: List of putatively *Giardia*-secreted proteins. Proteins most likely to be secreted are highlighted in blue; proteins with unclear secretior profile are highlighted in grey. The remaining proteins are the proteins considered as not likely secreted due to unreliable ratio and low peptide matches obtained during the proteomic analysis. The accession numbers in bold represent in which *Giardia* isolate the proteins were identified.

Protein Accession Number					QeXactive MS SP/P Ratio	
GS isolate	WB isolate	Protein name	SignalP[a]	Peptides ≥ 5 [b]	GS	WB
GL50581_3607	**GL50803_8742**	Extracellular nuclease, putative	1	Y	4.44	-
GL50581_3484	**GL50803_2661**	CDK-regulatory subunit	-	N	2.80	0.36
GL50581_1340	GL50803_14841	Phosphoglycolate phosphatase	-	N	-	
GL50581_1485	**GL50803_14546**					-
GL50581_1626	GL50803_10358	A-type flavoprotein lateral transfer candidate	-	Y	-	
GL50581_3056	GL50803_23888	FixW protein, putative	0.06	N	6.46	
GL50581_3210	GL50803_7760	Sentrin	-	N	-	
GL50581_1019	**GL50803_17163**	PPIB	-	Y		0.44
GL50581_31	**GL50803_16322**	Neurogenic locus Notch protein precursor	0.99	Y		0.69
GL50581_4017	**GL50803_480**	Translation initiation inhibitor	-	Y		0.43
GL50581_3399	**GL50803_8227**	ATP-binding cassette protein 5	-	Y		0.99
GL50581_3084	**GL50803_5180**	Peptide methionine sulfoxide reductase msrB	-	N		0.71
GL50581_4509	**GL50803_17476**	CXC-rich protein	1	Y		0.59
-	**GL50803_112312**	EF-1α	-	Y		-

[a] **Signal peptide**

[b] **5 or more peptides were matched to the protein during the proteomic analysis**

Chapter VI

General discussion and Future research.

6.1/ General discussion

In this thesis, data has been provided to determine the secretion profile of both *Giardia* assemblage A (WB isolate) and B (GS isolate) trophozoites from *in vitro* culture (Table 6.1), and to illustrate the effects of *Giardia*-secreted factors on intestinal cell morphology and on their main water and ion transporters. Giardiasis and associated symptoms have been thought to be mainly caused by trophozoites, at the site of infection. Therefore, most studies have focused on the effect of trophozoites on intestinal epithelia. The data presented demonstrates that *Giardia* trophozoites secrete proteins factors causing disruption of intestinal epithelia and suggests that these secreted proteins may act as possible mediators of symptoms even at distal points from the infection site.

Limitations in protein detection technology and the presence of exogenous bovine and yeast proteins contained in the *Giardia* growth medium has meant that the study of its secretion profile has hitherto been proved challenging. The data presented in this thesis gives for the first time a secretion profile for both *Giardia* GS and WB isolates and demonstrates that most *Giardia*-secreted protein families are conserved between these two isolates suggesting a conserved mechanism of pathogenesis. Future research should focus on the functions of these proteins in the mediation of host-parasite interactions and on using these markers to elucidate the *Giardia* secretory pathways.

6.1.1/ A *Giardia* supernatant free from exogenous proteins.

During *in vitro* growth, serum proteins such as BSA bind to the trophozoite surface which interferes with the characterisation of the parasite secretome (Skarin et al., 2011). Moreover, it indicates that serum proteins can affect detection. To address these problems, I removed indicators and most of the exogenous proteins from the *Giardia* culture media (Chapter IV). Nevertheless, I found that a high number of giardial proteins were present in the *in vitro* growth supernatant and that proteomic analysis was possible.

Chapter III showed that *Giardia* growth media can affect cultured intestinal cells, potentially biasing the results of the analysis. The use of an exogenous

protein-free *Giardia* supernatant showed a lesser but still significant effect on intestinal cells and their water and ion transporters. This data demonstrates how factors released during *Giardia* growth can affect cultured intestinal epithelia, suggesting that effects from *Giardia*-secreted factors on intestinal epithelia are best investigated only in exogenous protein-free milieu.

6.1.2/ Disruption of intestinal cell morphology and AQP3/CFTR translocation caused by *Giardia* supernatant

To understand the causes of giardiasis and giardial diarrhoea, various studies have focused their investigations on the effects of trophozoite infection (Cotton et al., 2011, Baldi et al., 2009, Yu et al., 2008, Buret, 2007, Troeger et al., 2007, Roxstrom-Lindquist et al., 2005, Chin et al., 2002, Scott et al., 2002). Data presented in this thesis shows shrinkage and ruffling of intestinal cells in presence of *Giardia* supernatant for more than 8 hours suggesting a loss of turgidity in intestinal cells; it also showed translocation of intestinal AQP3 and CFTR following one hour incubation with the parasite supernatant suggesting that these two intestinal transporters may be involved in the decrease of cell turgidity.

The loss in cell size and ruffling effect were assessed by measurement of intestinal cell surface area and perimeter via Image J software and calculation of a ruffling index using the formula: $\frac{\mathbf{Perimeter}}{\sqrt{(\mathbf{Surface\ area})}}$. Assemblage B (GS isolate) may have a greater inhibiting effect on the cell turgidity whereas assemblage A (WB isolate) may cause more mucosal damage to the intestinal cells suggesting that these two isolates may have different pathophysiologic effects on the host intestinal epithelia (Chapter III). It is worth considering that the morphology assay was done manually using Image J software which is susceptible to observer bias. The measurement of the cell volume changes using fluorescence self-quenching or scanning conductance ion microscopy (SCIM), which uses the increase of access resistance in a micro-pipette in an electrolyte containing environment when it approaches a poorly conducting surface to measure the ion efflux, would afford greater subjectivity and potentially increase the intestinal water efflux disruption and therefore cell volume changes occurring during incubation with parasite supernatant (Korchev et al.,

2000). Attribution of the intestinal cell morphological changes observed to a specific protein secreted by the parasite or to another parasite-derived factor was not assessed in this thesis and should be further investigated. The nature and identity of the soluble mediators involved could be investigated through the use of size selection of filtration/fractionation of *Giardia* supernatant and by selective deactivation/inhibition of possible identified proteins in the supernatants.

Intestinal AQP3 and CFTR were translocated after one hour incubation with *Giardia* GS isolate supernatant and after eight hours incubation with *Giardia* WB supernatant suggesting that assemblage B (GS isolate) may be more bioactive than assemblage A (WB isolate) early in the infection. The translocation of AQP3 and CFTR suggests that *Giardia* may secrete factors activating the translocation of these transporters. Whether these transporters may be involved in the decrease of cell turgidity and inhibition of Isc previously observed is not known and should be investigated. The inhibition of these two transporters while repeating the intestinal cell morphology study under the same conditions would either confirm or deny this theory. Interestingly, the inhibition of intestinal cell electrophysiology using chloride secretion inhibitors (DIDS and Gly101, Chapter I, section 1.5.4) implicates CFTR directly in the disruption of intestinal epithelium homeostasis observed when *Giardia* supernatants are applied. However, whether CFTR is involved in the intestinal cell shrinkage or the decrease of cell turgidity was not directly tested and should be further investigated. The quantification of this translocation was not possible while using the fluorescent microscope. The use of a confocal XZ plane imaging microscope and a peripheral stain as a cell shape marker to investigate the density of vesicles close to the plasma membrane or flow cytometry would better enable evaluation of the translocation effect. AQP3 and CFTR are two of the main water and ion transporters in the intestine, along with SGLT-1. Yet, many more are present on the apical side of intestinal cells and may interact with *Giardia*-secreted factors. The use of probes for different ions such as potassium (K^+), sodium (Na^+), or chloride (Cl^-) to observe the intestinal ion transport during infection with *Giardia* supernatant will give a better understanding of which intestinal transporter may be involved in the host-*Giardia*-secreted factors interactions. As with morphology assay, the possible *Giardia*-secreted factors involved in the translocation of

intestinal transporters were not investigated in this thesis and should be performed by deactivating proteins contained in *Giardia* supernatant and repeating the protein translocation assays.

6.1.3/ *Giardia* secretion profile

Only four proteins have been proposed to be released and involved in *Giardia* trophozoites pathogenicity so far; Arginine deiminase (ADI), Ornithine carbamoyltransferase (OCT), enolase and more recently Cathepsin B-like protease (Cotton et al., 2014, Ringqvist et al., 2008). Elongation factor 1-α (EF-1α) has also been proposed to be released by trophozoites during *in vitro* growth (Skarin et al., 2011). However, which proteins are secreted by *Giardia* trophozoites has not been previously explored. This thesis demonstrates that various proteins are secreted by *Giardia* trophozoites; 55 % of *Giardia* secretome is composed of proteins secreted either by one or both isolates and the 45 % remaining are cell-surface proteins such as VSPs and HCMPs (Chapter V) (Table 6.1). Interestingly, ADI, enolase and OCT are present in both isolate pellet and supernatant as some of the most abundant proteins but not as proteins most likely to be secreted by trophozoites. Their presence in *Giardia* supernatant represents either a generalised "leakiness" of some G*iardia* cells or arising from cell damage and patent levels of cell death. This suggests that these proteins may not be secreted into the intestinal lumen by the parasite but may still have a role in *Giardia* pathogenesis nonetheless, as previously shown by Ringqvist and collaborator.

6.1.4/ Conserved and lineage-specific protein secretion

6.1.4.1/ Identification of proteins secreted only by *Giardia* WB or GS isolates

During the identification of putatively secreted *Giardia* trophozoites proteins, five were shown to be secreted in a strain-specific manner. WB isolate trophozoites secrete CXC-rich protein and Neurogenic locus notch protein precursor, PPIB and one EF-1α; and GS isolate trophozoite secrete A-type flavoprotein lateral transfer candidate (Chapter V). These differences may go some

Table 6. 1: *Giardia* trophozoite secreted proteins and their functions in *Giardia* pathogenesis. Secretion profile obtained after proteomic analysis. Proteins with no known function in *Giardia* pathogenicity but associated with other micro-organisms pathogenicity are shown by the ? in the function column.

Virulence factor	*Giardia* isolate	Function	References
Cell-surface protein			
VSPs	GS and WB	Cellular signalling	Touz et al., 2005; Hiltpold et al.,2000
		Antigenic variation	Adam, 2010 and 2001
		Immune evasion	Singer et al., 2001
		Potential protection against luminal proteases, oxygen and free-radicals	Nash, 2002
		Host-parasite interaction	Touz et al., 2005
HCMPs	GS and WB	Protection against proteolysis	Davids et al., 2006, Nash et al., 2002
		Protection against oxidative damage?	Roquejor et al., 2010
CXC-Rich protein	WB	Antigenic variation	Davids et al., 2006
Neurogenic locus notch protein precursor	WB	Parasite proliferation, differentiation?	Morais et al., 2013, Loverde et al., 2007
Secreted protein			
Cathepsin B	GS and WB	Degradation of intestinal IL-8 leading to attenuation of host pro-inflammatory response	Cotton et al., 2014
		Degradation of mucin	Paget and James, 1994
Hypothetical proteins	GS and WB	Undetermined	
Extracellular nuclease	GS and WB	Nutrients acquisition?	Kiedrowski et al., 2014, Seper et al., 2011
		Salvage of preformed purines?	Gottlieb et al., 1989
		Degradation of intestinal outer mucus layer?	Harms et al., 1998, Brandt et al., 1995
Tenascin	GS and WB	Parasite growth, differentiation, anti-adhesion? Environmental change signalling?	Chiquet-Ehrishman and Chiquet, 2003
A-type flavoprotein lateral transfer candidate	GS	Oxygen reductase activity	Di Matteo et al., 2008
PPIB	WB	Stimulation of host IFN-γ to control acute phases of infection?	Bell et al., 2006
EF-1α	WB	Disruption of intestinal cell polarity?	Tan et al., 2009
		Down-regulation of host inflammatory cell signalling?	Nandan and Reiner, 2005
		Parasite attachment to intestinal epithelia?	Granato et al., 2004

way towards explaining differences in pathogenicity between GS and WB isolates

The coding genes of all the proteins have orthologs in multiple *Giardia* genome strains (Fig 6.1) suggesting that these proteins may have evolutionary conserved functions. Interestingly, EF-1α does not form an ortholog with GS isolate but is shared by WB, DH and P15 isolates which suggests that assemblages A and E share virulence factors that have not been acquired by assemblage B trophozoites. This correlates with a previous genomic comparative study which showed that assemblage A and B share 77 % genome sequence identity whereas assemblage A and E share 87 % sequence identity (Franzen et al., 2013).

The fact that GS and WB strains release different proteins provide potential explanation for differences in pathogenicity between the two assemblages; arising from differences in host-pathogen interactions. This could explain the dissimilarities of effect on intestinal cells observed in this thesis (Chapter III). Questions regarding relative levels of activities as well as the trigger for the secretion of these proteins by the different isolates during *Giardia* infection should be further investigated.

6.1.4.2/ Secretion of formerly Hypothetical proteins and a conserved extracellular nuclease

The work presented shows that a substantial group of hypothetical proteins and an extracellular nuclease are secreted by both GS and WB isolates. Interestingly, conserved hypothetical proteins show some degree of inter-relatedness (chapter V); their coding genes also show high sequence similarities with their DH (assemblage A) and P15 (assemblage E) orthologs. This suggests potentially conserved but yet unknown functions between the three *Giardia* assemblages for these proteins. Interestingly, one hypothetical protein secreted by WB isolate is a lineage-specific assemblage (Fig 6.1 B.). Although no syntenic orthologs have been associated to it, this protein appears to have some degree of relatedness with the HCMP super family and is therefore more likely to be a cell-surface protein. This demonstrates the importance of a correct annotation of proteins and genes when working on proteomics or genomics; further investigations are required to verify this.

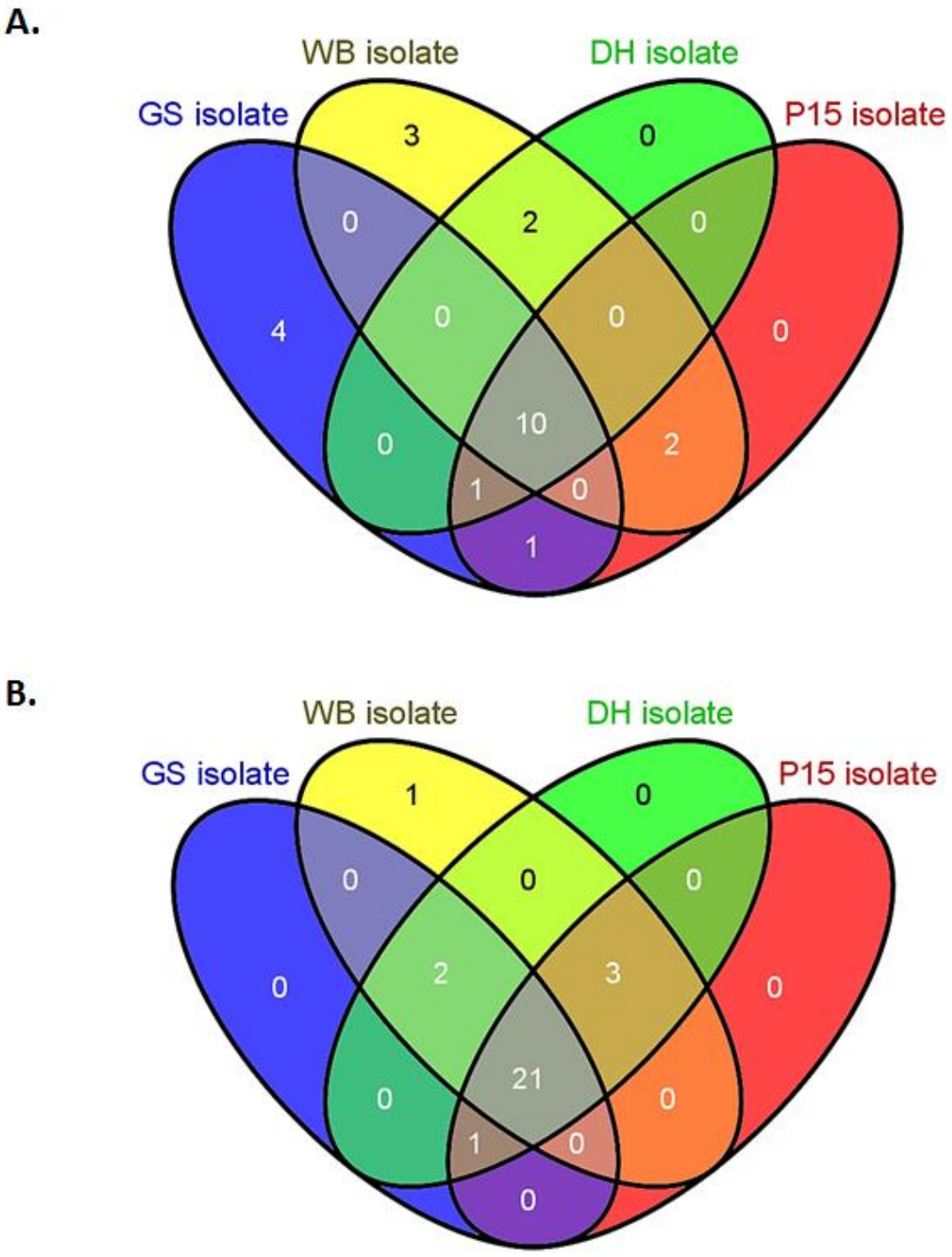


Fig 6. 1: Comparison of lineage-specific and orthologous proteins between assemblage B (GS isolate), A (WB and DH isolate) and E (P15 isolate). Orthologs were matched using GiardiaDB for the proteins identified as part of *Giardia* trophozoite secretome. Venn diagrams show the proteins having an orthologs in either one, or more assemblages, or the lineage-specific proteins. **A.** Orthologs and lineage-specific proteins located in the trophozoite surface: CXC-rich protein, VSPs, HCMPs, Neurogenic locus notch protein precursor. **B.** Orthologs and lineage-specific secreted proteins: Cathepsins B, Tenascins, A-type flavoprotein lateral transfer candidate, PPIB, EF-1α and hypothetical proteins.

The locus encoding for the extracellular nuclease is shared between GS (assemblage B), WB and DH (assemblage A) and P15 (assemblage E) isolates and highly conserved between these isolates strongly suggesting an evolutionary conserved function. Although DH and P15 isolates possess the coding gene for this protein, whether these two isolates also secrete this protein during *in vitro* growth or require environmental modifications to trigger its secretion should be investigated.

6.1.5/ HCMP super family

The HCMP superfamily is composed of Tenascins, VSPs and HCMPs. This thesis shows that this group of proteins is the most enriched in *Giardia* supernatant. VSPs being hypervariable, it is to be expected that they do not form syntenic orthologous pairs. Some VSPs identified showed true lineage-specificity for both GS and WB isolates as is to be expected from VSPs (Fig 6.1). This suggests that most VSPs conserved between isolates are not actual VSPs but misannotated HCMPs. The analysis of their function and encoding loci should determine what these proteins really are. The true lineage-specific VSPs may be considered as markers for specific *Giardia* strains and should be further investigated for a better understanding of *Giardia* isolate-specific pathogenicity.

I have shown that the Tenascin group is one of the most secreted groups of proteins by both *Giardia* GS and WB isolates. Transcript profiles during host-parasite interactions for these proteins were retrieved from the GiardiaDB database; and indicate that Tenascins conserved between GS and WB isolates have a decreasing transcript profile during host-parasite interactions suggesting a function in the early stage of the infection. These proteins also form orthologous groups with Tenascins in DH isolate (assemblage A) and P15 isolate (assemblage E) suggesting a conserved function in the early stage of giardial infection. Intriguingly, the Tenascins secreted only by WB isolate trophozoites show the opposite profile, their expression appears to increase throughout the infection. Similarly to the conserved Tenascins, these proteins form orthologous pairs with DH and P15 isolates also suggesting a certain conservation of function. The question of the secretion of these proteins by DH and P15 isolates trophozoites remains unknown

and should be investigated. Furthermore, the mechanisms triggering the expression and secretion of these Tenascins should be determined. Moreover, I observed a high degree of variation between Tenascin orthologs making then potential good markers for virulent and avirulent *Giardia* strains discrimination.

6.1.6/ *Giardia* Cathepsin B family

Along with the HCMP superfamily, the *G*CATB family contains some of the proteins the most secreted by both GS and WB isolates. Similarly to the Tenascins, their transcript profiles during host-parasite interactions were retrieved from GiardiaDB. The expression of most of the *G*CATBs conserved between GS and DH isolates increases during infection. This suggests that these proteins are required throughout the infection. However, the expression of one conserved *G*CATB decreases during infection suggesting that this *G*CATB may facilitate the attachment to the host intestinal epithelia for both isolates. All but one conserved *G*CATBs are also orthologous to Cathepsins B in DH and P15 isolates showing potential conserved functions for these proteins in *Giardia* pathogenicity.

Interestingly, the *G*CATBs only secreted by WB isolate trophozoites show either a constant transcript expression profile during host-pathogen interactions or a very high transcript expression profile early in the infection which greatly decreases later on during host-pathogen interactions. Here too, all *G*CATBs but one form orthologous pairs with Cathepsins B in DH and P15 isolates showing a potential mechanism by which these isolates may vary in term of pathogenicity. Whether environmental changes may trigger the expression and secretion of these proteins by the other isolates is not known yet and should be further investigated.

The genetic heterogeneity of these proteins between isolates is generally low compared with their orthologs (75 – 99 % similarities, data not shown). This suggests that *G*CATBs may have conserved functions between isolates. Yet, the fact that some are secreted during *in vitro* growth in one isolate and not the others may explain the variations in human pathogenicity between isolates. The question of their secretion in the early stage of infection in DH and P15 isolates should be investigated to verify if the secretion of several *G*CATBs is a pathogenic

characteristic of WB isolates or is shared with sub-assemblage AII and assemblage E.

6.2/ Summary and future work

Despite the recent advances in *Giardia* genomic, proteomics and pathogenesis, important questions still remain unanswered; including whether these newly identified secreted proteins play a role in *Giardia* pathogenicity. Further, the variation in protein secreted between isolates, and thus in pathogenicity is far from being understood. Similarly, the effects of these secreted virulence factors on host intestinal epithelia are still under debate and investigation.

In this thesis, I have presented microscopic and proteomic data to identify *Giardia*-secreted virulence factors and their effects on human intestinal epithelia. Morphological changes were shown to occur after only 8 hours. Variations in pathogenic effects on intestinal cells were also observed between the two *Giardia* isolates which could be explained by the difference in GS and WB secretion profiles. Yet, the question of these effects being caused by *Giardia*-secreted proteins could not be answered during my studies and should be further investigated. In the future, the use of inhibitors of targeted *Giardia*-secreted proteins will determine which of these proteins are responsible for the morphological changes occurring in intestinal cells during host-parasite interactions. Moreover, the use of "flat" organoids as a model of small intestinal tissue (Jabaji et al., 2013) allowing the inoculation on the apical side will give a better understanding of how *Giardia* – secreted virulence factors cause the disruption of the intestinal epithelia.

This thesis has exploited proteomic data from *Giardia* and compared their coding sequences to identify orthologs and sequence similarities. Several proteins identified were either shown to be secreted or paired with orthologs, such as FixW protein and VSPs respectively, when it should have not been the case. This illustrates the pitfalls of genomic comparative analysis and gene predictions in complete eukaryote genomes as it can lead to misannotation of proteins. Unfortunately, this was the case for a few proteins in this analysis and shows the importance of deeper analyses on these predicted features for annotated genes.

Nevertheless, proteomic analysis of *Giardia*-secreted proteins showed interesting results and potential new markers and mechanisms for *Giardia* pathogenicity.

In this analysis, proteins that are released and enriched to a substantial concentration outside of human infective *Giardia* trophozoites have been identified. It is most likely that these proteins are ones that have been classically secreted or shed. Just three groups form the majority of these proteins: *G*CATBs and the HCMP superfamily which encode known virulence factors and the SFHPs which are an intriguing new group awaiting functional studies. The elucidation of the SFHP group is likely to dramatically increase our understanding of the pathogenic mechanisms underlying giardiasis at molecular level. The genes encoding these proteins are among the most heterogeneous of all genes between assemblages and their probable role in interaction with the host and luminal environment ensure that their variation is likely to be associated with their survival ability in the host and thus to virulence. Demonstration of the positive selection pressure through examination of parameters such as dN/dS ratios should confirm this. Correlation of variation within assemblages at these loci with strain virulence is the essential next step for their use in diagnosis of virulent strains, risk assessment, informing therapeutic options and disease prognosis.

In this thesis, I showed that an extracellular nuclease, *G*CATBs and Tenascins were potentially involved in various pathogenic mechanisms (Table 6.1). The fact that both extracellular nuclease and *G*CATBs can be involved in the degradation of the intestinal mucus layers and that both *G*CATBs and Tenascins can be associated with intestinal intracellular junction disruption suggests concerted action between these proteins. Based on my analysis of the proteomic data, I propose a model of pathogenic mechanism for *Giardia*-induced damage to the enteric epithelium and by extension for giardiasis (Fig 6.2). The extracellular nuclease may contribute to reducing the viscosity of mucus while *G*CATBs may degrade mucins and disrupt the intestinal intracellular junctions. Lastly, Tenascins may be involved in maintaining the intestinal cell separation by attaching to the EGF receptors present at the surface of intestinal cells. Taken together overtime, this would lead to the apoptosis of these disrupted cells which have become isolated in the epithelium. Further

experimental investigations will be required to verify this as the mechanism of pathogenesis by *Giardia*.

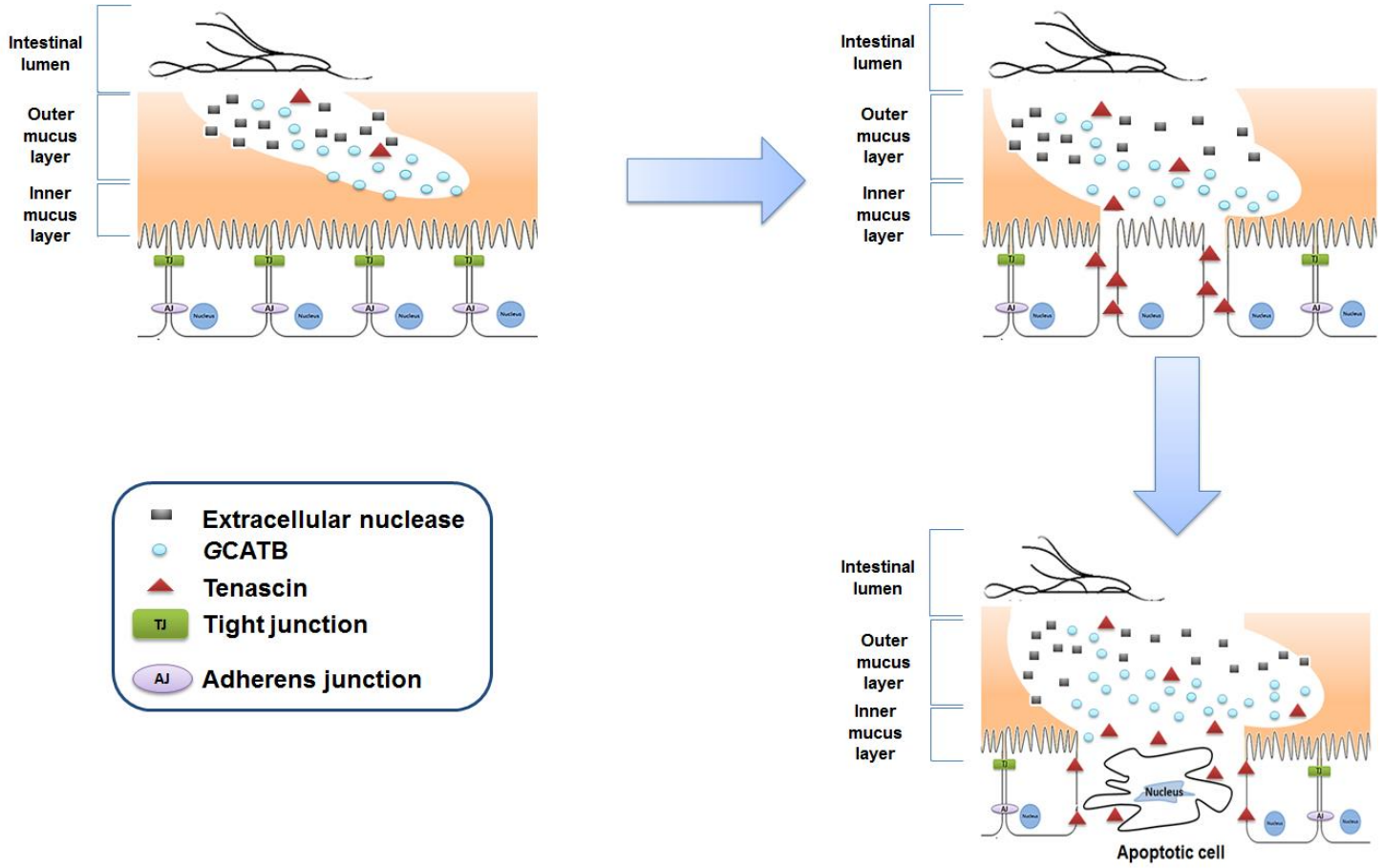


Fig 6.2: Proposed roles of *G*CATB, Tenascin and extracellular nuclease in host-parasite interactions. Extracellular nuclease ( ) may contribute to reducing the viscosity of the intestinal outer mucus layer, then *G*CATB ( ) may degrade mucins and disrupt intracellular junction. Finally, Tenascins ( ) may maintain intestinal cells apart by attaching to the EGF receptors present at the surface of intestinal cells which could over time lead to the apoptosis of these isolated intestinal cells.

Appendices

Appendix I:

Part of this work has been presented at conferences:

Abstract:

1. British Society of Immunology. June 2011. Norwich, UK. Enterocyte function is compromised by a soluble mediator from *Giardia*. **Audrey Dubourg**, Suha Al-Naimi. John Winpenny, Paul Hunter, Kevin Tyler. ***Poster presentation.***
2. Third Seminar of European PhD students on Water and Health. June 2011. Cannes, France. Enterocyte function is compromised by a soluble mediator from *Giardia*. Suha Al-Naimi, **Audrey Dubourg**, John Winpenny, Paul Hunter, Kevin Tyler. ***Oral presentation.***
3. British Society for Parasitology meeting. April 2012. Glasgow, UK. Enterocyte function is compromised by a *Giardia*-secreted mediator. **Audrey Dubourg**, Suha Al-Naimi. John Winpenny, Dong Xia, Jonathan Wastling, Paul Hunter, Kevin Tyler. ***Oral presentation.***
4. Cell, Development and Biomedicine seminar. May 2012. Norwich, UK. *Giardia*-released factors and their effect on enterocyte function. **Audrey Dubourg.** ***Oral presentation.***
5. Fifth International Conference on Anaerobic Protists. September 2012. Los Angeles, USA. Enterocyte function is compromised by a *Giardia*-secreted mediator. **Audrey Dubourg**, Suha Al-Naimi. John Winpenny, Dong Xia, Jonathan Wastling, Paul Hunter, Kevin Tyler. ***Poster presentation.***

Appendix II:

Paired *t*-test compared to controls for difference in length and width of Caco-2 cells after incubation with different concentrations of *Giardia* supernatant for 24 hours. A. Compared to incubation in DMEM. B. Compared to incubation in TYI-S-33.

A.

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	L_WB1.5000 - L_DMEM	-1.74710	4.02706	.72328	-3.22423	-.26996	-2.416	30	.022
Pair 2	W_WB1.5000 - W_DMEM	-2.27032	4.10484	.73725	-3.77599	-.76465	-3.079	30	.004
Pair 3	L_WB1.250 - L_DMEM	-8.63613	4.01366	.72087	-10.10835	-7.16391	-11.980	30	.000
Pair 4	W_WB1.250 - W_DMEM	-5.81903	3.24143	.58218	-7.00800	-4.63007	-9.995	30	.000
Pair 5	L_GS1.5000 - L_DMEM	-5.62741	4.74949	.91404	-7.50624	-3.74857	-6.157	26	.000
Pair 6	W_GS1.5000 - W_DMEM	-2.85370	3.77305	.72612	-4.34627	-1.36114	-3.930	26	.001
Pair 7	L_TYIS33 - L_DMEM	-1.08222	5.58667	1.07515	-3.29223	1.12779	-1.007	26	.323
Pair 8	W_TYS33 - W_DMEM	-1.32593	3.79605	.73055	-2.82759	.17574	-1.815	26	.081
Pair 9	L_GS1.250 - L_DMEM	-4.76935	3.98634	.71597	-6.23155	-3.30716	-6.661	30	.000
Pair 10	W_GS1.250 - W_DMEM	-2.38452	5.06927	.91047	-4.24394	-.52509	-2.619	30	.014

B.

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	L_WB1.5000 - L_TYIS33	-.97889	4.57605	.88066	-2.78911	.83133	-1.112	26	.277
Pair 2	W_WB1.5000 - W_TYS33	-.91148	3.42727	.65958	-2.26727	.44430	-1.382	26	.179
Pair 3	L_WB1.250 - L_TYIS33	-7.49741	4.21701	.81156	-9.16560	-5.82922	-9.238	26	.000
Pair 4	W_WB1.250 - W_TYS33	-4.20148	2.76416	.53196	-5.29495	-3.10802	-7.898	26	.000
Pair 5	L_GS1.5000 - L_TYIS33	-4.54519	4.16868	.80226	-6.19426	-2.89611	-5.665	26	.000
Pair 6	W_GS1.5000 - W_TYS33	-1.52778	3.31203	.63740	-2.83797	-.21758	-2.397	26	.024
Pair 7	L_DMEM - L_TYIS33	1.08222	5.58667	1.07515	-1.12779	3.29223	1.007	26	.323
Pair 8	W_DMEM - W_TYS33	1.32593	3.79605	.73055	-.17574	2.82759	1.815	26	.081
Pair 9	L_GS1.250 - L_TYIS33	-3.86556	4.44625	.85568	-5.62443	-2.10668	-4.518	26	.000
Pair 10	W_GS1.250 - W_TYS33	-.51444	3.42808	.65974	-1.87055	.84166	-.780	26	.443

Appendix III:

Non-parametric Mann-Whitney U-test compared to control (DMEM) for difference in surface area and perimeter of Caco-2 cells after incubation with different concentrations of *Giardia* supernatant for A. 1 hr. B. 8 hrs. C. 24 hrs.

A.

Test Statistics[a]

		SurfaceArea	Perimeter
DMEM GS1.250	Mann-Whitney U	63168	63112.5
	Wilcoxon W	131433	131377.5
	Z	-0.176	-0.196
	Asymp. Sig. (2-tailed)	0.86	0.845
DMEM GS1.1000	Mann-Whitney U	71076	71682.5
	Wilcoxon W	147321	139947.5
	Z	-0.291	-0.09
	Asymp. Sig. (2-tailed)	0.771	0.928
DMEM WB1.250	Mann-Whitney U	53705	56479.5
	Wilcoxon W	121970	124744.5
	Z	-1.848	-0.776
	Asymp. Sig. (2-tailed)	0.065	0.438
DMEM WB1.1000	Mann-Whitney U	69298	69154
	Wilcoxon W	140929	140785
	Z	-0.15	-0.199
	Asymp. Sig. (2-tailed)	0.881	0.842

a. Grouping Variable: GiardiaSupernatant

B.

Test Statistics[a]

		SurfaceArea	Perimeter
DMEM GS1.250	Mann-Whitney U	128157	131062
	Wilcoxon W	317577	320482
	Z	-4.633	-4.088
	Asymp. Sig. (2-tailed)	0	0
DMEM GS1.1000	Mann-Whitney U	124773	126950.5
	Wilcoxon W	295593	297770.5
	Z	-3.978	-3.552
	Asymp. Sig. (2-tailed)	0	0
DMEM WB1.250	Mann-Whitney U	116425.5	123959.5
	Wilcoxon W	259805.5	267339.5
	Z	-3.453	-1.879
	Asymp. Sig. (2-tailed)	0.001	0.06
DMEM WB1.1000	Mann-Whitney U	118127	119999.5
	Wilcoxon W	279723	281595.5
	Z	-4.597	-4.223
	Asymp. Sig. (2-tailed)	0	0

a. Grouping Variable: GiardiaSupernatant

C.

Test Statistics[a]

		SurfaceArea	Perimeter
DMEM GS1.250	Mann-Whitney U	272681.5	274346
	Wilcoxon W	605201.5	606866
	Z	-2.612	-2.421
	Asymp. Sig. (2-tailed)	0.009	0.015
DMEM GS1.1000	Mann-Whitney U	278887.5	287560
	Wilcoxon W	613858.5	622531
	Z	-2.019	-1.026
	Asymp. Sig. (2-tailed)	0.043	0.305
DMEM WB1.250	Mann-Whitney U	211163	215497.5
	Wilcoxon W	474338	478672.5
	Z	-2.156	-1.549
	Asymp. Sig. (2-tailed)	0.031	0.121
DMEM WB1.1000	Mann-Whitney U	217251	203756
	Wilcoxon W	480426	466931
	Z	-1.548	-3.427
	Asymp. Sig. (2-tailed)	0.122	0.001

a. Grouping Variable: GiardiaSupernatant

Appendix IV:

Panel data approach performed on Stata software (v 13.1) to compare the ruffling index of CaCo-2 cells under different *Giardia* supernatant treatment to the control (DMEM) at different times. A. 1 hour. B. 8 hours. C. 24 hours.

A.

	Ruffling coef.	Std. Err.	p-value
GS1.250	0.035	0.069	
GS1.1000	0.025	0.067	
constant	5.150	0.048	
			0.870
WB1.250	-0.182	0.067	
WB1.1000	-0.001	0.064	
constant	5.146	0.046	
			0.012

B.

	Ruffling coef.	Std. Err.	p-value
GS1.250	-0.011	0.053	
GS1.1000	-0.021	0.053	
constant	5.336	0.039	
			0.924
WB1.250	0.194	0.058	
WB1.1000	0.035	0.057	
constant	5.325	0.041	
			0.001

C.

	Ruffling coef.	Std. Err.	p-value
GS1.250	0.031	0.049	
GS1.1000	0.091	0.049	
constant	5.350	0.036	
			0.164
WB1.250	-0.048	0.051	
WB1.1000	0.185	0.051	
constant	5.351	0.035	
			0.000

Appendix V:

Non-parametric Mann-Whitney U-test compared to control (DMEM) for difference in AQP3 and SGLT-1 expression after incubation with different concentrations of *Giardia* supernatant for A. 1 hr. B. 8 hrs. C. 24 hrs.

A.

Test Statistics[b]

		AQP3	SGLT1
DMEM GS1.250	Mann-Whitney U	.000	.000
	Wilcoxon W	6.000	6.000
	Z	-2.087	-2.087
	Asymp. Sig. (2-tailed)	.037	.037
	Exact Sig. [2*(1-tailed Sig.)]	.100	.100
DMEM GS1.500	Mann-Whitney U	.000	3.000
	Wilcoxon W	6.000	9.000
	Z	-2.087	-.696
	Asymp. Sig. (2-tailed)	.037	.487
	Exact Sig. [2*(1-tailed Sig.)]	.100	.700
DMEM GS1.1000	Mann-Whitney U	.000	3.000
	Wilcoxon W	6.000	9.000
	Z	-2.087	-.696
	Asymp. Sig. (2-tailed)	.037	.487
	Exact Sig. [2*(1-tailed Sig.)]	.100	.700
DMEM GS1.5000	Mann-Whitney U	3.000	3.000
	Wilcoxon W	9.000	9.000
	Z	-.696	-.696
	Asymp. Sig. (2-tailed)	.487	.487
	Exact Sig. [2*(1-tailed Sig.)]	.700	.700
DMEM WB1.250	Mann-Whitney U	.000	3.000
	Wilcoxon W	6.000	9.000
	Z	-2.087	-.696
	Asymp. Sig. (2-tailed)	.037	.487
	Exact Sig. [2*(1-tailed Sig.)]	.100	.700
DMEM WB1.500	Mann-Whitney U	.000	3.000
	Wilcoxon W	6.000	9.000
	Z	-2.087	-.696
	Asymp. Sig. (2-tailed)	.037	.487
	Exact Sig. [2*(1-tailed Sig.)]	.100	.700
DMEM WB1.1000	Mann-Whitney U	.000	.000
	Wilcoxon W	6.000	6.000
	Z	-2.087	-2.087
	Asymp. Sig. (2-tailed)	.037	.037
	Exact Sig. [2*(1-tailed Sig.)]	.100	.100
DMEM WB1.5000	Mann-Whitney U	.000	.000
	Wilcoxon W	6.000	6.000
	Z	-2.087	-2.087
	Asymp. Sig. (2-tailed)	.037	.037
	Exact Sig. [2*(1-tailed Sig.)]	.100	.100

a. Not corrected for ties.

b. Grouping Variable: GiardiaSP

B.

Test Statisticsb

		AQP3	SGLT1
DMEM GS1.250	Mann-Whitney U	3.000	.000
	Wilcoxon W	9.000	6.000
	Z	-.696	-2.087
	Asymp. Sig. (2-tailed)	.487	.037
	Exact Sig. [2*(1-tailed Sig.)]	.700	.100
DMEM GS1.500	Mann-Whitney U	3.000	.000
	Wilcoxon W	9.000	6.000
	Z	-.696	-2.087
	Asymp. Sig. (2-tailed)	.487	.037
	Exact Sig. [2*(1-tailed Sig.)]	.700	.100
DMEM GS1.1000	Mann-Whitney U	3.000	.000
	Wilcoxon W	9.000	6.000
	Z	-.696	-2.087
	Asymp. Sig. (2-tailed)	.487	.037
	Exact Sig. [2*(1-tailed Sig.)]	.700	.100
DMEM GS1.5000	Mann-Whitney U	3.000	3.000
	Wilcoxon W	9.000	9.000
	Z	-.696	-.696
	Asymp. Sig. (2-tailed)	.487	.487
	Exact Sig. [2*(1-tailed Sig.)]	.700	.700
DMEM WB1.250	Mann-Whitney U	3.000	3.000
	Wilcoxon W	9.000	9.000
	Z	-.696	-.696
	Asymp. Sig. (2-tailed)	.487	.487
	Exact Sig. [2*(1-tailed Sig.)]	.700	.700
DMEM WB1.500	Mann-Whitney U	3.000	3.000
	Wilcoxon W	9.000	9.000
	Z	-.696	-.696
	Asymp. Sig. (2-tailed)	.487	.487
	Exact Sig. [2*(1-tailed Sig.)]	.700	.700
DMEM WB1.1000	Mann-Whitney U	3.000	.000
	Wilcoxon W	9.000	6.000
	Z	-.696	-2.087
	Asymp. Sig. (2-tailed)	.487	.037
	Exact Sig. [2*(1-tailed Sig.)]	.700	.100
DMEM WB1.5000	Mann-Whitney U	3.000	.000
	Wilcoxon W	9.000	6.000
	Z	-.696	-2.087
	Asymp. Sig. (2-tailed)	.487	.037
	Exact Sig. [2*(1-tailed Sig.)]	.700	.100

a. Not corrected for ties.

b. Grouping Variable: GiardiaSP

C.

Test Statisticsb

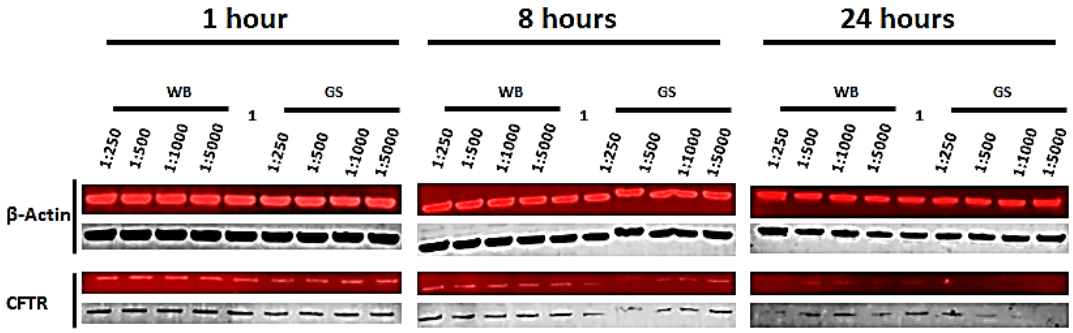
		AQP3	SGLT1
DMEM GS1.250	Mann-Whitney U	3.000	.000
	Wilcoxon W	9.000	6.000
	Z	-.696	-2.087
	Asymp. Sig. (2-tailed)	.487	.037
	Exact Sig. [2*(1-tailed Sig.)]	.700	.100
DMEM GS1.500	Mann-Whitney U	3.000	3.000
	Wilcoxon W	9.000	9.000
	Z	-.696	-.696
	Asymp. Sig. (2-tailed)	.487	.487
	Exact Sig. [2*(1-tailed Sig.)]	.700	.700
DMEM GS1.1000	Mann-Whitney U	3.000	3.000
	Wilcoxon W	9.000	9.000
	Z	-.696	-.696
	Asymp. Sig. (2-tailed)	.487	.487
	Exact Sig. [2*(1-tailed Sig.)]	.700	.700
DMEM GS1.5000	Mann-Whitney U	3.000	3.000
	Wilcoxon W	9.000	9.000
	Z	-.696	-.696
	Asymp. Sig. (2-tailed)	.487	.487
	Exact Sig. [2*(1-tailed Sig.)]	.700	.700
DMEM WB1.250	Mann-Whitney U	3.000	3.000
	Wilcoxon W	9.000	9.000
	Z	-.696	-.696
	Asymp. Sig. (2-tailed)	.487	.487
	Exact Sig. [2*(1-tailed Sig.)]	.700	.700
DMEM WB1.500	Mann-Whitney U	.000	3.000
	Wilcoxon W	6.000	9.000
	Z	-2.087	-.696
	Asymp. Sig. (2-tailed)	.037	.487
	Exact Sig. [2*(1-tailed Sig.)]	.100	.700
DMEM WB1.1000	Mann-Whitney U	3.000	.000
	Wilcoxon W	9.000	6.000
	Z	-.696	-2.087
	Asymp. Sig. (2-tailed)	.487	.037
	Exact Sig. [2*(1-tailed Sig.)]	.700	.100
DMEM WB1.5000	Mann-Whitney U	.000	.000
	Wilcoxon W	6.000	6.000
	Z	-2.087	-2.087
	Asymp. Sig. (2-tailed)	.037	.037
	Exact Sig. [2*(1-tailed Sig.)]	.100	.100

a. Not corrected for ties.

b. Grouping Variable: GiardiaSP

Appendix VI:

CFTR bands obtained using the AbCam monoclonal Mouse anti-CFTR antibody. Membranes were incubated overnight with the antibody diluted at 1:750. Bands were obtained for the three time points 1, 8 and 24 hours. However, 24hours, not all the bands were visible for CFTR.



Appendix VII:

Correlation between data obtained with Orbitrap and QeXactive MS via spearman correlation test using SPSS. **A.** GS Pellet replicates; **B.** GS Supernatant replicates; **C.** WB Pellet replicates; **D.** WB Supernatant replicates.

A.

***Giardia* replicate and technique**	GS P Orbitrap	GS P1 QeXactive	GS P2 QeXactive
GS P Orbitrap	1	0.681	0.682
GS P1 QeXactive	0.681	1	0.999
GS P2 QeXactive	0.682	0.999	1

B.

***Giardia* replicate and technique**	GS S Orbitrap	GS S1 QeXactive	GS S2 QeXactive
GS S Orbitrap	1	0.371	0.222
GS S1 QeXactive	0.371	1	0.602
GS S2 QeXactive	0.222	0.602	1

C.

***Giardia* replicate and technique**	WB P Orbitrap	WB P1 QeXactive	WB P2 QeXactive
WB P Orbitrap	1	0.652	0.651
WB P1 QeXactive	0.652	1	0.999
WB P2 QeXactive	0.651	0.999	1

D.

***Giardia* replicate and technique**	WB S Orbitrap	WB S1 QeXactive	WB S2 QeXactive
WB S Orbitrap	1	0.365	0.335
WB S1 QeXactive	0.365	1	0.977
WB S2 QeXactive	0.335	0.977	1

Appendix VIII:

Correlation between the first and second datasets obtained with Orbitrap and QeXactive MS via spearman correlation test using SPSS. **A.** GS Pellet replicates; **B.** GS Supernatant replicates; **C.** WB Pellet replicates; D. WB Supernatant replicates.

A.

Giardia replicate and technique	GS P Orbitrap 2012-2013	GS P 1 Orbitrap 2014	GS P 2 Orbitrap 2014	GS P 3 Orbitrap 2014	GS P 1 QeXactive 2012-2013	GS P 2 QeXactive 2012-2013	GS P 1 QeXactive 2014	GS P 2 QeXactive 2014	GS P 3 QeXactive 2014
GS P Orbitrap 2012-2013	1	0.588	0.584	0.583	0.681	0.682	0.574	0.571	0.569
GS P 1 Orbitrap 2014	0.588	1	0.996	0.993	0.703	0.703	0.862	0.858	0.859
GS P 2 Orbitrap 2014	0.584	0.996	1	0.998	0.7	0.7	0.859	0.858	0.859
GS P 3 Orbitrap 2014	0.583	0.993	0.998	1	0.697	0.697	0.857	0.855	0.857
GS P 1 QeXactive 2012-2013	0.681	0.703	0.7	0.697	1	0.999	0.716	0.712	0.71
GS P 2 QeXactive 2012-2013	0.682	0.703	0.7	0.697	0.999	1	0.716	0.712	0.71
GS P 1 QeXactive 2014	0.574	0.862	0.859	0.857	0.716	0.716	1	0.997	0.997
GS P 2 QeXactive 2014	0.571	0.858	0.858	0.855	0.712	0.712	0.997	1	0.999
GS P 3 QeXactive 2014	0.569	0.859	0.859	0.857	0.71	0.71	0.997	0.999	1

B.

Giardia replicate and technique	GS S Orbitrap 2012-2013	GS S 1 Orbitrap 2014	GS S 2 Orbitrap 2014	GS S 3 Orbitrap 2014	GS S 1 QeXactive 2012-2013	GS S 2 QeXactive 2012-2013	GS S 1 QeXactive 2014	GS S 2 QeXactive 2014	GS S 3 QeXactive 2014
GS S Orbitrap 2012-2013	1	0.588	0.561	0.568	0.371	0.222	0.546	0.518	0.526
GS S 1 Orbitrap 2014	0.588	1	0.991	0.993	0.61	0.364	0.835	0.836	0.84
GS S 2 Orbitrap 2014	0.561	0.991	1	0.998	0.638	0.381	0.83	0.847	0.846
GS S 3 Orbitrap 2014	0.568	0.993	0.998	1	0.631	0.376	0.832	0.845	0.846
GS S 1 QeXactive 2012-2013	0.371	0.61	0.638	0.631	1	0.602	0.635	0.68	0.671
GS S 2 QeXactive 2012-2013	0.222	0.364	0.381	0.376	0.602	1	0.403	0.435	0.426
GS S 1 QeXactive 2014	0.546	0.835	0.83	0.832	0.635	0.403	1	0.981	0.986
GS S 2 QeXactive 2014	0.518	0.836	0.847	0.845	0.68	0.435	0.981	1	0.994
GS S 3 QeXactive 2014	0.526	0.84	0.846	0.846	0.671	0.426	0.986	0.994	1

C.

Giardia replicate and technique	WB P Orbitrap 2012-2013	WB P 1 Orbitrap 2014	WB P 2 Orbitrap 2014	WB P 3 Orbitrap 2014	WB P 1 QeXactive 2012-2013	WB P 2 QeXactive 2012-2013	WB P 1 QeXactive 2014	WB P 2 QeXactive 2014	WB P 3 QeXactive 2014
WB P Orbitrap 2012-2013	1	0.723	0.73	0.7	0.652	0.651	0.7	0.706	0.709
WB P 1 Orbitrap 2014	0.723	1	0.996	0.981	0.675	0.676	0.877	0.877	0.874
WB P 2 Orbitrap 2014	0.73	0.996	1	0.98	0.673	0.674	0.877	0.879	0.877
WB P 3 Orbitrap 2014	0.7	0.981	0.98	1	0.669	0.67	0.86	0.861	0.859
WB P 1 QeXactive 2012-2013	0.652	0.675	0.673	0.669	1	0.999	0.686	0.686	0.692
WB P 2 QeXactive 2012-2013	0.651	0.676	0.674	0.67	0.999	1	0.686	0.686	0.692
WB P 1 QeXactive 2014	0.7	0.877	0.877	0.86	0.686	0.686	1	0.997	0.997
WB P 2 QeXactive 2014	0.706	0.877	0.879	0.861	0.686	0.686	0.997	1	0.998
WB P 3 QeXactive 2014	0.709	0.874	0.877	0.859	0.692	0.692	0.997	0.998	1

D.

Giardia replicate and technique	WB S Orbitrap 2012-2013	WB S 1 Orbitrap 2014	WB S 2 Orbitrap 2014	WB S 3 Orbitrap 2014	WB S 1 QeXactive 2012-2013	WB S 2 QeXactive 2012-2013	WB S 1 QeXactive 2014	WB S 2 QeXactive 2014	WB S 3 QeXactive 2014
WB S Orbitrap 2012-2013	1	0.461	0.454	0.451	0.365	0.335	0.483	0.475	0.473
WB S 1 Orbitrap 2014	0.461	1	0.997	0.997	0.411	0.37	0.836	0.844	0.839
WB S 2 Orbitrap 2014	0.454	0.997	1	0.999	0.419	0.377	0.836	0.848	0.841
WB S 3 Orbitrap 2014	0.451	0.997	0.999	1	0.416	0.375	0.835	0.847	0.841
WB S 1 QeXactive 2012-2013	0.365	0.411	0.419	0.416	1	0.977	0.445	0.459	0.454
WB S 2 QeXactive 2012-2013	0.335	0.37	0.377	0.375	0.977	1	0.402	0.413	0.41
WB S 1 QeXactive 2014	0.483	0.836	0.836	0.835	0.445	0.402	1	0.992	0.99
WB S 2 QeXactive 2014	0.475	0.844	0.848	0.847	0.459	0.413	0.992	1	0.996
WB S 3 QeXactive 2014	0.473	0.839	0.841	0.841	0.454	0.41	0.99	0.996	1

Appendix IX:

Secretion profile for each isolate obtained via Orbitrap MS and QeXactive MS. Proteins identified in all datasets are highlighted in grey, proteins identified via both MS technique for one isolate are highlighted in green. Annotated proteins are in red and hypothetical proteins are in blue. **A.** GS Orbitrap secretion profile ; **B.** GS QeXactive secretion profile; **C.** WB Orbitrap secretion profile; **D.** WB QeXactive secretion profile.

A.

Assemblage B										Assemblage A syntenic orthologs		
		Orbitrap										
Protein Accesion Number	Protein Description	P Expression-iBAQ	SP Expression-iBAQ	SP/P Ratio	GO function	GO location	SignalP[a]	MW[b] (Da)	TM[c]	Accession	Description	Ortholog identified in WB isolate dataset (Y/N) [d]
GL50581_3370	Protein disulfide isomerase PDI5	1.19E+06	2.57E+06	2.166598627	cell redox homeostasis	Cytoplasm	1	12558	0	GL50803_8064	Protein disulfide isomerase PDI5	Y (QeX and Orb)
GL50581_4139	Hypothetical protein	1.04E+06	1.31E+06	1.26274105	cell redox homeostasis	Cytoplasm	-	8586	-	GL50803_8394	hypothetical protein	Y (QeX and Orb)
GL50581_4133	Hypothetical protein	8.02E+05	1.25E+06	1.557456644	oxidation-reduction process	Cytoplasm	-	14486	-	GL50803_5810	hypothetical protein	Y (QeX and Orb)
GL50581_665	Hypothetical protein	-	1.01E+06	-	-	Undetermined	-	25064	-	GL50803_10315	Qb-SNARE 5 (previously Hypothetical pro	Y (QeX)
GL50581_73	Hypothetical protein	6.21E+05	8.18E+05	1.317835962	-	Undetermined	-	11478	-	GL50803_4239	Hypothetical protein	Y (QeX and Orb)
GL50581_1982	Tenascin-like protein	5.28E+05	7.63E+05	1.446030613	protein binding	Cytoplasm	0.99	61197	0	GL50803_95162	Tenascin-like	Y (QeX and Orb)
GL50581_4057	Tenascin precursor	2.57E+05	5.28E+05	2.056165222	protein binding	Cytoplasm	0.99	27781	0	GL50803_10330	Tenascin precursor	Y (QeX and Orb)
GL50581_438	Cathepsin B precursor	1.82E+05	3.56E+05	1.956460847	proteolysis	Cytoplasm	0.78	34066	0	GL50803_16468	Cathepsin B precursor	Y (QeX and Orb)
GL50581_832	Thioredoxin reductase	3.10E+05	3.52E+05	1.135417566	removal of superoxide radicals/oxidation-reduction process	Cytoplasm	-	33877	-	GL50803_9827	Thioredoxin reductase	Y (QeX and Orb)
GL50581_4499	Hypothetical protein	2.07E+05	3.38E+05	1.62819289	-	Undetermined	-	30003	-	GL50803_11354	Hypothetical protein	Y (QeX and Orb)
GL50581_2531	FKBP-type peptidyl-prolyl cis-trans isomeras	2.46E+05	2.94E+05	1.195097034	protein folding	Cytoplasm	-	11857	-	GL50803_10450	FKBP-type peptidyl-prolyl cis-trans isomerase	Y (QeX and Orb)
GL50581_4316	Tenascin precursor	1.02E+05	2.44E+05	2.396286096	protein binding	Cytoplasm	0.98	53521	0	GL50803_8687	Tenascin precursor	Y (QeX and Orb)
GL50581_1500	Macrophage migration inhibitory factor	1.10E+05	2.39E+05	2.163885764	-	Undetermined	-	12465	-	GL50803_12091	Macrophage migration inhibitory factor	Y (QeX and Orb)
GL50581_2954	VSP	1.05E+05	2.29E+05	2.179510702	-	Plasma membrane	-	17746	-	-	-	-
GL50581_1902	VSP	1.11E+05	1.62E+05	1.458110942	protein phosphorylation	Plasma membrane	-	96279	-	DHA2_150066	VSP	N
GL50581_1791	VSP	7.80E+04	1.46E+05	1.867654873	-	Plasma membrane	-	64182	-	-	-	-
GL50581_2110	Hypothetical protein	3.78E+04	1.33E+05	3.504543452	-	Undetermined	-	5900	-	GL50803_28477	Hypothetical protein	Y
GL50581_3909	Hypothetical protein	5.29E+04	1.07E+05	2.023756307	-	Undetermined	-	11343	-	GL50803_1875	Hypothetical protein	Y (QeX and Orb)
GL50581_1626	A-type flavoprotein lateral transfer candidate	9.22E+04	1.06E+05	1.154189146	oxidoreductase activity/hydrolase activity	Cytoplasm	-	46451	-	GL50803_10358	A-type flavoprotein lateral transfer candidate	Y (QeX and Orb)
GL50581_4524	Branched-chain amino acid aminotransferase lateral transfer candidate	7.49E+04	1.03E+05	1.37167137	metabolic process / branched-chain amino acid metabolic process	Cytoplasm	-	39188	-	GL50803_6184	Branched-chain amino acid aminotransferase lateral transfer candidate	Y (QeX and Orb)
GL50581_1838	Hypothetical protein	8.98E+04	9.86E+04	1.098898557	-	Undetermined	-	10368	-	GL50803_27918	Hypothetical protein	Y (QeX and Orb)
GL50581_2767	Hypothetical protein	8.82E+04	9.75E+04	1.105929765	-	Undetermined	-	11335	-	GL50803_5258	Hypothetical protein	Y (QeX and Orb)
GL50581_3607	Extracellular nuclease, putative	2.18E+04	7.62E+04	3.499864395	nuclease activity	Cytoplasm	1	30603	0	GL50803_8742	Extracellular nuclease, putative	Y (QeX and Orb)
GL50581_3436	Hypothetical protein	5.16E+04	6.64E+04	1.285235208	-	Undetermined	-	15150	-	GL50803_5604	Hypothetical protein	Y (QeX and Orb)
GL50581_3575	Tenascin-37	1.63E+04	6.48E+04	3.979611758	protein binding	Cytoplasm	0.99	62017	0	GL50803_16477	Tenascin-37	Y (QeX and Orb)
GL50581_4180	Tenascin precursor	1.81E+04	6.19E+04	3.416486052	protein binding	Cytoplasm	1	60692	0	GL50803_113038	hypothetical protein	Y (QeX and Orb)
GL50581_2256	Heat shock protein HSP 90-alpha	-	6.09E+04	-	protein folding/response to stress	Cytoplasm	-	42835	-	GL50803_98054	Heat shock protein HSP 90-alpha	Y (QeX and Orb)
GL50581_727	High cysteine membrane protein Group 1	-	4.58E+04	-	protein binding	Plasma membrane	0.99	73115	1	GL50803_7715	High cysteine membrane protein Group 1	Y (QeX and Orb)
GL50581_1808	RAN binding protein 1	4.10E+04	4.54E+04	1.106396617	intracellular transport	Signaling systems	-	18859	-	GL50803_16969	RAN binding protein 1	Y (QeX and Orb)
GL50581_2771	Protein phosphatase methylesterase-1	1.55E+04	4.46E+04	2.878943487	protein demethylation	Cytoplasm	-	40352	-	GL50803_16180	Protein phosphatase methylesterase-1	Y (QeX and Orb)
GL50581_2467	20S proteasome alpha subunit 4	3.76E+04	4.01E+04	1.067013298	ubiquitin-dependent protein catabolic process / proteolysis involved in cellular protein catabolic process	Cytoplasm	-	23281	-	GL50803_15099	20S proteasome alpha subunit 4	Y (QeX and Orb)
GL50581_1093	Hypothetical protein	3.33E+04	3.58E+04	1.072261722	-	Undetermined	-	14543	-	GL50803_14299	Hypothetical protein	Y (QeX and Orb)
GL50581_72	Dynein light chain	-	3.50E+04	-	-	Cytoskeleton	-	12101	-	GL50803_4236	Dynein light chain	Y (QeX and Orb)
GL50581_3484	Cyclin-dependent kinases regulatory subunit	1.82E+04	3.28E+04	1.798703468	cell cycle	Cytoplasm	-	10358	-	GL50803_2661	Cyclin-dependent Kinases regulatory subunit	Y (QeX and Orb)
GL50581_1	VSP	1.15E+04	2.76E+04	2.395475303	-	Plasma membrane	1	41559	1	-	-	-
GL50581_2622	Variant-specific surface protein	5.18E+03	2.55E+04	4.918394374	-	Plasma membrane	1	43681	0	GL50803_12063	Variant-specific surface protein	Y (QeX and Orb)
GL50581_2036	Cathepsin B precursor	-	2.51E+04	-	proteolysis	Cytoplasm	1	33027	0	GL50803_15564	Cathepsin B precursor	Y (QeX and Orb)
GL50581_235	Variant-specific surface protein	5.68E+03	2.42E+04	4.254710342	protein binding	Plasma membrane	0.95	50615	0	GL50803_9276	Variant-specific surface protein	Y (QeX and Orb)
GL50581_1690	NH3-dependent NAD synthetase	9.12E+03	2.34E+04	2.562445014	NAD biosynthetic process	Cytoplasm	-	31499	-	GL50803_31530	NH3-dependent 0D synthetase	Y (QeX)
GL50581_3698	Hypothetical protein	9.88E+03	2.18E+04	2.204103561	-	Undetermined	-	34091	-	GL50803_11129	Hypothetical protein	Y (QeX and Orb)
GL50581_1005	HAM1 protein	-	1.88E+04	-	deoxyribonucleoside triphosphate catabolic process/Nucleotide metabolic process/nucleoside-triphosphate diphosphatase activity	Cytoplasm	-	21421	-	GL50803_7511	HAM1 protein	N

GL50581_1828	Proteasome subunit beta type 3	4.44E+03	1.83E+04	4.120360009	proteolysis involved in cellular protein catabolic process	Cytoplasm	-	23051	-	GL50803_13756	Proteasome subunit beta type 3	Y (QeX and Orb)
GL50581_3117	Kinase, NEK	1.61E+04	1.81E+04	1.128074925	protein phosphorylation	Cytoplasm	-	33394	-	GL50803_24321	Kinase, NEK	Y (QeX and Orb)
GL50581_1186	Peptidyl-prolyl cis-trans isomerase B precursor	9.51E+03	1.81E+04	1.904021969	protein folding	Cytoplasm	1	21269	0	GL50803_17000	Peptidyl-prolyl cis-trans isomerase B precursor	Y (QeX and Orb)
GL50581_1655	Nucleotide-binding head-stalk protein, putative	-	1.49E+04	-	-	Cytoplasm	-	183844	-	GL50803_15995	Nucleotide-binding head-stalk protein, putative	Y (QeX and Orb)
GL50581_2952	Hypothetical protein	-	1.40E+04	-	-	Undetermined	-	13916	-	GL50803_11342	Hypothetical protein	N
GL50581_2137	Protein 21.1	-	1.38E+04	-	protein binding	Cytoplasm	-	24507	-	GL50803_6344	Protein 21.1	Y (QeX)
GL50581_3514	High cysteine membrane protein Group 1	9.28E+03	1.38E+04	1.484534126	protein binding	Cytoplasm	0.98	139382	1	GL50803_11309	High cysteine membrane protein Group 1	Y (QeX and Orb)
GL50581_1356	H-SHIPPO 1	-	1.21E+04	-	-	Undetermined	-	22142	-	GL50803_9148	H-SHIPPO 1	Y (Orb)
GL50581_3457	High cysteine protein	-	1.17E+04	-	cell adhesion / metalloendopeptidase activity	Plasma membrane	-	74175	-	-	-	-
GL50581_3401	VSP	4.40E+03	1.10E+04	2.501522001	-	Plasma membrane	1	38355	0	-	-	-
GL50581_789	Hypothetical protein	-	1.07E+04	-	protein binding	Cytoplasm	-	37664	-	GL50803_8789	hypothetical protein	Y (QeX and Orb)
GL50581_3818	VSP	-	1.06E+04	-	protein binding	Plasma membrane	-	61825	-	-	-	-
GL50581_2225	20S proteasome alpha subunit 2	-	9.97E+03	-	ubiquitin-dependent protein catabolic process/proteolysis involved in cellular protein catabolic process	Cytoplasm	-	27539	-	GL50803_11434	20S proteasome alpha subunit 2	Y (QeX and Orb)
GL50581_3056	FixW protein, putative	-	9.86E+03	-	oxidation-reduction process	Cytoplasm	0.06	23091	0	GL50803_23888	FixW protein, putative	Y (QeX and Orb)
GL50581_2776	Hypothetical protein	-	9.12E+03	-	-	Undetermined	0.5	28134	6	GL50803_12830	hypothetical protein	Y (QeX and Orb)
GL50581_352	Hypothetical protein	-	8.89E+03	-	-	Undetermined	-	16921	-	GL50803_16522	Hypothetical protein	Y (QeX and Orb)
GL50581_1462	Hypothetical protein	6.29E+03	8.32E+03	1.385329186	protein folding	Cytoplasm	-	31203	1	GL50803_16873	Hypothetical protein	N
GL50581_2416	Hypothetical protein	7.40E+03	8.32E+03	1.124017195	-	Undetermined	1	171076	1	GL50803_5813	hypothetical protein	N
GL50581_2872	Hypothetical protein	-	7.84E+03	-	-	Undetermined	1	30773	1	GL50803_16424	Hypothetical protein	Y (QeX and Orb)
GL50581_3608	Hypothetical protein	-	7.40E+03	-	protein folding	Cytoplasm	-	13790	1	GL50803_112870	Hypothetical protein	N
GL50581_1896	Hypothetical protein	2.51E+03	6.73E+03	2.677244317	-	Undetermined	-	31254	-	GL50803_112932	Hypothetical protein	
GL50581_316	Hypothetical protein	-	6.29E+03	-	-	Undetermined	0.93	44965	0	GL50803_16996	hypothetical protein	Y (QeX and Orb)
GL50581_3924	Serine-pyruvate aminotransferase	-	6.15E+03	-	metabolic process/transaminase activity	Cytoplasm	-	41399	-	GL50803_3313	Serine-pyruvate aminotransferase	Y (QeX and Orb)
GL50581_680	Lysosomal acid phosphatase precursor	2.24E+03	6.06E+03	2.70354486	acid phosphatase activity	Lysosome	0.91	45898	0	GL50803_7556	Lysosomal acid phosphatase precursor	Y (QeX and Orb)
GL50581_4139	Hypothetical protein	-	6.01E+03	-	-	Undetermined	-	20433	-	GL50803_86960	hypothetical protein	Y (QeX and Orb)
GL50581_3418	Nudix hydrolase, putative	3.57E+03	5.56E+03	1.488051079	hydrolase activity	Cytoplasm	-	13784	-	GL50803_16623	hypothetical protein	N (Orb)
GL50581_1487	VSP	-	4.37E+03	-	-	Plasma membrane	0.91	30423	0	-	-	N
GL50581_2318	Cathepsin B precursor	-	4.30E+03	-	cysteine-type peptidase activity	Cytoplasm	1	34583	0	GL50803_17516	Cathepsin B precursor	Y (QeX and Orb)
GL50581_572	VSP	1.48E+03	4.04E+03	2.735426398	protein binding	Plasma membrane	-	67208	-	-	-	-
GL50581_1528	PLP dependent enzymes class III	-	3.75E+03	-	-	Undetermined	-	-	-	GL50803_15041	hypothetical protein, enzyme with a TIM-barrel fold	Y (QeX and Orb)
GL50581_322	High cysteine membrane protein Group 4	-	3.68E+03	-	protein binding	Plasma membrane	0.61	87867	1	-	-	N
GL50581_1565	Nif3-related protein	-	2.95E+03	-	-	Undetermined	-	23958	-	GL50803_4355	Nif3-related protein	N
GL50581_1879	Hypothetical protein	-	2.59E+03	-	-	Undetermined	-	21886	-	GL50803_2555	hypothetical protein	Y (QeX)
GL50581_1340	Phosphoglycolate phosphatase	-	2.52E+03	-	metabolic process	Cytoplasm	-	25403	-	GL50803_14841	Phosphoglycolate phosphatase	Y (QeX and Orb)
GL50581_3466	Alpha-1 guanin	2.11E+03	2.47E+03	1.171033934	cytoskeleton organization	Cytoskeleton	-	33641	-	GL50803_4213	Alpha-1 guanin	Y (QeX)
GL50581_3259	Hypothetical protein	5.33E+02	2.43E+03	4.552205509	-	Undetermined	-	32676	-	GL50803_14213	hypothetical protein	Y (QeX and Orb)
GL50581_3788	Coiled-coil protein				-	Undetermined	-	97042	-	GL50803_14963	hypothetical protein	Y (QeX and Orb)
GL50581_587	5'-methylthioadenosine nucleosidase, S-adenosylhomocysteine nucleosidase	-	2.39E+03	-	nucleoside metabolic process/nucleoside catabolic process /L-methionine salvage from methylthioadenosine	Cytoplasm	-	28831	-	GL50803_20195	5-methylthioadenosine nucleosidase, S-adenosylhomocysteine nucleosidase	Y (QeX and Orb)
GL50581_1694	Hypothetical protein	-	2.30E+03	-	-	Undetermined	-	32730	-	GL50803_6896	hypothetical protein	Y (QeX and Orb)
GL50581_2090	Hypothetical protein	-	2.18E+03	-	-	Undetermined	-	33903	-	GL50803_8423	Hypothetical protein	N
GL50581_2012	Hypothetical protein	1.70E+03	2.01E+03	1.179324363	ATP hydrolysis coupled proton transport	Signaling systems	-	23367	-	GL50803_13603	Hypothetical protein	Y (QeX and Orb)
GL50581_3224	Kinase, NEK	-	1.61E+03	-	protein serine / threonine kinase activity / protein phosphorylation	Cytoplasm	-	79509	-	-	-	-
GL50581_3469	Hypothetical protein	-	1.48E+03	-	-	Undetermined	-	20431	-	GL50803_16234	hypothetical protein	Y (QeX)
GL50581_3518	Hypothetical protein	-	1.42E+03	-	protein binding	Undetermined	-	43680	-	GL50803_24528	Hypothetical protein	N
GL50581_342	WD-repeat protein	-	1.32E+03	-	protein binding	Cytoplasm	-	44805	-	GL50803_16526	WD-repeat protein	Y (QeX and Orb)
GL50581_4293	Hypothetical protein	-	1.29E+03	-	-	Undetermined	-	58856	-	GL50803_9899	hypothetical protein	Y (QeX and Orb)
GL50581_3210	Sentrin	9.08E+02	1.08E+03	1.077741204	protein binding	Undetermined	-	11495	-	GL50803_2760	Sentrin	Y (QeX)
GL50581_3748	Hypothetical protein		9.78E+02		-	Undetermined	-	35538	-	GL50803_16527	Hypothetical protein	N
GL50581_250	High cysteine membrane protein Group 6	-	8.57E+02	-	protein binding	Plasma membrane	1	66359	1	GL50803_41942	High cysteine membrane protein Group 6	Y (QeX and Orb)
GL50581_2882	Vacuolar ATP synthase subunit C	-	8.12E+02	-	ATP hydrolysis coupled proton transport	Signaling systems	-	56271	-	GL50803_87058	Vacuolar ATP synthase subunit C	Y (QeX and Orb)
GL50581_1330	Hypothetical protein	-	6.81E+02	-	-	Undetermined	-	28912	-	GL50803_15847	Hypothetical protein	Y (QeX and Orb)
GL50581_3019	Protein 21.1	-	6.17E+02	-	protein binding	Undetermined	-	81573	-	GL50803_12028	Protein 21.1	N
GL50581_1160	Hypothetical protein	-	5.33E+02	-	protein binding	Undetermined	-	89016	-	GL50803_14509	Hypothetical protein	N
GL50581_395	High cysteine membrane protein Group 2	-	4.81E+02	-	-	Plasma membrane	1	75129	1	GL50803_16842	High cysteine membrane protein Group 2	N
GL50581_791	UDP-N-acetylglucosamine pyrophosphorylase	-	4.25E+02	-	nucleotidyltransferase activity	Cytoplasm	-	48552	-	GL50803_16217	UDP-N-acetylglucosamine pyrophosphorylase	N
GL50581_4421	Hypothetical protein	-	4.20E+02	-	-	Undetermined	-	59573	-	GL50803_13133	hypothetical protein	Y (QeX)
GL50581_719	Nitroreductase family protein	-	2.07E+02	-	oxidoreductase activity	Cytoplasm	-	19486	-	GL50803_15307	Nitroreductase family protein	Y (QeX)
GL50581_4522	Protein 21.1	-	1.98E+02	-	protein binding	Undetermined	-	28275	-	GL50803_14133	Protein 21.1	Y (QeX and Orb)
GL50581_2985	Long-flagella protein, Kinase, CMGC RCK	-	9.19E+01	-	protein phosphorylation	Cytoplasm	-	60278	-	GL50803_14004	Long-flagella protein, Kinase, CMGC RCK	Y (QeX and Orb)

[a] Signal peptide
[b] Molecular Weight
[c] Transmembrane Domain
[d] QeX: QeXactive dataset; Orb: Orbitrap dataset

B.

Protein Accesion Number	Assemblage B									Assemblage A syntenic orthologs		
	QeXactive											
	Protein Description	P Expression-iBAQ	SP Expression-iBAQ	SP/P Ratio	GO function	GO location	Signal[a]	MW[b] (Da)	TM[c]	Accesion	Description	Ortholog identified in WB isolate dataset (Y/N)[d]
GL50581_3386	Protein disulfide isomerase PDI5	6.56E+07	1.48E+08	2.25296	glycerol ether metabolic process / cell redox homeostasis	Cytoplasm	1	12643	0	GL50803_14670	Protein disulfide isomerase PDI3	Y (QeX and Orb)
GL50581_727	High cysteine membrane protein Group 1	-	2.54E+06	-	protein binding	Plasma membrane	0.99	73115	1	GL50803_7715	High cysteine membrane protein Group 1	Y (QeX and Orb)
GL50581_111	Hypothetical protein	-	1.28E+06	-	-	Undetermined	-	776415	-	GL50803_35341	Hypothetical protein	N
GL50581_3593	FKBP-type peptidyl-prolyl cis-trans isomerase	-	1.08E+06	-	protein folding	Cytoplasm	1	23933	0	GL50803_101339	FKBP-type peptidyl-prolyl cis-trans isomerase	Y (QeX and Orb)
GL50581_2355	Prefoldin subunit 3, putative	-	8.32E+05	-	protein folding	Cytoplasm	-	22034	-	GL50803_7030	Prefoldin subunit 3, putative	Y (QeX and Orb)
GL50581_2318	Cathepsin B precursor	-	7.81E+05	-	cysteine-type peptidase activity	Cytoplasm	1	34583	0	GL50803_17516	Cathepsin B precursor	Y (QeX and Orb)
GL50581_123	5'-methylthioadenosine nucleosidase, S-adenosylhomocysteine nucleosidase	-	4.19E+05	-	nucleoside metabolic process	Cytoplasm	-	28319	-	GL50803_4059	5'-methylthioadenosine nucleosidase, S-adenosylhomocysteine nucleosidase	Y (QeX and Orb)
GL50581_1356	H-SHIPPO 1	-	3.48E+05	-	-	Undetermined	-	22142	-	GL50803_9148	H-SHIPPO 1	Y (Orb)
GL50581_2090	Hypothetical protein	-	2.69E+05	-	-	Undetermined	-	33903	-	GL50803_8423	Hypothetical protein	N
GL50581_66	Hypothetical protein	-	2.56E+05	-	protein folding	Cytoplasm	-	15445	-	GL50803_1383	Hypothetical protein	Y (QeX only)
GL50581_1485	Phosphoglycolate phosphatase	-	1.76E+05	-	metabolic process	Cytoplasm	-	29657	-	GL50803_14546	Phosphoglycolate phosphatase	Y (QeX and Orb)
GL50581_1325	Hypothetical protein	-	1.64E+05	-	-	Undetermined	0.14	22047	1	GL50803_14458	Hypothetical protein	Y (Orb only)
GL50581_1340	Phosphoglycolate phosphatase	-	1.46E+05	-	metabolic process	Cytoplasm	-	25403	-	GL50803_14841	Phosphoglycolate phosphatase	Y (QeX and Orb)
GL50581_1671	Alpha-5 giardin	-	1.29E+05	-	cytoskeleton organization/calcium ion binding	Cytoskeleton	-	34049	-	GL50803_7797	Alpha-5 giardin	Y (QeX and Orb)
GL50581_250	High cysteine membrane protein Group 6	-	9.23E+04	-	protein binding	Plasma membrane	1	66359	1	GL50803_41942	High cysteine membrane protein Group 6	Y (QeX and Orb)
GL50581_2768	High cysteine protein	-	6.85E+04	-	-	Plasma membrane	-	36885	-	GL50803_113987	High cysteine membrane protein Group 3	Y (QeX and Orb)
GL50581_2781	Peptide methionine sulfoxide reductase msrA	-	6.10E+04	-	oxidation-reduction process/protein repair	Cytoplasm	-	22158	-	GL50803_4946	Peptide methionine sulfoxide reductase msrA	Y (QeX and Orb)
GL50581_3358	Hypothetical protein	-	4.66E+04	-	hydrolase activity	Undetermined	-	28388	-	GL50803_15311	Hypothetical protein	N
GL50581_3210	Sentrin	-	3.67E+04	-	protein binding	Undetermined	-	11495	-	GL50803_7760	Sentrin	Y (QeX)
GL50581_1527	Cleavage stimulation factor 50K chain	-	2.49E+04	-	protein binding	Undetermined	-	46821	-	GL50803_15042	Cleavage stimulation factor 50K chain	Y (QeX)
GL50581_490	Hypothetical protein	-	2.00E+04	-	-	Undetermined	-	58187	-	GL50803_6709	Hypothetical protein	N
GL50581_1548	Hypothetical protein	-	9.72E+03	-	ATP hydrolysis coupled proton transport	Signaling systems	-	23081	-	GL50803_10868	Hypothetical protein	Y (QeX and Orb)
GL50581_690	Hypothetical protein	-	8.22E+03	-	nucleic acid binding	Cytoplasm	-	55531	-	GL50803_13075	Hypothetical protein	Y (QeX and Orb)
GL50581_1753	Hypothetical protein	-	6.38E+03	-	-	Undetermined	-	367622	-	GL50803_113873	Hypothetical protein	N
GL50581_3106	Hypothetical protein	-	3.81E+03	-	protein binding	Undetermined	-	154345	-	GL50803_9807	Hypothetical protein	N
GL50581_1777	RNA binding protein, putative	-	3.54E+03	-	nucleic acid binding	Cytoplasm	-	116152	-	GL50803_92031	RNA binding protein, putative	Y (QeX)
GL50581_1737	Hypothetical protein	-	1.13E+03	-	-	Undetermined	0.7	39136	1	GL50803_13668	hypothetical protein	Y (QeX)
GL50581_3457	High cysteine protein	3.43E+04	3.48E+05	10.1259	cell adhesion / metalloendopeptidase activity	Plasma membrane	-	74175	-	-	-	-
GL50581_572	VSP	6.50E+04	5.39E+05	8.2832	protein binding	Plasma membrane	-	67208	-	-	-	-
GL50581_3056	FixW protein, putative	1.92E+05	1.24E+06	6.46307	oxidation-reduction process	Cytoplasm	0.06	23091	0	GL50803_23888	FixW protein, putative	Y (QeX and Orb)
GL50581_1387	WD-40 repeat protein	3.44E+04	2.22E+05	6.45943	protein binding	Cytoplasm	-	114460	-	GL50803_16264	WD-40 repeat protein	Y (QeX and Orb)
GL50581_3818	VSP	7.85E+04	5.06E+05	6.44979	protein binding	Plasma membrane	-	61825	-	-	-	-
GL50581_1	VSP	2.36E+05	1.46E+06	6.16205	-	Plasma membrane	1	41559	1	-	-	-
GL50581_235	Variant-specific surface protein	3.25E+05	1.78E+06	5.48508	protein binding	Plasma membrane	0.95	50615	0	GL50803_9276	Variant-specific surface protein	Y (QeX and Orb)
GL50581_2622	Variant-specific surface protein	3.62E+05	1.94E+06	5.35467	-	Plasma membrane	1	43681	0	GL50803_12063	Variant-specific surface protein	Y (QeX and Orb)
GL50581_3575	Tenascin-37	8.35E+05	4.03E+06	4.83096	protein binding	Cytoplasm	0.99	62017	0	GL50803_16477	Tenascin-37	Y (QeX and Orb)
GL50581_3190	Spindle pole protein, putative	1.39E+05	6.35E+05	4.55977	-	Cytoskeleton	-	60105	-	GL50803_16013	Spindle pole protein, putative	Y (QeX)
GL50581_3607	Extracellular nuclease, putative	1.03E+06	4.57E+06	4.43619	nuclease activity	Cytoplasm	1	30603	0	GL50803_8742	Extracellular nuclease, putative	Y (QeX and Orb)
GL50581_4057	Tenascin precursor	4.66E+06	2.02E+07	4.34229	protein binding	Cytoplasm	0.99	27781	0	GL50803_10330	Tenascin precursor	Y (QeX and Orb)
GL50581_3748	Hypothetical protein	7.75E+04	3.28E+05	4.23873	-	Undetermined	-	35528	-	GL50803_10527	Hypothetical protein	N
GL50581_2036	Cathepsin B precursor	1.10E+06	3.95E+06	3.58961	proteolysis	Cytoplasm	1	33027	0	GL50803_15564	Cathepsin B precursor	Y (QeX and Orb)
GL50581_2367	Hypothetical protein	6.25E+05	2.12E+06	3.38648	protein binding	Cytoplasm	-	45126	-	GL50803_9099	Hypothetical protein	Y (QeX and Orb)
GL50581_1730	Hypothetical protein	2.63E+05	7.58E+05	2.88486	-	Undetermined	-	15356	-	GL50803_4624	hypothetical protein	Y (QeX)

GL50581_3484	Cyclin-dependent Kinases regulatory subunit	1.14E+06	3.20E+06	2.80306	cell cycle	Cytoplasm	-	10358	-	GL50803_2661	Cyclin-dependent Kinases regulatory subunit	Y (QeX and Orb)
GL50581_2767	Hypothetical protein	3.93E+06	1.09E+07	2.78092	-	Undetermined	-	11335	-	GL50803_5258	Hypothetical protein	Y (QeX and Orb)
GL50581_1902	VSP	1.58E+05	4.27E+05	2.70029	protein phosphorylation	Plasma membrane	-	96279	-	DHA2_150066	VSP	N
GL50581_4180	Tenascin precursor	1.20E+06	3.14E+06	2.62093	protein binding	Cytoplasm	1	60692	0	GL50803_113038	hypothetical protein	Y (QeX and Orb)
GL50581_1186	Peptidyl-prolyl cis-trans isomerase B precursor	1.04E+06	2.70E+06	2.59707	protein folding	Cytoplasm	1	21269	0	GL50803_17000	Peptidyl-prolyl cis-trans isomerase B precursor	Y (QeX and Orb)
GL50581_3401	VSP	4.93E+05	1.22E+06	2.48425	-	Plasma membrane	1	38355	0	-	-	-
GL50581_85	High cysteine membrane protein Group 4	1.13E+05	2.69E+05	2.38565	protein binding	Cytoplasm	0.99	85154	1	-	-	-
GL50581_349	Hypothetical protein	3.57E+04	8.29E+04	2.32251	protein transport	Signaling systems	-	20869	-	GL50803_24947	hypothetical protein	Y (QeX and Orb)
GL50581_3386	Protein disulfide isomerase PDI5	6.56E+07	1.48E+08	2.25296	glycerol ether metabolic process / cell redox homeostasis	Cytoplasm	1	12643	0	GL50803_14670	Protein disulfide isomerase PDI3	Y (QeX and Orb)
GL50581_3619	Cathepsin B precursor	9.52E+05	2.11E+06	2.21892	proteolysis	Cytoplasm	1	33265	0	-	-	-
GL50581_3909	Hypothetical protein	8.91E+06	1.85E+07	2.07789	-	Undetermined	-	11343	-	GL50803_1875	hypothetical protein	Y (QeX and Orb)
GL50581_4139	Hypothetical protein	3.65E+07	7.57E+07	2.07538	cell redox homeostasis	Cytoplasm	-	8586	-	GL50803_8394	hypothetical protein	Y (QeX and Orb)
GL50581_4133	Hypothetical protein	5.71E+07	1.18E+08	2.06309	oxidation-reduction process	Cytoplasm	-	14486	-	GL50803_5810	hypothetical protein	Y (QeX and Orb)
GL50581_1500	Macrophage migration inhibitory factor	1.94E+07	3.93E+07	2.02248	-	Undetermined	-	12465	-	GL50803_12091	Macrophage migration inhibitory factor	Y (QeX and Orb)
GL50581_1626	A-type flavoprotein lateral transfer candidate	4.35E+06	8.70E+06	2.00174	oxidoreductase activity / hydrolase activity	Cytoplasm	-	46451	-	GL50803_10358	A-type flavoprotein lateral transfer candidate	Y (QeX and Orb)
GL50581_352	Hypothetical protein	1.15E+06	2.21E+06	1.92883	-	Undetermined	-	16921	-	GL50803_16522	hypothetical protein	Y (QeX and Orb)
GL50581_2954	VSP	1.18E+07	2.25E+07	1.90956	-	Plasma membrane	-	17746	-	-	-	-
GL50581_438	Cathepsin B precursor	9.63E+06	1.79E+07	1.86131	proteolysis	Cytoplasm	0.78	34066	0	GL50803_16468	Cathepsin B precursor	Y (QeX and Orb)
GL50581_3979	Hypothetical protein	3.30E+04	5.97E+04	1.81083	-	Undetermined	-	33265	-	GL50803_14748	Hypothetical protein	N
GL50581_4316	Tenascin precursor	6.22E+06	1.10E+07	1.77053	protein binding	Cytoplasm	0.98	53521	0	GL50803_8687	Tenascin precursor	Y (QeX and Orb)
GL50581_4162	Hypothetical protein	4.97E+05	8.75E+05	1.76017	protein folding	Cytoplasm	-	13497	-	DHA2_152684	Hypothetical protein	N
GL50581_3387	Hypothetical protein	2.07E+06	3.62E+06	1.7504	-	Undetermined	-	17492	-	GL50803_8250	hypothetical protein	Y (QeX and Orb)
GL50581_2893	Aspartate aminotransferase, cytoplasmic	1.58E+06	2.72E+06	1.72045	biosynthetic process	Cytoplasm	-	47645	-	GL50803_91056	Aspartate aminotransferase, cytoplasmic	Y (QeX and Orb)
GL50581_2606	Hypothetical protein, similar to AMMECR1	1.98E+06	3.40E+06	1.71691	-	Undetermined	-	21608	-	GL50803_9349	hypothetical protein, similar to AMMECR1	Y (QeX and Orb)
GL50581_1838	Hypothetical protein	9.86E+06	1.68E+07	1.70234	-	Undetermined	-	10368	-	GL50803_27918	hypothetical protein	Y (QeX and Orb)
GL50581_3459	Hypothetical protein	4.17E+04	7.00E+04	1.67867	protein binding	Cytoplasm	-	59837	-	GL50803_6139	hypothetical protein	Y (QeX)
GL50581_3133	Hypothetical protein	5.56E+06	9.23E+06	1.65844	tubulin complex assembly	Cytoskeleton	-	12799	-	GL50803_1649	hypothetical protein	Y (QeX and Orb)
GL50581_4115	Fructose-bisphosphate aldolase	5.77E+07	9.55E+07	1.65392	carbohydrate metabolic process / glycolysis/fructose 1,6-bisphosphate metabolic process	Cytoplasm	-	35155	-	GL50803_11043	Fructose-bisphosphate aldolase	Y (QeX and Orb)
GL50581_3156	Proteasome subunit beta type 2	1.80E+05	2.95E+05	1.64247	proteolysis involved in cellular protein catabolic process carbohydrate metabolic	Cytoplasm	-	18958	-	GL50803_3209	Proteasome subunit beta type 2	Y (QeX and Orb)
GL50581_275	1,4-alpha-glucan branching enzyr	3.74E+05	6.10E+05	1.6309	process/glycogen biosynthetic process	Cytoplasm	-	91696	-	GL50803_15823	1,4-alpha-glucan branching enzyme	Y (QeX and Orb)
GL50581_1982	Tenascin-like protein	2.97E+07	4.77E+07	1.60702	protein binding	Cytoplasm	0.99	61197	0	GL50803_95162	Tenascin-like	Y (QeX and Orb)
GL50581_4194	FixW protein, putative	7.22E+06	1.16E+07	1.60176	-	Cytoplasm	-	14961	-	GL50803_6289	FixW protein, putative	Y (QeX and Orb)
GL50581_1095	Brix domain containing protein	9.82E+04	1.55E+05	1.58325	-	Undetermined	-	37698	-	GL50803_6835	Brix domain containing protein	Y (QeX and Orb)
GL50581_1019	Peptidyl-prolyl cis-trans isomerase B precursor	4.81E+07	7.48E+07	1.55515	protein folding	Cytoplasm	-	18030	-	GL50803_17163	Peptidyl-prolyl cis-trans isomerase B precursor	Y (QeX and Orb)
GL50581_3370	Protein disulfide isomerase PDI5	5.98E+07	9.29E+07	1.55381	glycerol ether metabolic process /cell redox homeostasis	Cytoplasm	1	12558	0	GL50803_8064	Protein disulfide isomerase PDI5	Y (QeX and Orb)
GL50581_2531	FKBP-type peptidyl-prolyl cis-trans isomerase	2.31E+07	3.58E+07	1.55181	protein folding	Cytoplasm	-	11857	-	GL50803_10450	FKBP-type peptidyl-prolyl cis-trans isomerase	Y (QeX and Orb)

[a] **Signal peptide**
[b] **Molecular Weight**
[c] **Transmembrane Domain**
[d] **QeX: QeXactive dataset, Orb: Orbitrap dataset**

C.

		Assemblage A								Assemblage B syntenic orthologs		
		Orbitrap										
Protein Accession Number	Protein Description	P Expression-iBAQ	SP Expression-iBAQ	SP/P Ratio	GO function	GO location	Signal P[a]	MW[b] (Da)	TM[c]	Accession	Description	Ortholog identified in GS isolate dataset (Y/N) [d]
GL50803_33279	VSP	6.18E+06	3.61E+06	0.58459	protein binding	Plasma membrane	0.98	75981	1	-	-	-
GL50803_5810	hypothetical protein	1.32E+06	5.03E+05	0.37975	oxidation-reduction process	Cytoplasm	-	14500	-	GL50581_4133	hypothetical protein	Y (QeX and Orb)
GL50803_16779	Cathepsin B precursor	1.03E+05	3.58E+05	3.48645	proteolysis	Cytoplasm	1	32761	0	GL50581_78	Cathepsin B precursor	Y (QeX and Orb)
GL50803_16322	Neurogenic locus Notch protein precursor	4.46E+05	2.97E+05	0.66533	calcium ion binding / protein binding	Cytoplasm	0.99	63070	0	GL50581_31	Neurogenic locus Notch protein precursor	Y (QeX and Orb)
GL50803_17163	Peptidyl-prolyl cis-trans isomerase B precursor	9.04E+05	2.91E+05	0.32235	protein folding	Cytoplasm	-	18044	-	GL50581_1019	Peptidyl-prolyl cis-trans isomerase B precursor	Y (QeX and Orb)
GL50803_8394	hypothetical protein	8.61E+05	2.77E+05	0.32177	cell redox homeostasis	Cytoplasm	-	9378	-	GL50581_4139	hypothetical protein	Y (QeX and Orb)
GL50803_4239	hypothetical protein	8.93E+05	2.69E+05	0.30155	-	Undetermined	-	11494	-	GL50581_73	hypothetical protein	Y (QeX and Orb)
GL50803_137740	VSP-3	2.57E+05	2.61E+05	1.01435	protein binding	Plasma membrane	-	46480	-	-	-	-
GL50803_14019	Cathepsin B precursor	4.40E+05	2.53E+05	0.57417	proteolysis	Cytoplasm	1	33056	0	GSB_14019	Cathepsin B-like cysteine proteinase	N
GL50803_10330	Tenascin precursor	2.04E+05	2.47E+05	1.21209	protein binding	Cytoplasm	0.98	28090	0	GL50581_4057	Tenascin precursor	Y (QeX and Orb)
GL50803_8687	Tenascin precursor	1.41E+03	2.27E+05	160.408	protein binding	Cytoplasm	0.99	53969	0	GL50581_4316	Tenascin precursor	Y (QeX and Orb)
GL50803_16468	Cathepsin B precursor	2.28E+04	2.23E+05	9.79739	proteolysis	Cytoplasm	0.86	33951	0	GL50581_438	Cathepsin B precursor	Y (QeX and Orb)
GL50803_95162	Tenascin-like	1.95E+05	1.61E+05	0.82605	protein binding	Cytoplasm	0.98	61787	0	GL50581_1982	Tenascin-like	Y (QeX and Orb)
GL50803_16160	Cathepsin B precursor	3.39E+05	1.35E+05	0.39773	proteolysis	Cytoplasm	0.3	39199	0	GL50581_2946	Cathepsin B precursor	Y (QeX and Orb)
GL50803_114815	Tenascin precursor	1.69E+04	1.17E+05	6.90495	calcium ion binding / protein binding	Cytoplasm	0.89	63018	0	-	-	-
GL50803_11354	hypothetical protein	3.28E+05	1.10E+05	0.33391	-	Undetermined	-	30099	-	GL50581_4499	Hypothetical protein	Y (QeX and Orb)
GL50803_13272	hypothetical protein	2.71E+05	9.06E+04	0.33379	-	Undetermined	-	40412	-	GL50581_3472	Inosine-uridine nucleoside N-ribohydrolase	Y (QeX and Orb)
GL50803_8227	ATP-binding cassette protein 5	2.80E+04	8.18E+04	2.92188	ATPase activity	Cytoplasm	-	118852	-	GL50581_3399	ATP-binding cassette protein 5	N
GL50803_28477	hypothetical protein	4.48E+04	6.77E+04	1.5093	-	Undetermined	-	5786	-	GL50581_2110	Hypothetical protein	Y
GL50803_480	Translation initiation inhibitor	8.24E+04	5.15E+04	0.62574	deaminase activity	Cytoplasm	-	12893	-	GL50581_4017	Translation initiation inhibitor	Y (QeX and Orb)
GL50803_17516	Cathepsin B precursor	1.32E+04	4.96E+04	3.76073	proteolysis	Cytoplasm	1	34681	0	GL50581_2318	Cathepsin B precursor	Y (QeX and Orb)
GL50803_136021	Beta tubulin	-	4.12E+04	-	GTP catabolic process / protein polymerization	Cytoskeleton	-	50050	-	GSB_154277	Tubulin beta chain	N
GL50803_112312	Elongation factor 1-alpha	-	4.11E+04	-	translational elongation	Cytoplasm	-	49105	-	-	-	N
GL50803_10217	Cathepsin B precursor	-	3.97E+04	-	proteolysis	Cytoplasm	1	33478	0	GSB_155190	Cathepsin B-like cysteine proteinase	N
GL50803_16522	hypothetical protein	3.92E+04	2.47E+04	0.62911	-	Undetermined	-	16666	-	GL50581_352	Hypothetical protein	Y (QeX and Orb)
GL50803_137613	VSP with INR	4.51E+04	2.35E+04	0.52085	-	Plasma membrane	0.99	40536	1	-	-	-
GL50803_15250	High cysteine membrane protein Group 6	5.10E+03	2.26E+04	4.44197	protein binding	Plasma membrane	1	66839	1	GSB_154625	Variant-specific surface protein	N
GL50803_5258	hypothetical protein	-	2.16E+04	-	-	Undetermined	-	11782	-	GL50581_2767	Hypothetical protein	Y (QeX and Orb)
GL50803_103454	High cysteine membrane protein Group 1	1.98E+04	1.99E+04	1.00358	-	Plasma membrane	1	79409	1	-	-	-
GL50803_16833	Tenascin-like	1.18E+04	1.92E+04	1.6281	protein binding	Cytoplasm	0.99	64667	0	GL50581_1227	Tenascin-like protein	Y (QeX and Orb)
GL50803_3910	hypothetical protein	4.44E+04	1.57E+04	0.35375	cell redox homeostasis	Cytoplasm	-	13724	-	GL50581_661	Hypothetical protein	Y (QeX and Orb)
GL50803_15564	Cathepsin B precursor	8.92E+03	1.57E+04	1.75911	proteolysis	Cytoplasm	-	36985	-	GL50581_2036	Cathepsin B precursor	Y (QeX and Orb)
GL50803_14299	hypothetical protein	2.46E+04	1.50E+04	0.61163	-	Undetermined	-	15554	-	GL50581_1093	Hypothetical protein	Y (QeX and Orb)
GL50803_102813	Protein 21.1	4.28E+04	1.33E+04	0.31151	protein binding	Cytoplasm	-	36989	-	-	-	-
GL50803_10238	hypothetical protein	3.91E+02	8.70E+03	22.2276	carbohydrate metabolic process	Cytoplasm	1	113230	1	-	-	-
GL50803_102438	Plasma membrane calcium-transporting ATPase 2	-	8.70E+03	-	calcium-transporting ATPase activity	Signaling systems	-	85318	-	GL50581_4552	Plasma membrane calcium-transporting ATPase 2	N
GL50803_2661	Cyclin-dependent Kinases regulato	-	6.93E+03	-	cell cycle	Cytoplasm	-	10358	-	GL50581_3484	Cyclin-dependent Kinases regulatory	Y (QeX and Orb)
GL50803_135886	hypothetical protein	-	6.88E+03	-	catalytic activity	Undetermined	-	24039	-	GSB_150820	Putative histidine phosphatase domain protein	N
GL50803_8742	Extracellular nuclease, putative	-	6.28E+03	-	nuclease activity	Cytoplasm	0.99	30781	0	GL50581_3607	Extracellular nuclease, putative	Y (QeX and Orb)
GL50803_17476	CXC-rich protein	5.50E+03	6.12E+03	1.11393	protein binding	Cytoplasm	1	224253	1	GL50581_4509	CXC-rich protein	Y (QeX)
GL50803_15574	Alanyl dipeptidyl peptidase	1.00E+04	5.54E+03	0.5525	proteolysis	Cytoplasm	0.92	82780	1	GL50581_1499	Alanyl dipeptidyl peptidase	Y (QeX and Orb)
GL50803_12063	Variant-specific surface protein	6.22E+03	5.42E+03	0.87146	-	Plasma membrane	1	43782	0	GL50581_2622	Variant-specific surface protein	Y (QeX and Orb)
GL50803_14573	Tenascin-X	9.85E+02	5.10E+03	5.17242	protein binding	-	1	53351	0	GL50581_1475	Tenascin-X	N
GL50803_5604	hypothetical protein	1.61E+04	4.93E+03	0.30638	-	Undetermined	-	14802	-	GL50581_3436	Hypothetical protein	Y (QeX and Orb)
GL50803_16477	Tenascin-37	4.67E+03	4.52E+03	0.9683	protein binding	Cytoplasm	0.95	61934	0	GL50581_3575	Tenascin-37	Y (QeX and Orb)
GL50803_16716	High cysteine membrane protein Group 5	8.18E+03	4.50E+03	0.55017	protein binding	Plasma membrane	1	72587	1	GSB_153259	Variant-specific surface protein	N

GL50803_7843	hypothetical protein	1.19E+04	4.29E+03	0.35876	-	Undetermined	-	21923	-	GL50581_1952	Hypothetical protein		Y (QeX and Orb)
GL50803_17277	Phospholipase B	1.17E+04	3.97E+03	0.33919	-	Cytoplasm	1	65608	0	GL50581_128	Phospholipase B		Y (QeX and Orb)
GL50803_8789	hypothetical protein	-	3.06E+03	-	protein binding	Undetermined	-	36237	-	GL50581_789	Hypothetical protein		Y (QeX and Orb)
GL50803_101010	VSP with INR	9.67E+02	2.99E+03	3.08712	-	Plasma membrane	0.98	42851	1	GSB_155559	Variant-specific surface protein		N
GL50803_9148	H-SHIPPO 1	-	2.90E+03	-	-	Undetermined	-	22258	-	GL50581_1356	H-SHIPPO 1		Y (QeX and Orb)
GL50803_15004	NADPH oxidoreductase, putative	-	2.43E+03	-	-	Cytoplasm	-	18610	-	-	-		N
GL50803_4059	5-methylthioadenosine nucleosidase, S-adenosylhomocysteine nucleosidase	2.03E+03	2.42E+03	1.19132	nucleoside metabolic process / nucleoside catabolic process /L-methionine salvage from methylthioadenosine	Cytoplasm	-	31648	-	GL50581_123	5'-methylthioadenosine nucleosidase, S-adenosylhomocysteine nucleosidase TOPep_Count=13		Y (QeX and Orb)
GL50803_9899	hypothetical protein	7.59E+03	2.30E+03	0.30301	-	Undetermined	-	59015	-	GL50581_4293	Hypothetical protein		Y (QeX and Orb)
GL50803_101832	High cysteine protein	4.57E+03	2.07E+03	0.45211	-	Plasma membrane	0.77	57049	1	GL50581_3963	High cysteine protein		Y (QeX and Orb)
GL50803_33689	Dipeptidyl-peptidase III	1.57E+03	2.02E+03	1.28621	proteolysis	Cytoplasm	-	78645	-	GSB_33689	Dipeptidyl-peptidase III		N
GL50803_14546	Phosphoglycolate phosphatase	-	1.98E+03	-	phosphoglycolate phosphatase activity	Cytoplasm	-	25277	-	GL50581_1485	Phosphoglycolate phosphatase		Y(QeX only)
GL50803_16795	Topoisomerase II	5.72E+03	1.82E+03	0.31798	nucleotide binding/ATP binding	Cytoplasm	-	57018	-	GL50581_3137	Topoisomerase II		Y (QeX and Orb)
GL50803_101074	VSP with INR	4.59E+03	1.78E+03	0.38754	-	Plasma membrane	1	76764	1	-	-		-
GL50803_16623	hypothetical protein	-	1.76E+03	-	hydrolase activity	Cytoplasm	-	19182	-	GL50581_1418	Nudix hydrolase, putative		Y (QeX and Orb)
GL50803_112135	High cysteine membrane protein VSP-like	-	1.65E+03	-	protein binding	Plasma membrane	1	120767	1	GSB_152985	Variant-specific surface protein		N
GL50803_114210	hypothetical protein	3.81E+03	1.56E+03	0.41008	protein binding	Cytoplasm	0.95	114009	1	GL50581_3969	Hypothetical protein		Y (QeX)
GL50803_137617	VSP -77	8.33E+02	1.54E+03	1.85102	protein binding	Plasma membrane	1	75503	1	-	-		-
GL50803_7715	High cysteine membrane protein Group 1	-	1.14E+03	-	-	Plasma membrane	0.79	72830	1	GL50581_727	High cysteine membrane protein Group 1		Y (QeX and Orb)
GL50803_112063	hypothetical protein	-	9.74E+02	-	transmembrane transport	Plasma membrane	-	72525	-	GSB_112063	Transporter, MFS superfamily protein		N
GL50803_114674	Hypothetical protein	-	7.63E+02	-	protein binding	Undetermined	0.99	73024	1	-	-		-
GL50803_137618	VSP	1.75E+03	7.47E+02	0.42671	protein binding	Plasma membrane	-	88660	-	-	-		N
GL50803_16936	High cysteine membrane protein EGF-like	-	7.38E+02	-	protein binding	Plasma membrane	0.96	162611	1	GL50581_4085	High cysteine membrane protein EGF-like		N
GL50803_91348	Purine nucleoside phosphorylase lateral transfer candidate	-	6.98E+02	-	nucleoside metabolic process/Cytoplasm		-	88455	-	GL50581_1357	Purine nucleoside phosphorylase lateral transfer candidate		Y (QeX and Orb)
GL50803_5180	Peptide methionine sulfoxide reductase msrB	-	5.35E+02	-	protein repair / response to oxidative stress / peptide-methionine (R or S)-S-oxide reductase activity/	Cytoplasm	-	17407	-	GL50581_3084	Peptide methionine sulfoxide reductase msrB		Y (QeX and Orb)
GL50803_115338	hypothetical protein	-	5.22E+02	-	protein binding	Undetermined	-	36491	-	GSB_152063	hypothetical protein		N
GL50803_137612	VSP-160	-	4.81E+02	-	-	Plasma membrane	0.87	89600	1	GSB_150872	Variant-specific surface protein		N
GL50803_113970	hypothetical protein	-	4.10E+02	-	-	Undetermined	-	30970	-	GL50581_90	Hypothetical protein		N
GL50803_113038	hypothetical protein	-	4.06E+02	-	calcium ion binding /protein binding	Undetermined	1	60841	0	GL50581_4180	Tenascin precursor		Y (QeX and Orb)
GL50803_11309	High cysteine membrane protein Group 1	-	3.47E+02	-	protein binding	Plasma membrane	0.72	140405	1	GL50581_3514	High cysteine membrane protein Group 1		Y (QeX and Orb)
GL50803_95593	Kinase, NEK	-	3.32E+02	-	protein phosphorylation	Cytoplasm	-	42363	-	GL50581_449	Kinase, NEK		Y (QeX and Orb)
GL50803_137708	VSP-48.2	-	2.49E+02	-	protein binding	Plasma membrane	1	60304	1	-	-		N
GL50803_91354	H-SHIPPO 1	-	2.26E+02	-	-	Undetermined	-	63489	-	GL50581_1356	H-SHIPPO 1		Y (QeX and Orb)
GL50803_112801	VSP-169	-	1.55E+02	-	protein binding	Plasma membrane	0.99	76443	1	GSB_154903	Variant-specific surface protein		N
GL50803_116477	VSP-149	-	1.36E+02	-	protein binding	Plasma membrane	1	76114	1	GSB_153909	Variant-specific surface protein		N
GL50803_24947	hypothetical protein	-	1.15E+02	-	protein transport	Signaling systems	-	20953	-	GL50581_349	Hypothetical protein		Y(QeX only)
GL50803_14586	VSP with INR	-	8.03E+01	-	protein binding	Plasma membrane	1	75531	1	-	-		N
GL50803_9239	NI13-related protein	-	6.89E+01	-	-	Undetermined	-	29852	-	GL50581_4241	NIf3-related protein		Y(QeX only)
GL50803_115797	VSP-54 with INR	-	2.16E+01	-	protein binding	Plasma membrane	0.98	70227	1	-	-		N
GL50803_113319	High cysteine membrane protein TMK-like	-	1.34E+01	-	cell adhesion	Plasma membrane	0.99	265709	1	-	-		N

[a] Signal peptide
[b] Molecular Weight
[c] Transmembrane Domain
[d] QeX: QeXactive dataset, Orb: Orbitrap dataset

D.

	Assemblage A									Assemblage B syntenic orthologs		
Protein Accession Number	Protein Description	QeXactive P Expressio n:iBAQ	SP Expressio n:iBAQ	SP/P Ratio	GO function	GO location	SignalP[a]	MW[b] (Da)	TM[c]	Accession	Description	Ortholog identified in GS isolate dataset (Y/N) [d]
GL50803_33279	VSP	3.68E+08	2.98E+08	0.80993	protein binding	Plasma membrane	0.98	75981	1	-	-	-
GL50803_5810	hypothetical protein	8.79E+07	3.76E+07	0.42761	oxidation-reduction process	Cytoplasm	-	14500	-	GL50581_4133	hypothetical protein	Y (QeX and Orb)
GL50803_17163	Peptidyl-prolyl cis-trans isomerase B precursor	6.89E+07	3.03E+07	0.43974	protein folding	Cytoplasm	-	18044	-	GL50581_1019	Peptidyl-prolyl cis-trans isomerase B precursor	Y (QeX and Orb)
GL50803_14670	Protein disulfide isomerase PDI3	6.11E+07	2.37E+07	0.38784	glycerol ether metabolic process / cell redox homeostasis	Cytoplasm	1	12669	0	GL50581_3386	Protein disulfide isomerase PDI5	Y (QeX and Orb, top 50
GL50803_10330	Tenascin precursor	7.37E+06	1.70E+07	2.30439	protein binding	Cytoplasm	0.98	28090	0	GL50581_4057	Tenascin precursor	Y (QeX and Orb)
GL50803_16779	Cathepsin B precursor	6.73E+06	1.56E+07	2.31252	proteolysis	Cytoplasm	1	32761	0	GL50581_78	Cathepsin B precursor	Y (QeX and Orb)
GL50803_16322	Neurogenic locus Notch protein precursor	2.12E+07	1.46E+07	0.69085	calcium ion binding / protein binding	Cytoplasm	0.99	63070	0	GL50581_31	Neurogenic locus Notch protein precursor	Y (QeX and Orb)
GL50803_95162	Tenascin-like	1.62E+07	1.28E+07	0.79473	protein binding	Cytoplasm	0.98	61787	0	GL50581_1982	Tenascin-like	Y (QeX and Orb)
GL50803_137740	VSP-3	1.47E+07	1.18E+07	0.80039	protein binding	Plasma membrane	-	46480	-	-	-	-
GL50803_16468	Cathepsin B precursor	1.34E+06	1.11E+07	8.32112	proteolysis	Cytoplasm	0.86	33951	0	GL50581_438	Cathepsin B precursor	Y (QeX and Orb)
GL50803_13272	hypothetical protein	1.83E+07	9.33E+06	0.50938	-	Undetermined	-	40412	-	GL50581_3472	Inosine-uridine nucleoside h	Y (QeX and Orb)
GL50803_103676	Alpha-tubulin	-	8.46E+06	-	GTP catabolic process / protein polymerization	Cytoskeleton	-	50553	-	GL50581_1380	Alpha-tubulin	Y (QeX and Orb)
GL50803_8687	Tenascin precursor	3.77E+05	7.17E+06	19.0496	protein binding	Cytoplasm	0.99	53969	0	GL50581_4316	Tenascin precursor	Y (QeX and Orb, top 50
GL50803_14019	Cathepsin B precursor	1.55E+07	6.56E+06	0.42377	proteolysis	Cytoplasm	1	33056	0	GSB_14019	Cathepsin B-like cysteine pr	N
GL50803_6289	FixW protein, putative	1.65E+07	5.74E+06	0.34674	-	Cytoplasm	-	14958	-	GL50581_4194	FixW protein, putative	Y (QeX and Orb)
GL50803_480	Translation initiation inhibitor	9.47E+06	4.08E+06	0.43038	deaminase activity	Cytoplasm	-	12893	-	GL50581_4017	Translation initiation inhibit	Y (QeX and Orb)
GL50803_113038	hypothetical protein	-	3.56E+06	-	calcium ion binding / protein binding	Undetermined	1	60841	0	GL50581_4180	Tenascin precursor	Y (QeX and Orb)
GL50803_8227	ATP-binding cassette protein 5	2.93E+06	2.89E+06	0.98546	ATPase activity	Cytoplasm	-	118852	-	GL50581_3399	ATP-binding cassette protein 5	N
GL50803_12091	Macrophage migration inhibitory factor	6.39E+06	2.62E+06	0.40966	-	Undetermined	-	12368	-	GL50581_1500	Macrophage migration inhibitory factor	Y (QeX and Orb)
GL50803_16522	hypothetical protein	4.85E+06	2.53E+06	0.52138	-	Undetermined	-	16666	-	GL50581_352	Hypothetical protein	Y (QeX and Orb)
GL50803_112312	Elongation factor 1-alpha	-	2.22E+06	-	translational elongation	Cytoplasm	-	49105	-	-	-	N
GL50803_17516	Cathepsin B precursor	3.36E+05	2.04E+06	6.06677	proteolysis	Cytoplasm	1	34681	0	GL50581_2318	Cathepsin B precursor	Y (QeX and Orb)
GL50803_5258	hypothetical protein	3.42E+06	1.99E+06	0.58272	-	Undetermined	-	11782	-	GL50581_2767	Hypothetical protein	Y (QeX and Orb)
GL50803_114815	Tenascin precursor	1.69E+06	1.76E+06	1.04278	calcium ion binding / protein binding	Cytoplasm	0.89	63018	0	-	-	-
GL50803_113892	Long chain fatty acid CoA ligase, putative	1.26E+06	1.68E+06	1.33546	metabolic process	Cytoplasm	-	84549	-	GL50581_2104	Long chain fatty acid CoA lig	Y (QeX and Orb)
GL50803_9353	Hypothetical protein	-	1.49E+06	-	phosphatidylinositol phosphorylation	Cytoplasm	-	49541	-	GL50581_1611	hypothetical protein	N
GL50803_137613	VSP with INR	2.41E+06	1.36E+06	0.56346	-	Plasma membrane	0.99	40536	1	-	-	-
GL50803_103454	High cysteine membrane protein Group 1	2.11E+06	1.33E+06	0.62925	-	Plasma membrane	1	79409	1	-	-	-
GL50803_16833	Tenascin-like	1.16E+06	1.08E+06	0.93724	protein binding	Cytoplasm	0.99	64667	0	GL50581_1227	Tenascin-like protein	Y (QeX and Orb)
GL50803_15564	Cathepsin B precursor	3.58E+05	9.63E+05	2.69212	proteolysis	Cytoplasm	-	36985	-	GL50581_2036	Cathepsin B precursor	Y (QeX and Orb)
GL50803_41942	High cysteine membrane protein Group 6	5.29E+05	9.38E+05	1.77272	protein binding	Plasma membrane	-	61723	-	GL50581_250	High cysteine membrane protein Group 6	Y (QeX and Orb)
GL50803_17000	Peptidyl-prolyl cis-trans isomerase B precursor	2.14E+06	7.71E+05	0.36034	protein folding	Cytoplasm	1	21210	0	GL50581_1186	Peptidyl-prolyl cis-trans isomerase B precursor	Y (QeX and Orb)
GL50803_10217	Cathepsin B precursor	-	7.56E+05	-	proteolysis	Cytoplasm	1	33478	0	GSB_155190	Cathepsin B-like cysteine proteinase	N
GL50803_114462	Axonemal p66.0	5.21E+05	6.28E+05	1.20656	-	Cytoskeleton	-	62607	-	GL50581_769	Axonemal p66.0	Y (QeX and Orb)
GL50803_102841	hypothetical protein	1.08E+06	5.66E+05	0.52276	-	Undetermined	-	17549	-	-	-	-
GL50803_15250	High cysteine membrane protein Group 6	1.69E+05	5.09E+05	3.0102	protein binding	Plasma membrane	1	66839	1	GSB_154625	Variant-specific surface protein	N
GL50803_12063	Variant-specific surface protein	3.45E+05	4.46E+05	1.29337	-	Plasma membrane	1	43782	0	GL50581_2622	Variant-specific surface protein	Y (QeX and Orb)
GL50803_2661	Cyclin-dependent kinases regulatory subunit	1.17E+06	4.18E+05	0.35838	cell cycle	Cytoplasm	-	10358	-	GL50581_3484	Cyclin-dependent Kinases regulatory subunit	Y (QeX and Orb)
GL50803_8742	Extracellular nuclease, putative	-	4.06E+05	-	nuclease activity	Cytoplasm	0.99	30781	0	GL50581_3607	Extracellular nuclease, putative	Y (QeX and Orb)

GL50803_9276	Variant-specific surface protein	-	3.56E+05	-	protein binding	Plasma membrane	0.98	50500	0	GL50581_235	Variant-specific surface protein	-	Y (QeX and Orb)
GL50803_114671	Protein 21.1	2.17E+05	3.15E+05	1.448	protein binding	Cytoplasm	-	69474	-	-	-	-	-
GL50803_3760	Protein 21.1	5.48E+05	3.15E+05	0.5749	protein binding	Cytoplasm	-	50095	-	GL50581_1174	Protein 21.1	Y	Y (QeX and Orb)
GL50803_13825	Kinesin-1	1.73E+05	3.15E+05	1.81925	microtubule-based movement	Cytoskeleton	-	107938	-	GL50581_3851	Kinesin-1	Y	Y (QeX and Orb)
GL50803_17478	Dynein heavy chain	1.48E+04	3.15E+05	21.3456	nucleotide binding / nucleoside-triphosphatase activity	Cytoplasm	0.55	601122	0	GL50581_4498	Dynein heavy chain	N	
GL50803_113616	Hypothetical protein	-	3.15E+05	-	-	Cytoplasm	-	49083	-	GSB_153361	Hypothetical protein	N	
GL50803_16477	Tenascin-37	2.46E+05	2.96E+05	1.20191	protein binding	Cytoplasm	0.95	61934	0	GL50581_3575	Tenascin-37	Y	Y (QeX and Orb)
GL50803_17476	CXC-rich protein	4.80E+05	2.83E+05	0.58945	protein binding	Cytoplasm	1	224253	1	GL50581_4509	CXC-rich protein	Y	Y (QeX)
GL50803_14573	Tenascin-X	3.05E+05	2.79E+05	0.91504	protein binding		1	53351	0	GL50581_1475	Tenascin-X	N	
GL50803_9116	Kinase, CMGC GSK	-	2.51E+05	-	protein phosphorylation	Cytoplasm	-	42242	-	GL50581_1445	Kinase, CMGC GSK	Y	Y(QeX only)
GL50803_14546	Phosphoglycolate phosphatase	-	2.45E+05	-	phosphoglycolate phosphatase activity	Cytoplasm	-	25277	-	GL50581_1485	Phosphoglycolate phosphatase	Y	Y(QeX only)
GL50803_7556	Lysosomal acid phosphatase precursor	1.96E+05	1.53E+05	0.78049	acid phosphatase activity	Lysosome	0.7	45771	0	GL50581_680	Lysosomal acid phosphatase precursor	Y	Y (QeX and Orb)
GL50803_115066	High cysteine membrane protein VSP-like	4.18E+05	1.49E+05	0.35525	protein binding	Plasma membrane	0.98	122459	1	GSB_155460 / GSB_155461 / GSB_155462	Variant-specific surface protein	N	
GL50803_112135	High cysteine membrane protein VSP-like	-	1.40E+05	-	protein binding	Plasma membrane	1	120767	1	GSB_152985	Variant-specific surface protein	N	
GL50803_101074	VSP with INR	4.04E+05	1.40E+05	0.34619	-	Cytoplasm	1	76764	1	-	-	-	-
GL50803_112473	hypothetical protein	-	1.35E+05	-	-	Undetermined	1	42511	0	-	-	N	
GL50803_8367	VATPase	3.69E+05	1.30E+05	0.3521	ion transmembrane transport	Signaling systems	-	15435	-	GL50581_375	VATPase	N	
GL50803_10238	hypothetical protein	-	1.21E+05	-	carbohydrate metabolic process	Extracellular region	1	113230	1	-	-	N	
GL50803_16267	hypothetical protein	2.81E+05	1.05E+05	0.37293	-	Undetermined	-	63116	-	GL50581_3075	Coiled-coil protein	Y (QeX and Orb)	
GL50803_5180	Peptide methionine sulfoxide reductase msrB	1.24E+05	8.80E+04	0.70952	protein repair / response to oxidative stress / peptide-methionine (R or S)-S-oxide reductase activity	Cytoplasm	-	17407	-	GL50581_3084	Peptide methionine sulfoxide reductase msrB	Y (QeX and Orb)	
GL50803_10016	hypothetical protein	-	7.24E+04	-	-	Undetermined	0.77	35331	0	GL50581_3566	Hypothetical protein	Y (QeX and Orb)	
GL50803_7715	High cysteine membrane protein	-	6.61E+04	-	-	Plasma membrane	0.79	72830	1	GL50581_727	High cysteine membrane protein Group 1	Y (QeX and Orb)	
GL50803_16936	High cysteine membrane protein	-	4.82E+04	-	protein binding	Plasma membrane	0.96	162611	1	GL50581_4085	High cysteine membrane protein EGF-like	N	
GL50803_113450	VSP with INR	3.79E+04	2.09E+04	0.55201	protein binding	Plasma membrane	0.99	74848	1	GSB_152640	Variant-specific surface protein	N	
GL50803_41472	VSP-49	-	2.08E+04	-	protein binding	Plasma membrane	1	59856	1	-	-	N	
GL50803_111936	VSP-162.1	-	1.86E+04	-	protein binding	Plasma membrane	1	75411	1	-	-	N	
GL50803_114674	Hypothetical protein	2.98E+04	1.40E+04	0.46856	protein binding	Undetermined	0.99	73024	1	-	-	-	
GL50803_137617	VSP -77	2.07E+04	1.09E+04	0.52804	protein binding	Plasma membrane	1	75503	1	-	-	-	
GL50803_116477	VSP-149	-	7.98E+03	-	protein binding	Plasma membrane	1	76114	1	GSB_153909	Variant-specific surface protein	N	
GL50803_14586	VSP with INR	1.95E+04	7.81E+03	0.40055	protein binding	Plasma membrane	1	75531	1	-	-	N	
GL50803_112633	High cysteine membrane protein EGF-like	-	5.60E+03	-	protein binding	Plasma membrane	1	179884	1	-	-	N	
GL50803_112208	VSP-98.1 with INR	-	5.41E+03	-	protein binding	Plasma membrane	0.98	60519	1	-	-	N	
GL50803_11521	VSP-126.1	-	5.04E+03	-	protein binding	Plasma membrane	1	64144	1	GSB_150196	Variant-specific surface protein	N	
GL50803_113319	High cysteine membrane protein TMK-like	-	3.95E+03	-	cell adhesion	Plasma membrane	0.99	265709	1	-	-	N	
GL50803_11309	High cysteine membrane protein Group 1	-	3.14E+03	-	protein binding	Plasma membrane	0.72	140405	1	GL50581_3514	High cysteine membrane protein Group 1	Y (QeX and Orb)	

[a] Signal peptide
[b] Molecular Weight
[c] Transmembrane Domain
[d] QeX: QeXactive dataset, Orb: Orbitrap dataset

Appendix X:

ClustalW sequence alignment of the Cathepsin B precursor genes coding for the Cathepsin B proteins identified in *Giardia* proteome. All the GL50581_ accession numbers represent GS isolate proteins and all the GL50803_ represent WB isolate proteins.

```
CLUSTAL 2.1 multiple sequence alignment


GL50581_2318        ------------------------------------------------------------
GL50803_17516       ------------------------------------------------------------
GL50581_159         ------------------------------------------------------------
GL50581_3619        ------------------------------------------------------------
GL50803_10217       ------------------------------------------------------------
GL50581_2036        ------------------------------------------------------------
GL50803_15564       ------------------------------------------------------------
GL50581_78          ------------------------------------------------------------
GL50803_16779       ------------------------------------------------------------
GL50803_14019       ------------------------------------------------------------
GL50581_2946        ------------------------------------------------------------
GL50803_16160       ATGGATGCATTGATTGAGCCAAACATGCTTTTAGCTAAAGTCTTTATCACTATCTCTTGG 60
GL50581_438         ------------------------------------------------------------
GL50803_16468       ------------------------------------------------------------


GL50581_2318        ------------------------------------------------------------
GL50803_17516       ------------------------------------------------------------
GL50581_159         ------------------------------------------------------------
GL50581_3619        ------------------------------------------------------------
GL50803_10217       ------------------------------------------------------------
GL50581_2036        ------------------------------------------------------------
GL50803_15564       ---------------------------------------------------------ATG 3
GL50581_78          ------------------------------------------------------------
GL50803_16779       ------------------------------------------------------------
GL50803_14019       ------------------------------------------------------------
GL50581_2946        ------------------------------------------------------------
GL50803_16160       CTCAGAACAGCTTGCAGAGCTAGATCTTTGTCCGCGCTACCTCTCAGTGCATCAAGCAGA 120
GL50581_438         ------------------------------------------------------------
GL50803_16468       ------------------------------------------------------------


GL50581_2318        ------------------------------------------------------------
GL50803_17516       ------------------------------------------------------------
GL50581_159         ------------------------------------------------------------
GL50581_3619        ------------------------------------------------------------
GL50803_10217       ------------------------------------------------------------
GL50581_2036        ------------------------------------------------------------
GL50803_15564       CGCGCGGGAATCCTCTGGTGGCGCTGTCGGTCTAGAGATAGACGTATGTACAGCCCCGAA 63
GL50581_78          ------------------------------------------------------------
GL50803_16779       ------------------------------------------------------------
GL50803_14019       ------------------------------------------------------------
GL50581_2946        ------------------------------------------------------------
GL50803_16160       CAACGATTTGTAGAAAAGTCACGTACCTGTCATACTTGCCTGTACACACAAACGGGCAAA 180
GL50581_438         ------------------------------------------------------------
GL50803_16468       ------------------------------------------------------------

GL50803_17516       ---------------ATGAGGCTTCT---------GATCCTTCTGATT-------GCATC 29
GL50581_159         ---------------ATGATTCTAGC---------TTTACTTTTAGCT-------GTGGT 29
GL50581_3619        ---------------ATGATTCTAGC---------TTTACTTTTAGCT-------GTRGT 29
GL50803_10217       ---------------ATGGCTTTATC---------TTTGCTCTTGGCT-------GTAGT 29
GL50581_2036        ---------------ATGCTACCTGTGGCCACCTGCTTGCTA-TTGGTGTTGGGCACCGG 44
GL50803_15564       CGCGTATCGTTTAAAATGTTGTCTGTGATCACGTATCTCTTAGCTGGGCTTGGG-GTGGC 122
GL50581_78          ---------------ATGAAGCTCTT---------CCTCCTTGCTGCG-------GCAGC 29
GL50803_16779       ---------------ATGAAGCTCTT---------TCTCCTCGCCGCT-------GCAGC 29
GL50803_14019       ---------------ATGAAGCTCTT---------TCTCCTCGCCGCT-------GCAGC 29
GL50581_2946        ------------ATGATAGGGCTGTC---------TCTTTTATTGGGC-------GCACT 32
GL50803_16160       ATCGAAATTAAAATGATAGGGGCTTC---------TCTTTTGCTAGGT-------GCAGT 224
GL50581_438         ---------------ATGCTACTTGC---------TGCCTTTGCTGTC-------GCAGT 29
GL50803_16468       ---------------ATGCTATTTGC---------TGTCCTCGTTGTT-------GCAGT 29
```

```
GL50581_2318        CTTAGCCGAGCC-CCTTTTAACTCAAGCGAAGCTAAGGCAGATACAAGCACTTGGT---C 85
GL50803_17516       CCTAGCCGAGCC-CCTTTTGACCCAGGTGGAGTTGAGGCAAATTCAAGCACTTGCT---C 85
GL50581_159         CTGCGCAAAGCC-CCTGGTAAGTCGTGCCGAGCTGAGACGCATTCAGGCCCTCAAC---C 85
GL50581_3619        CTGCGCAAAGCC-CCTGGTAAGTCGTGCCGAGCTGAGACGCATTCAGGCCCTCAAC---C 85
GL50803_10217       TTGTGCAAAGCC-CCTGGTGAGCCGCGCCGAGCTAAGGCGGATTCAGGCTCTCAAT---C 85
GL50581_2036        CCTTTCCCGGCC-TCTCCTTTCCGCAAGAGAGCTGAGGGCCATTCTCTCTCTCCAT---C 100
GL50803_15564       CCTCTCCAAGCC-CCTTCTCTCCAGAAGGGAGTTACAGGAGATCCGCGCCCTGCAA---C 178
GL50581_78          CTTCTCCGCGCC-CGCCCTGACGGTTTCCGAGCTCAACCACATCAAGTCCCTGAAC---C 85
GL50803_16779       TTTCTCTGCGCC-CGCCCTGACGGTCTCCGAGCTTAACCACATCAAGTCCCTGAAT---C 85
GL50803_14019       TTTCTCTGCGCC-CGCCCTGACGGTCTCCGAGCTTAACCACATCAAGTCCCTGAAT---C 85
GL50581_2946        TCTTGCTGCTCC-CGCCGTCTCCCGCATAGATCTCCGTACGATCCAGGCGCTTGATGGTC 91
GL50803_16160       TCTTGCCGCTCC-GGCTGTCTCCCACGCCGATCTCCACACAATTAAGGCGCTCGATGGTC 283
GL50581_438         CTTGTCAACTCCGTTCTATAGCCCGCATC---TCCTGAAATATTTGGAGAAGAAGGAGAA 86
GL50803_16468       TCTATCAACTCCGTTCTATAGCCCCCATC---TCCTGAAATATTTGAACAAGAAGGAAAA 86
                          *    **      *                    **

GL50581_2318        CGA--TATGGAAGGCTGGCATCCCGGAGCGACTTAAGAATCTCACGGAAACTGACTTCAA 143
GL50803_17516       CTG--CATGGAAAGCTGGCATCCCGGAGCGACTGAAGAACCTGACGAAGAACGACTTCAA 143
GL50581_159         CCT--CATGGGTGGCTGCGATGCCCAAGAGGTTTGAAAACGTTACAGAGGATGAGTTCCG 143
GL50581_3619        CCC--CATGGGTGGCTGCGATGCCCAAGAGGTTTGAAAACGTTACAGAGGATGAGTTCCG 143
GL50803_10217       CTC--CATGGAAAGCAGGGATGCCAAAAAGGTTTGAAAACGTCACAGAAGATGAATTTCG 143
GL50581_2036        CAG--CATGGGACGCTGGGATGCCCGACCGGTTTGCGGGTCTCTCCGAGGACGACCTGCA 158
GL50803_15564       CAC--CCTGGACCGCTGGGATCTCCGACCGGCTTGTGGGCCTCTCCGAGGACGACCTGCG 236
GL50581_78          CGA--GGTGGAAGGCCGGGATCCCGAAGCGCTTCGAGGGCCTCACGAAGGACGAG----A 139
GL50803_16779       CCA--GGTGGAAGGCGGGCATCCCGAAGCGCTTCGAGGGGCTCACGAAGGATGAG----A 139
GL50803_14019       CCA--GGTGGAAGGCGGGCATCCCGAAGCGCTTCGAGGGGCTCACGAAGGACGAG----A 139
GL50581_2946        TCA--CTTGGGTCCCGGAGCTTTCGAAGCGTTTTGAGGGCAAGTCCCTGGATGAAGTCAA 149
GL50803_16160       TCA--CCTGGGTCCCGGAGCTTCCGAAGCGTTTTATGGGCAAGTCCCTGGACGAGGTCAA 341
GL50581_438         TAAGCTTTGGGAGGCGGGAATACCGATAAAGTTCGCAGATCGTACACGTGATGAGATCAC 146
GL50803_16468       CAAGCTCTGGGAGGCGGGAATACCGGCTAAATTCGCTAATCGAACGCATGATGAAGTTAC 146
                          ***    * *   *  *         *           *       **

GL50581_2318        AAGACTGG------T-CTCGG--CGAAGGATCCACGCGGCCAGATA--CC-CA---CATT 188
GL50803_17516       GAAGATGC------T-ATCAG--CAGGCAGTCCACGTACGCAATCA--TC-CA---TTGT 188
GL50581_159         TGGTATGC------TAATTAACCCCGACCGTTT--GAAAGCAAG-A--TC-CGGATCTAT 191
GL50581_3619        TGGTATGC------TRATTAAYCCCGACCGTTT--GAAAGCAAG-A--TC-CGGATCTAT 191
GL50803_10217       TAGCATGC------TGATCCGTCCCGACCGTCT--TAGAGCAAG-A--TC-TGGCTCTCT 191
GL50581_2036        CGCCATGT-----TTCCTCGCCACGACCTGGCCGCTAGCGTACCAG--C---AGAG-TGC 207
GL50803_15564       CGCCATGT-----TCCCTCGCCATGGCCAGCCCACCCGTCCGTCCG--C---AGAG-TGT 285
GL50581_78          TTTCGAGC-----CTCCTGAT----GCCCATCTCCTTCCTCAACCG--CGACAGAGCTGC 188
GL50803_16779       TCTCGAGC-----CTCCTGAT----GCCTGTTTCCTTCCTCAAGCG--CGATAGGGCTGC 188
GL50803_14019       TCTCGAGT-----CTCCTTAT----GCCTGTTTCCTTCCTTAAGAA--CGCAAAGGGCGC 188
GL50581_2946        GGCCATGCTCGGCCCCCTCATCA--ACACATCTAGACCGGCGATCA--CCAGAAGG---- 201
GL50803_16160       GGCTATGTTCGGTCCCCTTGTTG--ACACGTCCAGACCTGCAATTA--CCATGAGG---- 393
GL50581_438         AAAGATGT------TTTCTC--TCACACGTTTCTCAAGCCAACCATTCCTCGCTACTAT 198
GL50803_16468       AAAGATGT------TCTTCCC--CCATGCGTTCCTTAGACCAAATATTCCCCGCTATTAT 198
                          *         *

GL50581_2318        GCATCTTATTCACACGTACGAAAGCGAG---------GATCCTATACCAGATCACTTTGA 239
GL50803_17516       TCGACCTGTTCGTGTCCCTGAAAACGAA---------GATCCTGTACCAGATCACTTTGA 239
GL50581_159         GCCATCAGCTCCTCTGAAAGAGATTAATGATCCTACCGATCCTCTTCCGGCTCAGTTTGA 251
GL50581_3619        GCCATCAGCTCCTCTGAAAGAGATTAATGATCCTACCGATCCTCTTCCGGCTCAGTTTGA 251
GL50803_10217       ACCACCAATTTCGATAACAGAAGTCCAGGAACTCGTAGATCCTATTCCGCCTCAGTTCGA 251
GL50581_2036        ----CCGCGCGGCG----AGCCGTCGGG---------GAGCAT--CCCCGCCTCATTCGA 248
GL50803_15564       ----CCGCGCGCTG----AACCATCAGG---------ACCCAT--TCCCGACGCGTTTGA 326
GL50581_78          TGTCCCCCGTGGCACTATAGCCGACACA---------AAGGTC---CCCGATTCCTTTGA 236
GL50803_16779       GGTTCCCCGTGGGACCGTGTCCGCCACG---------CAGGCC---CCCGACTCCTTTGA 236
GL50803_14019       CGCACCCCGTGGCACTTTTACCGATAAG---------GACGACGTGCCTGAGTCTTTTGA 239
GL50581_2946        -------CATAGCACCA-AGCCTCCAGT---------CGGGGC--TCCAGAGAGCTACGA 242
GL50803_16160       -------CGCAGCACCA-CGCCTCCTGT---------TGGGGC--CCCGGAGAGTTATGA 434
GL50581_438         GCTGC-AGATCTCACTGAGGATGACCTCTATCCTCCGGCTGACTCCCCTGATCGCCTTGA 257
GL50803_16468       GGCGT-GAATATTACTGAGGATGATCTCTATCCTCCGGCCGGCTCCCCTGACCGTCTTGA 257
                                                              **          **
```

.

```
GL50581_2318       CTTCAGGGAGGAGTATCCACAGTGCATAACAGAAGTTATTGACATGGGTACTTGCTCATC 299
GL50803_17516      CTTCCGGGAAGAGTATCCACAGTGCATAACGGAGGTTATTGACATAGGCCTTTGCTCGTC 299
GL50581_159        CTTTCGCGACGAGTATCCTCACTGTGTCTCACCAGTTTTTGATCAGGGTAGTTGTGGAGG 311
GL50581_3619       CTTTCGCGACGAGTATCCTCACTGTGTCTCACCAGTTTTTGATCAGGGTAGTTGTGGAGG 311
GL50803_10217      CTTCCGTGATGAGTACCCTCAGTGCGTGAAACCAGCGCTGGACCAAGGAAGTTGCGGGGG 311
GL50581_2036       TTTCCGCGAGGAGTACCCGCAGTGCATCACTCCGGTCTACGACCAGGGCCACTGCGGCTC 308
GL50803_15564      CCTCCGCGAAGAGTACCCGCAGTGCATCACCCCGGTCTATGACCAAGGCTACTGCGGCGC 386
GL50581_78         TTTCCGCGAGGAGTACCCCCACTGCATCCCCGAGGTCGTCGACCAGGGCAGCTGCGGCTC 296
GL50803_16779      CTTCCGCGAGGAGTATCCCCACTGCATCCCCGAGGTTGTCGACCAGGGCGGATGCGGCTC 296
GL50803_14019      CTTCCGCGAGGAGTATCCTCACTGCATCCCCGAGGTTGTCGACCAGGGCGGGTGCGGCTC 299
GL50581_2946       CTTCCGCGATGAGTACCCCCACTGCATCACAGAGGTCGTCGACCAGGGCAGCTGCGGCTC 302
GL50803_16160      CTTCCGCGACGAGTATCCTCACTGCATCACAGAGGTCGTTGATCAGGGCAACTGCGGATC 494
GL50581_438        CTATCGTCAAACTCATCCTGAGTGCTTCTTCGAACCGGAGGATCAGAGCGACTGCTCATG 317
GL50803_16468      CTATCGCCAGACCCATCCGGAGTGCTTTTTTGAACCGGAAGACCAGAAAGAATGCTCCTG 317
                         *  *     * **  * **  *             **          **

GL50581_2318       TAGCTGGGCACATTCTCCTGTGGAGGCCTTCGGCCACAGACGATGTAT-GAATGGTGTTG 358
GL50803_17516      CAGCTGGGCGTACTCTGCTGTTGATGCCTTTAGCCACAGACGCTGTCT-AACCGGCCTGG 358
GL50581_159        ATGCTGGGCATTTTCGGCTATCGGAATGTTCGGGAGCAGACGCTGTGC-AGTAGGAATCG 370
GL50581_3619       ATGCTGGGCATTTTCGGCTATCGGAATGTTCGGGAGCAGACGCTGTGC-AGTAGGAATCG 370
GL50803_10217      ATGTTGGGCATTCTCAGCTATAGGAGTGTTCGGGGACAGGCGCTGTGC-GATGGGCATCG 370
GL50581_2036       TTGCTGGGCGTTCTCCGCCACTTCAGCCTTTGGAGACCGAAGGTGCATGCAG-GGCTTGG 367
GL50803_15564      CTGTTGGGCGTTCTCGGCGACTGGGGCGTTCGGTGACCGACGATGCATGCAGTGGCTTG- 445
GL50581_78         CTGCTGGGCCTTCTCCTCCGTTGCCTCTCTTGGTGACCGCCGCTGCTT-TGCCGGCCTCG 355
GL50803_16779      CTGCTGGGCCTTCTCCTCCGTAGCTTCCGTCGGCGACCGCCGATGCTT-TGCTGGCCTCG 355
GL50803_14019      CTGCTGGGCCTTCTCCTCTGTTGCCACCTTTGGTGATCGCCGCTGCGT-CGCTGGGCTCG 358
GL50581_2946       CTGCTGGGCCTTCTCCTCTATCCAGACCTTCGCGGACCACCGCTGCCG-AAGTGGGTTGG 361
GL50803_16160      CTGCTGGGCCTTCTCCTCCGTCCAGACCTTCGCAGATCACCGCTGCAG-GAGCGGGTTGG 553
GL50581_438        CTGCTATGCTTTTGCAACGCTTGGCGCACTATCTACACGCCGGTGCAT-TGCAAAGCTAG 376
GL50803_16468      TTGCTATGCATTTGCAACCCTGGGCGCACTATCTACACGTCGCTGCAT-CGCAAAGCTTG 376
                     * *  **     *  *           *           * **           *

GL50581_2318       ATCAGGAAGCGACGCGCTACTCTGCGCAATA-CATCTTATCCTGCGCTACAACAAACGGT 417
GL50803_17516      ATCAGGAAGCCACACGTTACTCCGCCCAGTA-CATCCTGTCTTGTTCTTCAACGAATGGT 417
GL50581_159        ATAAGGCGGCCGTTCTTTATTC---GCAACAGCATCTCATCTCATGCT-CAACGGAGAAT 426
GL50581_3619       ATAAGGCGGCCGTTCTTTATTC---GCAACAGCATCTCATCTCATGCT-CAACGGAGAAT 426
GL50803_10217      ATAAGGAGGCCGTCTCCTACTC---GCAGCAACACCTCATATCCTGCT-CACTAGAGAAC 426
GL50581_2036       ACTCAGCTGGGGTGCCCTACTC---GCAGCAGTACACCATATCGTGCGACTACCTCGATC 424
GL50803_15564      ACCCTGTTGGCGTGCCCTACTC---ACAGCAGTACACGGTCTCGTGTGATGACTTGGATC 502
GL50581_78         ACAAGAAGGCTGTCACCTACTC---CCCACAGTATGTCGTCTCGTGCGACCACGGCGATA 412
GL50803_16779      ACAAGAAAGCCGTCAAGTACTC---CCCGCAGTATGTGGTCTCCTGCGACCGTGGAGACA 412
GL50803_14019      ACAAGAAGCCTGTCAAGTACTC---CCCACAGTACGTGGTCTCCTGTGATCATGGCGACA 415
GL50581_2946       ACGCCACAGGCGTAAGCTACTC---TGTCCAATACGTGCTCGATTGTGACCGCAAGGACC 418
GL50803_16160      ATGCCACGGGCGTTAGCTACTC---GGTTCAGTACGTGCTTGACTGCGACCGCAAGGACC 610
GL50581_438        ACGCAAGTGTAGTTCCCCTTTC---TGCCCAGCACATGGTTTCATGCGATCACGGCGAGG 433
GL50803_16468      ATCCCCAAGCTGTCTCTCTTTC---CGTTCAGCACATGGTCTCGTGTGACAGCGGCGAGG 433
                   *                    **        *  *     *

GL50581_2318       TGTCTTGCGTTTCCAGGCCAGGG-GGTTGT-----AT--CCTGGGACTTTATCGCGACAA 469
GL50803_17516      TGCTTTGGATTTTCAACCCGAGA-GAGTAT-----AG--CGTGGGACTTTATTGCGACGA 469
GL50581_159        T---TTGG--CTGCAGCGGAGGC-GACTTCTTTCCAA--CCTGGAGCTTTCTGACTCAGA 478
GL50581_3619       T---TTGG--CTGCAGCGGAGGC-GACTTCTTTCCAA--CCTGGAGCTTTCTGACTCAGA 478
GL50803_10217      T---TTGG--CTGTGATGGAGGC-GATTTTCAACCGA--CATGGAGCTTTCTAACTTTCA 478
GL50581_2036       T----AGG--GTGTGCCGGCG---GGCTTTCGTTCAGCGTCTGGACCTTTCTCACAGAGC 475
GL50803_15564      T----CGG--GTGCGCTGGAG---GGACTTCGTTCAATGTTTGGACCTTCCTTACGGAGC 553
GL50581_78         T----GGC--CTGTGATGGCGGATGGCTTCAG---AGTGTCTGGAGGTTCCTCACCAAGA 463
GL50803_16779      T----GGC--TTGCGACGGCGGCTGGCTCCCG---AGCGTCTGGAGATTCCTCACCAAGA 463
GL50803_14019      T----GGC--TTGCAATGGGGGCTGGCTCCCG---AATGTCTGGAAGTTCCTTACCAAGA 466
GL50581_2946       A----CGG--CTGCAACGGCGGCGAGCCCACT---AAGGCCTTCGACTTCCTTCACAGCA 469
GL50803_16160      A----CGG--CTGCAACGGCGGCGAGCCAGTG---AACGCCTTCAATTTCCTCCACAACA 661
GL50581_438        C----AGG--CTGTCAAGGAGGC-GGCTTCAATACTT--CTTGGGCTTTTCTTGAGACAG 484
GL50803_16468      C----CGG--TTGTCAGGGAGGC-GAGTTCGAATCAT--CCTGGGCTTTTCTCGAAACAG 484
                         *    *                              *     **  *
```

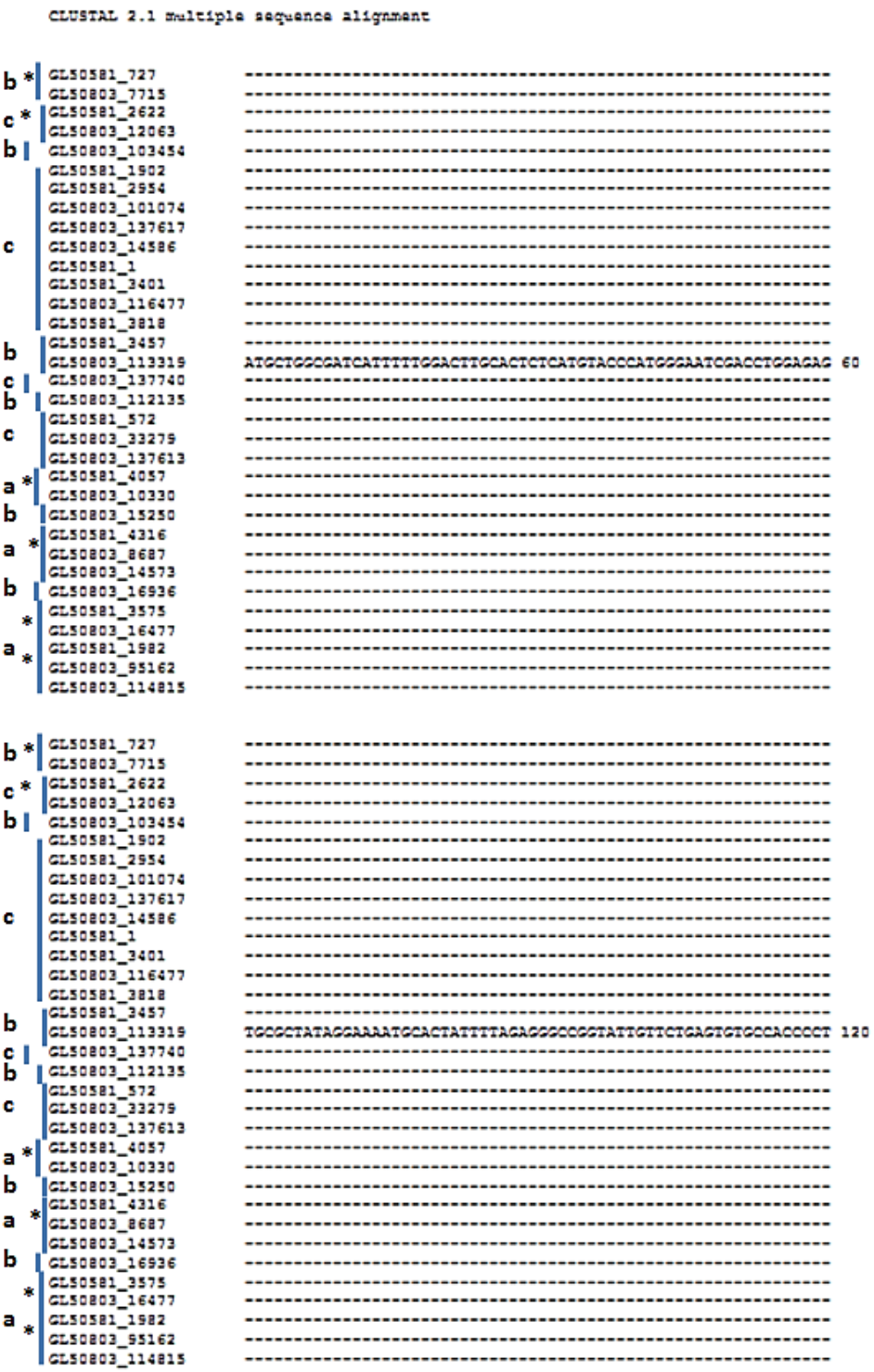
```
GL50581_2318        CAGGTATCCCCTTAGAAAGCTGTGTGAAGTATACAGACTATGATAA-GACAGAGTCCAGC 528
GL50803_17516       CGGGTATCCCCTTAGAAAGCTGTGTCAAGTATACAGATTATGACCA-AACACAGTCCAG- 527
GL50581_159         CTGGAGCAACCACAGCCGAATGCGTTAAATATGTTGACTACGGTAGCAGCGTGG--CAGC 536
GL50581_3619        CTGGAGCAACCACAGCCGAATGCGTTAAATATGTTGACTACGGTAGCAGCGTGG--CAGC 536
GL50803_10217       CAGGAGCAACCACAGCTGAATGCGTTAAGTACGTTGACTATGGTCACACCGTAG--CGTC 536
GL50581_2036        ACGGCACCACAACACTTGAGTGCGTGCCCTACACGGATGCTAATAAAGATATTT--CGTC 533
GL50803_15564       ACGGTACCACGACTCTCGAGTGTGTCCGCTACACAGACGCCGATAAGGACCTCT--CGTC 611
GL50581_78          CCGGAACAACCACCAACGAATGCGTGCCCTACCAGAGCGGAACTACCGGCGCTCG-CGGC 522
GL50803_16779       CTGGAACGACCACCGACGAGTGCGTCCCTTACCAGAGCGGCAGCACCGGCGCCCG-CGGC 522
GL50803_14019       CTGGAACGACCACCGACGAGTGCGTCCCCTACAAGAGCGGGTCGACAACTCTCCG-TGGC 525
GL50581_2946        CGGGGACTGTTTTGACCAGCTGCGTCGACTACACCGCCGGAGCCGACAATGTTGT-TAAG 528
GL50803_16160       CGGGGACTGTTCTGGCTAGCTGCGTGGGGTACACGGCCGGAGACGATGCTGTCGT-CAAG 720
GL50581_438         AAGGGGCAATAATGAGAGATTGCCTTCCCTACGTTAGCGGCGAGACTGGCCTTT--CGGG 542
GL50803_16468       AGGGAGCAGTAAAAAGTGACTGTCTTCCCTACACCAGTGGCGAAACTGGTAAAT--CAGG 542
                      **                  **  *    **

GL50581_2318        TACCCGTGTCCGAGTCTGTGCAACGATAATTC---TTCCCTAGTGCTTTACAAATCT--G 583
GL50803_17516       --GCCATGCCCGAGTACGTGTGATGACGATTC---CTTCCTCGAAGTTTATAAGCCC--G 580
GL50581_159         T--GCATGTCCCACGACCTGTGATGACGGGTC---TCAGATCCAGTTCTATAAAGCG--C 589
GL50581_3619        T--GCATGTCCCACGACCTGTGATGACGGGTC---TCAGATCCAGTTCTATAAAGCG--C 589
GL50803_10217       T--CCGTGCCCTGCGGTCTGTGACGACGGATC---TCCGATACAGCTCTACAAGGCA--C 589
GL50581_2036        C--CCGTGCCCAGACGCCTGTGCCGATGGGTC---AGAGATCCGGCTTGTCAAGGCA--G 586
GL50803_15564       T--CCATGCCCCGCATTATGCGACGATGGGTC---CGAGATTCAACTCGTCAAGGCC--G 664
GL50581_78          A--CC-TGCCCGACGAAGTGCGCTGATGGT---GGCGAGCTCTCTACGGTCAAGGCCAAG 576
GL50803_16779       A--CC-TGCCCGACGAAGTGCGCCGATGGC---AGCGACCTGCCCATCTACAAGGCCACA 576
GL50803_14019       A--CC-TGCCCGACGAAGTGCGCCGATGGATCGAGCAAGGTTCACTTGGCCACCGCCACG 582
GL50581_2946        T--TC-TGCCCCAAGACGTGCGATGACGGATC---TGCCGTGGAGAATGTCTTCGCT--- 579
GL50803_16160       T--TT-TGCCCCCAGAAGTGCGACGACGGATC---CGCTGTGGAGAACGTCGTCGCC--- 771
GL50581_438         T--GAGTGTCCCACAACATGCCAAGATGGGAC---TCTTCTTAATGATACCATTCACTAT 597
GL50803_16468       T--GAGTGCCCCACGACGTGCCAGGATGGAAC---CCCTGTCGAGAGCGCCTTCCATTAC 597
                          ** **       **    **              *

GL50581_2318        ATGGTTATGAGGGCGT-TGGATTTAATCCTGAAA-AGCTGAGACGCGCTATTGCTC-TCA 640
GL50803_17516       ACGGTTATGAAGGCGT-CGGACTCAATTGTGAAA-GGCTGAAGCGTGCAGTTGCTC-TCA 637
GL50581_159         ATGGATATGGTCAAGT--GAGCAAGAGCGTGCC--TGCTATAA--TGCAGATGCTCGTCA 643
GL50581_3619        ATGGATATGGTCAACT--GAGCAAGAGCGTGCC--TGCTATAA--TGCAGATGCTCGTCA 643
GL50803_10217       ATGGATATGGACAGGT--GAGCAAGAGTGTTCC--GGCCATAA--TGGGAATGCTCGTCG 643
GL50581_2036        ATGGCTGCCTGGATTA-CAGCGGAAAC-GTAAC--CGCCATCA--TGCAGGCGCTCGCCA 640
GL50803_15564       ACGGCTGCCTCGACTA-CAGCGGCAAC-GTGAC--CGCCATCA--TGCAGACCCTCGCCA 718
GL50581_78          AAGGCTG--TAGACTA-CGGCCTGGACTGCGAT--C-TGATCA--TGAAGGCTCTTGTTA 628
GL50803_16779       AAGGCTG--TCGACTA-TGGTCTGGACTGCGAC--C-TTATCA--TGAAGGCTCTCGCCA 628
GL50803_14019       TCTTACA--AGGATTA-CGGGCTTGAC-ATACC--CGCCATGA--TGAAGGCCCTGTCCA 634
GL50581_2946        ---GCCAG-TGGCTCG-AAGTCCGGAT-----C--TGCCATCG-----ATGTGCTTCTTT 622
GL50803_16160       ---ACTAG-CGGGTCG-AAGTCCGGGT-----C--TGCAATCG-----ACGTCCTTCTCG 814
GL50581_438         AAAGCTGTTTCGGCTAGCCATCTTAAGAACTATAATGAAATTA--TGACGTCTCTTCTTA 655
GL50803_16468       AAAGCTGCCTCGGCCAGCCGTCTTAGCAACTACAATGAGATCA--TGGTGTCCCTTCTTG 655
                                                                    **

GL50581_2318        ---GAGGTCCTATGCAAGCTATGTTTACTGTTTATGAGGACTTCGCATACTACTTAGAGG 697
GL50803_17516       ---GAGGACCCATGCAGGCTATGTTCACCGTTTATGAGGATTTCACATACTATTTGGAGG 694
GL50581_159         GTGGAGGACCGGTGCAGACTATGATCGTTGTTTACGCTGATTTATTATACTACGCTGGCG 703
GL50581_3619        GTGGAGGACCGGTGCAGACTATGATCGTTGTTTACGCTGATTTATTATACTACGCTGGCG 703
GL50803_10217       CTGGCGGACCGCTGCAGACCATGATCGTCGTGTACGCTGACCTGTCGTACTATGAAAGCG 703
GL50581_2036        ACGATGGCCCTGTCCAGGCAAGTATGGCTGTCTACCGGGACTTCCTGTACTACAGGAGCG 700
GL50803_15564       ACGACGGGCCTGTGCAGGCAGTCATGAGCGTCTATCGTGACTTCCTTTACTACCGAGGAG 778
GL50581_78          CGGGTGGTCCTCTCCAGACCGCTTTCACCGTTTACTCCGACTTCATGTACTATGAGGGCG 688
GL50803_16779       CGGGCGGCCCTCTCCAGACCGCCTTCACCGTTTACTCCGATTTCATGTACTATGAGGGCG 688
GL50803_14019       CGAGCGGCCCTCTCCAGGTCGCCTTCTTGGTCTACTCCGACTTCATGTACTACGAGAGCG 694
GL50581_2946        CTCACGGCCCCGTCGTGGCGACTTTCAACGTTGCGCAGGACTTCATGTACTACAAGAGTG 682
GL50803_16160       CTCACGGCCCCGTTGTAGCGACCTTCAACGTCGCGCAGGACTTCATGTATTACAAGAGCG 874
GL50581_438         ATGAGGGACCAGTGCAGACAGGGTTCTATGTCCATGAAGACTTTCTGTACTATGTCGGTG 715
GL50803_16468       CCGATGGGCCAGTGCAAACGGGCTTCTACGTCCACGAGGACTTTCTATATTATGTCGGTG 715
                         ** **  *           *    **       **  *    ** **       *
```

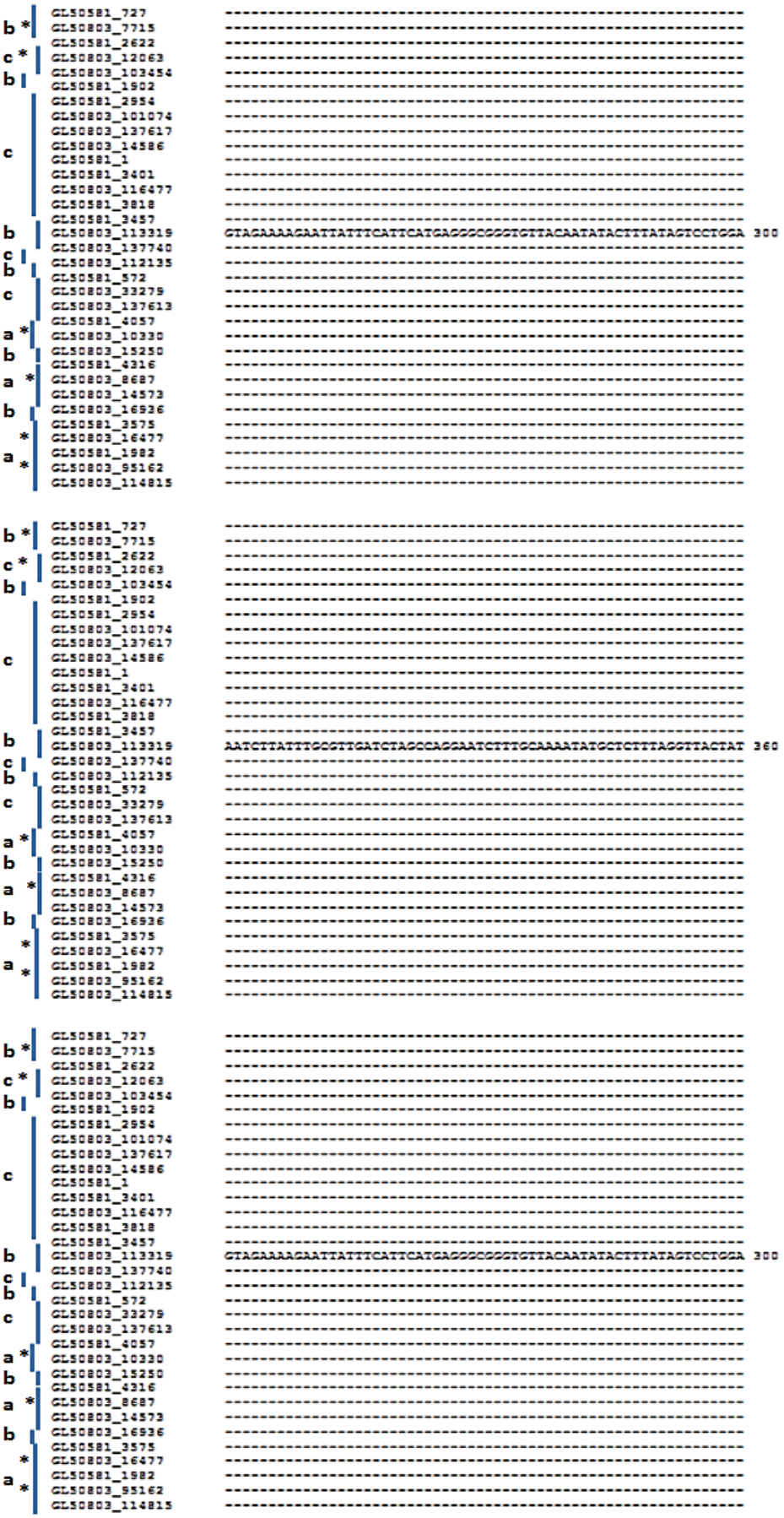
```
GL50581_2318        GCATTTATTCCCATGTATATGG---TGGTACTGCTGGTTATCTCTCCGTTGAGATAGTTG 754
GL50803_17516       GCATCTACTCCTACACATATGG---TAACAGAGTTGGTTTTCTTTCTGTTGAGATAGTCG 751
GL50581_159         GTGTCTATAGACACACCTACGGACCAATTAGCAACGGGCTTCATGCGCTAGAAATGGTGG 763
GL50581_3619        GTGTCTATAGACACACCTACGGACCAATTAGCAACGGGCTTCATGCGCTAGAAATGGTGG 763
GL50803_10217       GTGTCTATAAACACACCTATGGAACCATCAATCTTGGGTTCCATGCTTTGGAAATAGTTG 763
GL50581_2036        GGGTCTACCGCCACGTCTATGGAT---CGCAGATCTCCAGCCACGCCGTCGAAATCATTG 757
GL50803_15564       GCGTCTACAAGCACGTCTACGGGA---TCCAGATCTCCAGCCATGCGGTTGAGATCATCG 835
GL50581_78          GCGTCTACCAGCACATGTCTGGGAGAGTTGAGGGC---GGCCACGCCGTTGAGATGGTCG 745
GL50803_16779       GCGTCTACCAGCACACCTACGGAAGAGTCGAGGGT---GGTCACGCAGTCGAGATGGTGG 745
GL50803_14019       GCGTCTACCAGCACACCTACGGATACATGGAGGGT---GGCCACGCAGTCGAGATGGTGG 751
GL50581_2946        GTGTCTATCAGCACAGGTGGGGTG---TCTGGCTCGGCGGTCACGCCGTAGAGGTTGTCG 739
GL50803_16160       GTGTTTACCAGCACAGGTGGGGGC---TCTGGCTTGGCGGCCACGCCGTAGAGATAATTG 931
GL50581_438         GTATCTATCACAAGACCTACGGTT---CGAGCATTGGTGGACATGCGGTTCTTATCGTCG 772
GL50803_16468       GGATCTATCACAAGGTTTATGGCA---CGAGCCTTGGTGGGCATGCAGTTCTTATTGTTG 772
                    *  * **     *    *  **                   *   *  *     *  * *

GL50581_2318        GCTATGGCACA-------AGCGACGAAGGACAAGACTATTGGATCGTAAAGAATTACTGG 807
GL50803_17516       GCTATGGCACA-------AGCGACGAAGGGCAAGACTATTGGATTGTAAAGAACTACTGG 804
GL50581_159         GCTATGGCACT-------ACAGACGATGGGACAGACTATTGGACTATAAAAAACTCCTGG 816
GL50581_3619        GCTATGGCACT-------ACAGACGATGGGACAGACTATTGGACTATAAAAAACTCCTGG 816
GL50803_10217       GCTACGGGACC-------ACTGATGACGGGACGGACTACTGGATTATAAAGAACTCCTGG 816
GL50581_2036        GCTACGGAGCGGCCGATGAT-GAGGACAGCACTCCATACTGGATTGTGAAGAACTCTCTC 816
GL50803_15564       GCTACGGGACGACTGACGAC-GAGGAGCGCATCCCGTACTGGATCGTGAAGAACTCTCTC 894
GL50581_78          GCTACGGCACG------GAT-GAGTACGATGTCGACTACTGGATCATCCGCAACTCCTGG 798
GL50803_16779       GCTACGGCACA------GAC-GAGTACGACGTTGACTACTGGATCATCCGCAACTCCTGG 798
GL50803_14019       GCTACGGCACA------GAC-GACGACGGCGTTGACTACTGGATCATCCGCAACTCCTGG 804
GL50581_2946        GCTATGGTGTG-------ACAGACTCTGGGCTCGACTACTGGACCGTCCGCAACTCCTGG 792
GL50803_16160       GGTATGGCGTG-------ACAGACTCCGGCCTCGACTACTGGACTGTCCGCAACTCCTGG 984
GL50581_438         GCTACGGGAGC-------AT-GA--ATAATCACGATTACTGGATTGTACGTAATTCTTGG 822
GL50803_16468       GCTATGGAAGC-------AT-GA--ACAATCACGACTATTGGATTGTTCGTAATTCTTGG 822
                    * ** **            *  **              ** ****   *    ** *

GL50581_2318        GGTTCTAATTGGGGCGAAGATGGCTACTTCCGCATTGTCCGTGGCCAGAATGAATGCCAG 867
GL50803_17516       GGCCCGGGCTGGGGAGAAGATGGCTATTTCCGTATTGTTAGGGGCCAGAATGAATGCCAG 864
GL50581_159         GGGTCCGACTGGGGAGAGGATGGTTACTTCCGAATTGTTCGCGGCGTCAATGAGTGCAGG 876
GL50581_3619        GGGTCCGACTGGGGAGAGGATGGTTACTTCCGAATTGTTCGCGGCGTCAATGAGTGCAGG 876
GL50803_10217       GGCCCCGATTGGGGCGAAAACGGGTACTTCCGGATTGTCCGCGGAGTCAATGAGTGCAGG 876
GL50581_2036        GGATCCGGCTGGGGAGAGGAGGGGTACTTTAACATAGTCCGGGGGTCTAACGAATGTGAC 876
GL50803_15564       GGACCCAACTGGGGCGAGGAGGGCTACTTCAACATAGTCCGCGGGTCGAACGAGTGCGAC 954
GL50581_78          GGTCCCGACTGGGGCGAGGACGGCTACTTCCGCATCATCCGTATGACGAATGAGTGCGGT 858
GL50803_16779       GGCCCTGACTGGGGTGAGGACGGCTACTTCCGCATCATTCGCATGACGAACGAGTGCGGC 858
GL50803_14019       GGTCCTGACTGGGGCGAGGACGGCTACTTCCGCATGATCCGCGGCATAAACGACTGTAGC 864
GL50581_2946        GGTCCTGACTGGGGCGAGGACGGCTATTTTAGGATTGTTCGCGGGTCGGACGAATGCGGT 852
GL50803_16160       GGCCCTGACTGGGGTGAGGACGGCTACTTTAGAATCGTTCGCGGCGGCGACGAGTGTGGC 1044
GL50581_438         GGATCGGATTGGGGTGAGAATGGCTACTTCAGGATACTCCGCGGAACCAATGAATGCGGG 882
GL50803_16468       GGGTCGGATTGGGGTGAAAATGGCTACTTCAGGATACTCCGCGGAACAAACGAGTGTGGC 882
                    **  *    ***** **  * ** ** **    **  *  *         * ** **

GL50581_2318        ATTGAGGA-GGCTGTCTATGGTGCCATTATATACGCCAATACACAGTCTGCAACTACTCA 926
GL50803_17516       ATTGAGAA-CTCTGCCTATGGTGCTATAATAAGCCCTAACAAACCATCCATGGCCACCAA 923
GL50581_159         ATAGAGGA-CGAGATCTATGCTGC----GTA----------------CTTTGAGTAG--- 912
GL50581_3619        ATAGAGGA-CGAGATCTATGCTGC----GTA----------------CTTTGAGTAG--- 912
GL50803_10217       ATAGAGGA-TGAAATCTATGCTGT----ATA----------------TCTTGACTAG--- 912
GL50581_2036        ATTGA-GAGYGCTGTGTACA-GTGGCCTCTTCCTTTC----------CTCTGAGTGA--- 921
GL50803_15564       ATCGA-GAGCGCCGTGTATA-GCGGGCTCTTCCTCTC----------CTCTAAGTGA--- 999
GL50581_78          ATCGAGGAGCAG-GT-TATGGGCGGCATCTTCGA----------------AGAGTGA--- 897
GL50803_16779       ATTGAAGAGCAA-GT-AATAGGTGGGTTCTTTGA----------------GAACTAA--- 897
GL50803_14019       ATCGAGGAGCAG-GCCTATGCG-GGCTTCTTCGA----------------TGAGTAA--- 903
GL50581_2946        ATAGAGCAGGAAGGTTTCCA------CTATGCAGATG----------T--AGAGTGA--- 891
GL50803_16160       ATAGAGCACGAGGGCTTCCA------CTATGCAGACG----------T--AGAGTGA--- 1083
GL50581_438         ATCGAAAA-TAATGCATGGCAAGTTGCTGTG----------------AAGTGA--- 918
GL50803_16468       ATTGAGAA-GAATGCGTGGCAGGTAGCAGTG----------------AAGTAA--- 918
                    ** **  *                      *
```

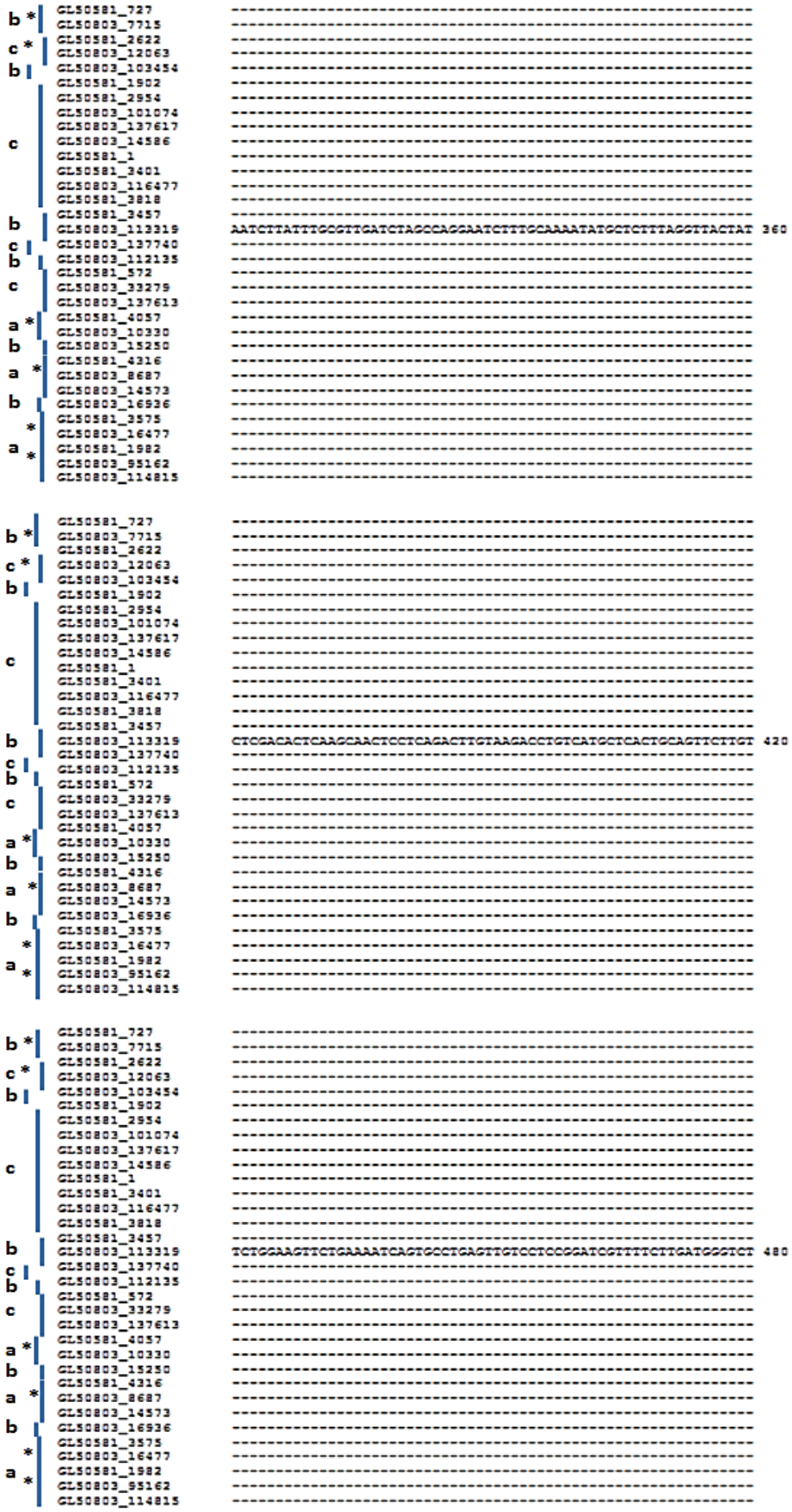
```
GL50581_2318        GTAA 930
GL50803_17516       GTGA 927
GL50581_159         ----
GL50581_3619        ----
GL50803_10217       ----
GL50581_2036        ----
GL50803_15564       ----
GL50581_78          ----
GL50803_16779       ----
GL50803_14019       ----
GL50581_2946        ----
GL50803_16160       ----
GL50581_438         ----
GL50803_16468       ----
```

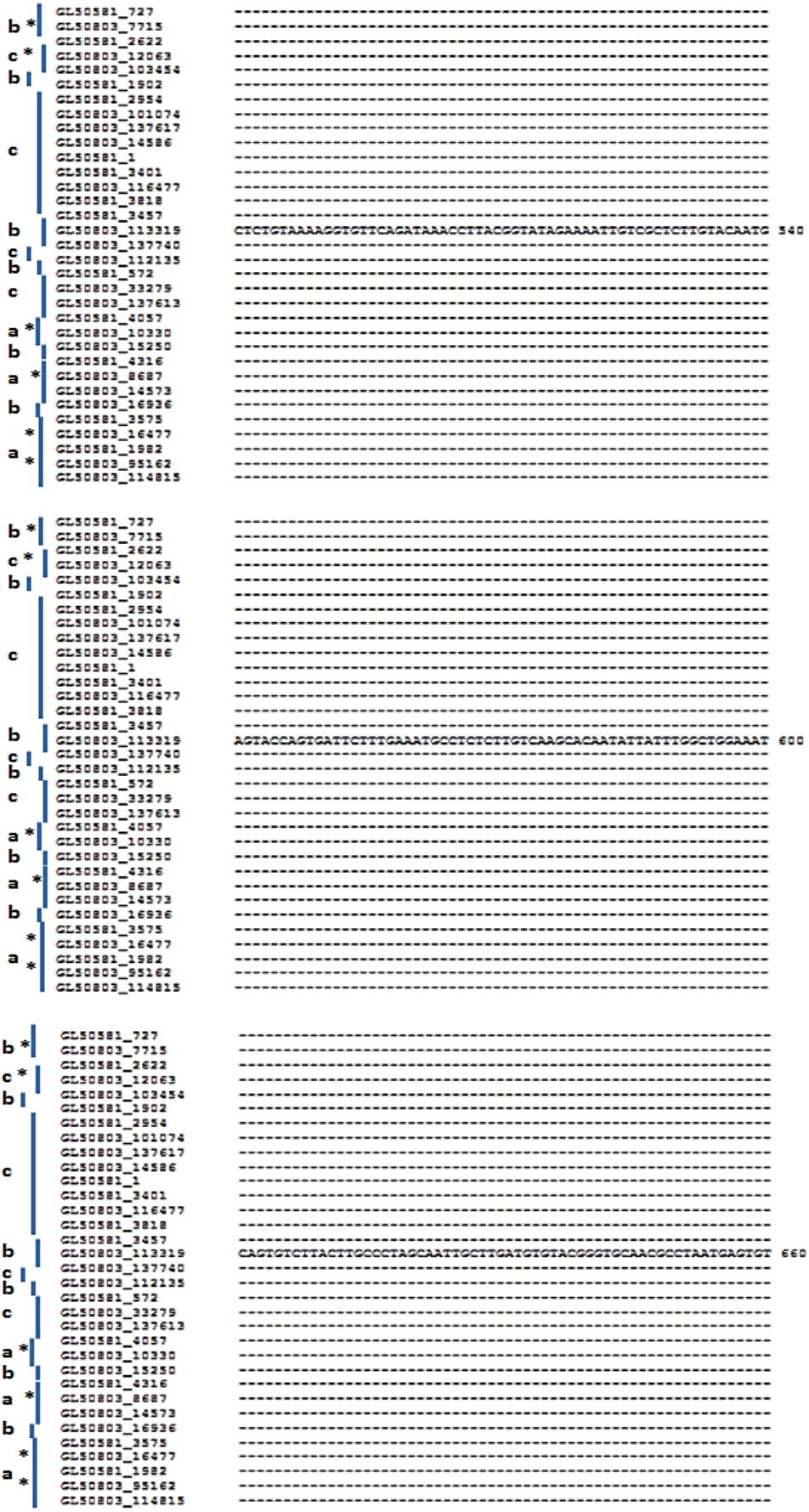
Appendix XI:

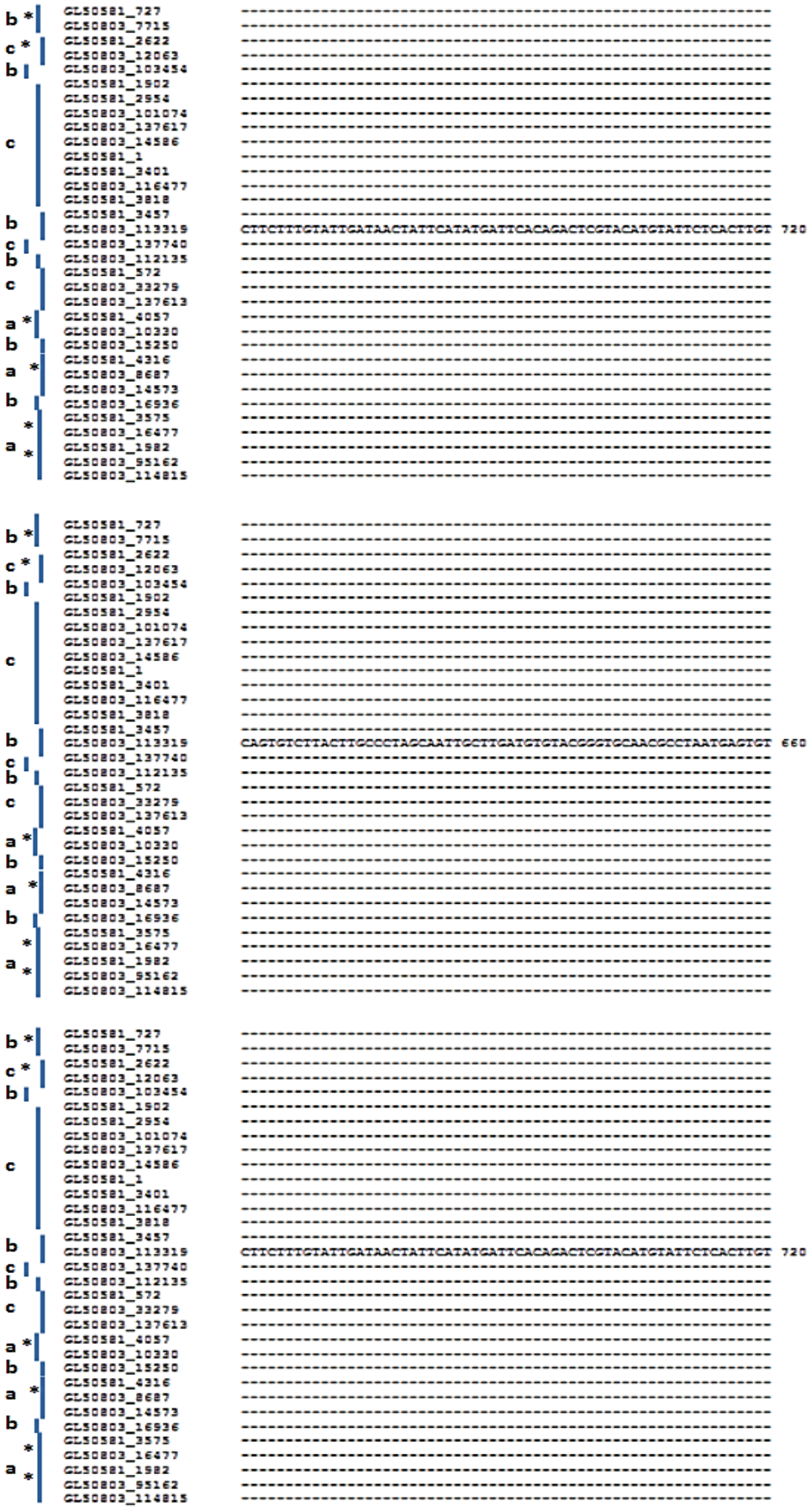
CLustalW sequence alignments of the conserved secreted tenascin, high cysteine proteins and VSPs. All the GL50581_ accession numbers represent GS isolate proteins and all the GL50803_ represent WB isolate proteins. Orthologs are shown by *. **a**: Tenascin, **b**: High cysteine membrane protein, **c:** VSP.

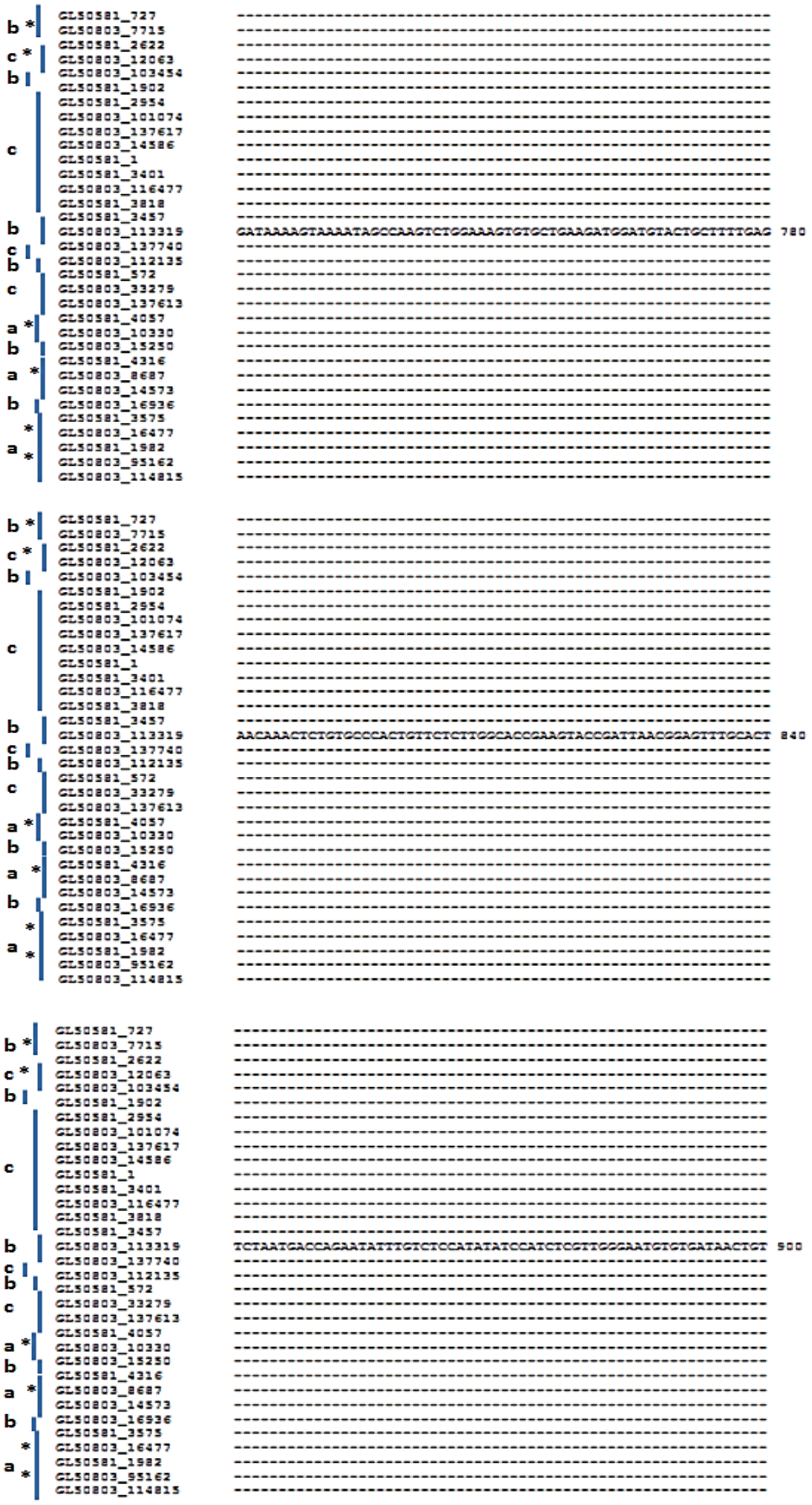



GLS0581_727
GLS0803_7715
GLS0581_2622
GLS0803_12063
GLS0803_103454
GLS0581_1902
GLS0581_2954
GLS0803_101074
GLS0803_137617
GLS0803_14586
GLS0581_1
GLS0581_3401
GLS0803_116477
GLS0581_3818
GLS0581_3457
GLS0803_113319 GTAGAAAAGAATTATTTCATTCATGAGGGCGGGTGTTACAATATACTTTATAGTCCTGGA 300
GLS0803_137740
GLS0803_112135
GLS0581_572
GLS0803_33279
GLS0803_137613
GLS0581_4057
GLS0803_10330
GLS0803_15250
GLS0581_4316
GLS0803_8687
GLS0803_14573
GLS0803_16936
GLS0581_3575
GLS0803_16477
GLS0581_1982
GLS0803_95162
GLS0803_114815
GLS0581_727
GLS0803_7715
GLS0581_2622
GLS0803_12063
GLS0803_103454
GLS0581_1902
GLS0581_2954
GLS0803_101074
GLS0803_137617
GLS0803_14586
GLS0581_1
GLS0581_3401
GLS0803_116477
GLS0581_3818
GLS0581_3457
GLS0803_113319 AATCTTATTTGCGTTGATCTAGCCAGGAATCTTTGCAAAATATGCTCTTTAGGTTACTAT 360
GLS0803_137740
GLS0803_112135
GLS0581_572
GLS0803_33279
GLS0803_137613
GLS0581_4057
GLS0803_10330
GLS0803_15250
GLS0581_4316
GLS0803_8687
GLS0803_14573
GLS0803_16936
GLS0581_3575
GLS0803_16477
GLS0581_1982
GLS0803_95162
GLS0803_114815
GLS0581_727
GLS0803_7715
GLS0581_2622
GLS0803_12063
GLS0803_103454
GLS0581_1902
GLS0581_2954
GLS0803_101074
GLS0803_137617
GLS0803_14586
GLS0581_1
GLS0581_3401
GLS0803_116477
GLS0581_3818
GLS0581_3457
GLS0803_113319 GTAGAAAAGAATTATTTCATTCATGAGGGCGGGTGTTACAATATACTTTATAGTCCTGGA 300
GLS0803_137740
GLS0803_112135
GLS0581_572
GLS0803_33279
GLS0803_137613
GLS0581_4057
GLS0803_10330
GLS0803_15250
GLS0581_4316
GLS0803_8687
GLS0803_14573
GLS0803_16936
GLS0581_3575
GLS0803_16477
GLS0581_1982
GLS0803_95162
GLS0803_114815

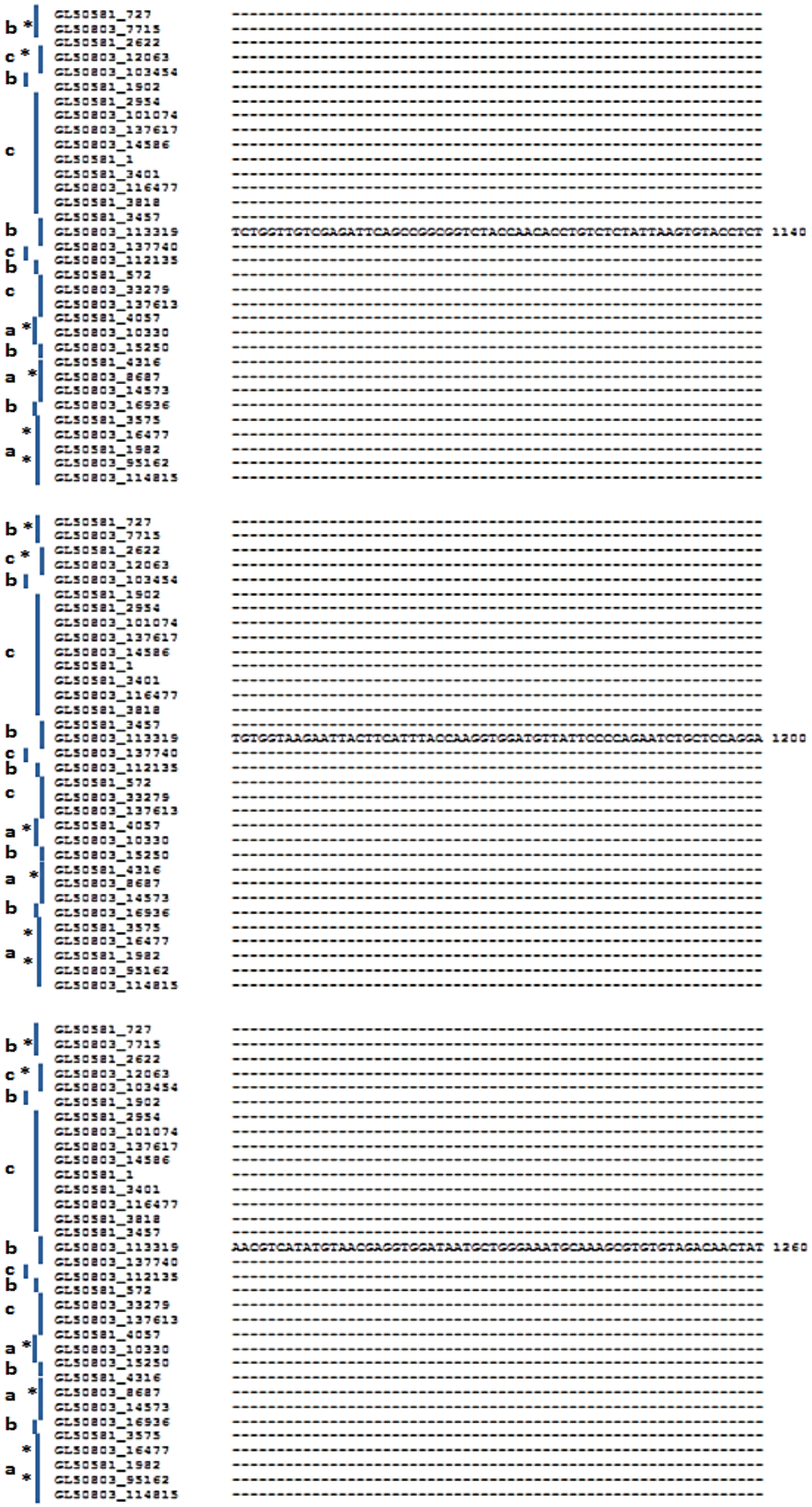

b * GL50581_727 ----------
GL50803_7715 ----------
c * GL50581_2622 ----------
GL50803_12063 ----------
b GL50803_103454 ----------
GL50581_1902 ----------
GL50581_2954 ----------
GL50803_101074 ----------
GL50803_137617 ----------
c GL50803_14586 ----------
GL50581_1 ----------
GL50581_3401 ----------
GL50803_116477 ----------
GL50581_3818 ----------
GL50581_3457 ----------
b GL50803_113319 AATCTTATTTGCGTTGATCTAGCCAGGAATCTTTGCAAAATATGCTCTTTAGGTTACTAT 360
c GL50803_137740 ----------
b GL50803_112135 ----------
GL50581_572 ----------
c GL50803_33279 ----------
GL50803_137613 ----------
a * GL50581_4057 ----------
GL50803_10330 ----------
b GL50803_15250 ----------
GL50581_4316 ----------
a * GL50803_8687 ----------
GL50803_14573 ----------
b GL50803_16936 ----------
GL50581_3575 ----------
* GL50803_16477 ----------
a * GL50581_1982 ----------
GL50803_95162 ----------
GL50803_114815 ----------
GL50581_727 ----------
b * GL50803_7715 ----------
GL50581_2622 ----------
c * GL50803_12063 ----------
b GL50803_103454 ----------
GL50581_1902 ----------
GL50581_2954 ----------
GL50803_101074 ----------
GL50803_137617 ----------
c GL50803_14586 ----------
GL50581_1 ----------
GL50581_3401 ----------
GL50803_116477 ----------
GL50581_3818 ----------
GL50581_3457 ----------
b GL50803_113319 CTCGACACTCAAGCAACTCCTCAGACTTGTAAGACCTGTCATGCTCACTGCAGTTCTTGT 420
c GL50803_137740 ----------
b GL50803_112135 ----------
GL50581_572 ----------
c GL50803_33279 ----------
GL50803_137613 ----------
a * GL50581_4057 ----------
GL50803_10330 ----------
b GL50803_15250 ----------
GL50581_4316 ----------
a * GL50803_8687 ----------
GL50803_14573 ----------
b GL50803_16936 ----------
GL50581_3575 ----------
* GL50803_16477 ----------
a * GL50581_1982 ----------
GL50803_95162 ----------
GL50803_114815 ----------
GL50581_727 ----------
b * GL50803_7715 ----------
GL50581_2622 ----------
c * GL50803_12063 ----------
b GL50803_103454 ----------
GL50581_1902 ----------
GL50581_2954 ----------
GL50803_101074 ----------
GL50803_137617 ----------
c GL50803_14586 ----------
GL50581_1 ----------
GL50581_3401 ----------
GL50803_116477 ----------
GL50581_3818 ----------
GL50581_3457 ----------
b GL50803_113319 TCTGGAAGTTCTGAAAATCAGTGCCTGAGTTGTCCTCCGGATCGTTTTCTTGATGGGTCT 480
c GL50803_137740 ----------
b GL50803_112135 ----------
GL50581_572 ----------
c GL50803_33279 ----------
GL50803_137613 ----------
a * GL50581_4057 ----------
GL50803_10330 ----------
b GL50803_15250 ----------
GL50581_4316 ----------
a * GL50803_8687 ----------
GL50803_14573 ----------
b GL50803_16936 ----------
GL50581_3575 ----------
* GL50803_16477 ----------
a * GL50581_1982 ----------
GL50803_95162 ----------
GL50803_114815 ----------

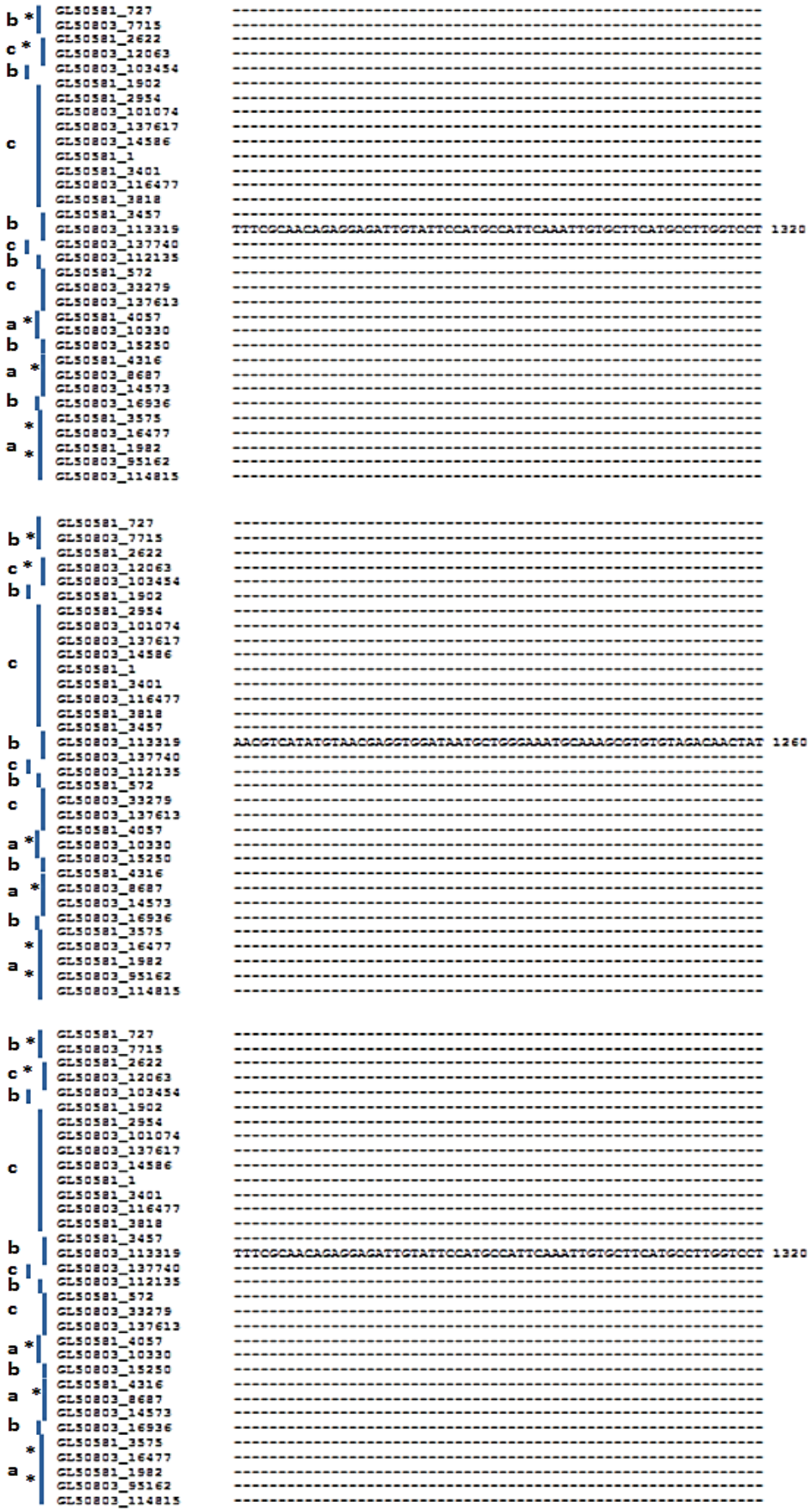

GLS0581_727
GLS0803_7715
GLS0581_2622
GLS0803_12063
GLS0803_103454
GLS0581_1902
GLS0581_2954
GLS0803_101074
GLS0803_137617
GLS0803_14586
GLS0581_1
GLS0581_3401
GLS0803_116477
GLS0581_3818
GLS0581_3457
GLS0803_113319
GLS0803_137740
GLS0803_112135
GLS0581_572
GLS0803_33279
GLS0803_137613
GLS0581_4057
GLS0803_10330
GLS0803_15250
GLS0581_4316
GLS0803_8687
GLS0803_14573
GLS0803_16936
GLS0581_3575
GLS0803_16477
GLS0581_1982
GLS0803_95162
GLS0803_114815
CTCTGTAAAAGGTGTTCAGATAAACCTTACGGTATAGAAAATTGTCGCTCTTGTACAATG 540
AGTACCAGTGATTCTTTGAAATGCCTCTCTTGTCAAGCACAATATTATTTGGCTGGAAAT 600
CAGTGTCTTACTTGCCCTAGCAATTGCTTGATGTGTACGGGTGCAACGCCTAATGAGTGT 660

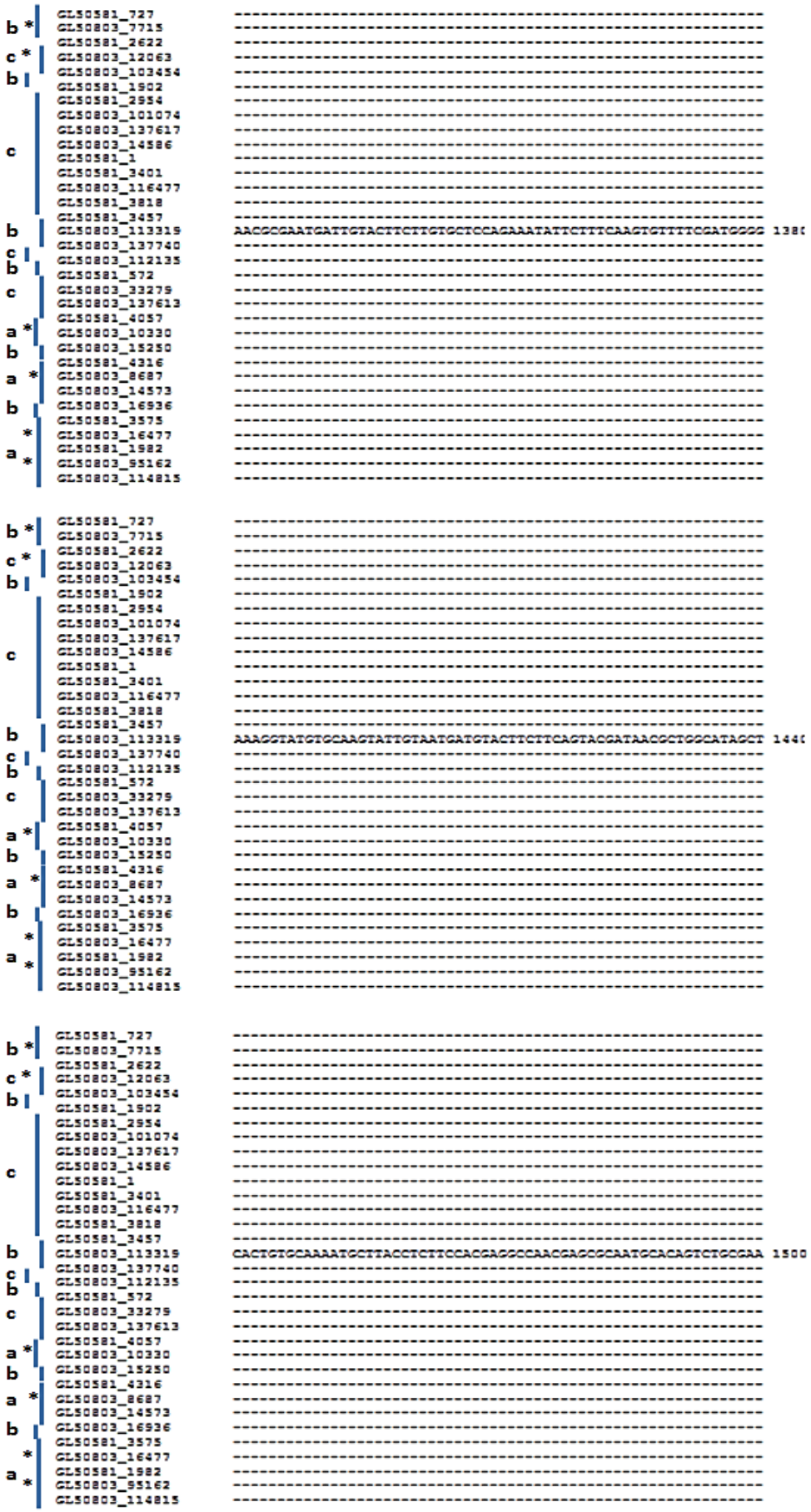

b * GLS0581_727
GLS0803_7715
c * GLS0581_2622
GLS0803_12063
b GLS0803_103454
GLS0581_1902
GLS0581_2954
GLS0803_101074
GLS0803_137617
c GLS0803_14586
GLS0581_1
GLS0581_3401
GLS0803_116477
GLS0581_3818
GLS0581_3457
b GLS0803_113319 CTTCTTTGTATTGATAACTATTCATATGATTCACAGACTCGTACATGTATTCTCACTTGT 720
c GLS0803_137740
b GLS0803_112135
GLS0581_572
c GLS0803_33279
GLS0803_137613
a * GLS0581_4057
GLS0803_10330
b GLS0803_15250
GLS0581_4316
a * GLS0803_8687
GLS0803_14573
b GLS0803_16936
GLS0581_3575
* GLS0803_16477
a * GLS0581_1982
GLS0803_95162
GLS0803_114815
b * GLS0581_727
GLS0803_7715
c * GLS0581_2622
GLS0803_12063
b GLS0803_103454
GLS0581_1902
GLS0581_2954
GLS0803_101074
GLS0803_137617
c GLS0803_14586
GLS0581_1
GLS0581_3401
GLS0803_116477
GLS0581_3818
GLS0581_3457
b GLS0803_113319 CAGTGTCTTACTTGCCCTAGCAATTGCTTGATGTGTACGGGTGCAACGCCTAATGAGTGT 660
c GLS0803_137740
b GLS0803_112135
GLS0581_572
c GLS0803_33279
GLS0803_137613
a * GLS0581_4057
GLS0803_10330
b GLS0803_15250
GLS0581_4316
a * GLS0803_8687
GLS0803_14573
b GLS0803_16936
GLS0581_3575
* GLS0803_16477
a * GLS0581_1982
GLS0803_95162
GLS0803_114815
b * GLS0581_727
GLS0803_7715
c * GLS0581_2622
GLS0803_12063
b GLS0803_103454
GLS0581_1902
GLS0581_2954
GLS0803_101074
GLS0803_137617
c GLS0803_14586
GLS0581_1
GLS0581_3401
GLS0803_116477
GLS0581_3818
GLS0581_3457
b GLS0803_113319 CTTCTTTGTATTGATAACTATTCATATGATTCACAGACTCGTACATGTATTCTCACTTGT 720
c GLS0803_137740
b GLS0803_112135
GLS0581_572
c GLS0803_33279
GLS0803_137613
a * GLS0581_4057
GLS0803_10330
b GLS0803_15250
GLS0581_4316
a * GLS0803_8687
GLS0803_14573
b GLS0803_16936
GLS0581_3575
* GLS0803_16477
a * GLS0581_1982
GLS0803_95162
GLS0803_114815

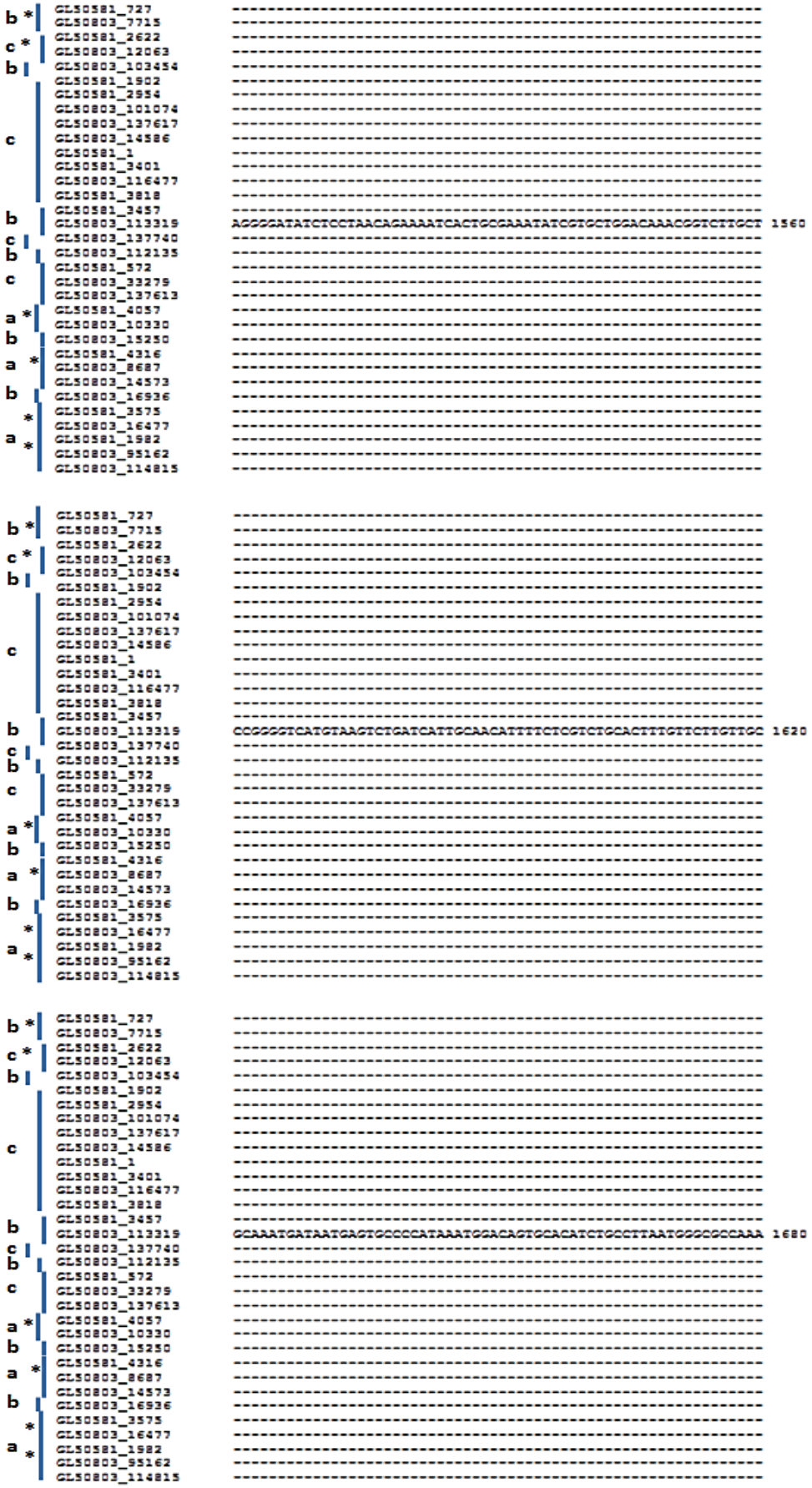

b * GL50581_727 ---------- 780
b * GL50803_7715 ---------- 780
c * GL50581_2622 ---------- 780
c * GL50803_12063 ---------- 780
b GL50803_103454 ---------- 780
GL50581_1902 ---------- 780
GL50581_2954 ---------- 780
GL50803_101074 ---------- 780
GL50803_137617 ---------- 780
c GL50803_14586 ---------- 780
GL50581_1 ---------- 780
GL50581_3401 ---------- 780
GL50803_116477 ---------- 780
GL50581_3818 ---------- 780
GL50581_3457 ---------- 780
b GL50803_113319 GATAAAAGTAAAATAGCCAAGTCTGGAAAGTGTGCTGAAGATGGATGTACTGCTTTTGAG 780
c GL50803_137740 ---------- 780
b GL50803_112135 ---------- 780
GL50581_572 ---------- 780
c GL50803_33279 ---------- 780
GL50803_137613 ---------- 780
a * GL50581_4057 ---------- 780
GL50803_10330 ---------- 780
b GL50803_15250 ---------- 780
GL50581_4316 ---------- 780
a * GL50803_8687 ---------- 780
GL50803_14573 ---------- 780
b GL50803_16936 ---------- 780
GL50581_3575 ---------- 780
* GL50803_16477 ---------- 780
a * GL50581_1982 ---------- 780
GL50803_95162 ---------- 780
GL50803_114815 ---------- 780
b * GL50581_727 ---------- 840
b * GL50803_7715 ---------- 840
c * GL50581_2622 ---------- 840
c * GL50803_12063 ---------- 840
b GL50803_103454 ---------- 840
GL50581_1902 ---------- 840
GL50581_2954 ---------- 840
GL50803_101074 ---------- 840
GL50803_137617 ---------- 840
c GL50803_14586 ---------- 840
GL50581_1 ---------- 840
GL50581_3401 ---------- 840
GL50803_116477 ---------- 840
GL50581_3818 ---------- 840
GL50581_3457 ---------- 840
b GL50803_113319 AACAAACTCTGTGCCCACTGTTCTCTTGGCACCGAAGTACCGATTAACGGAGTTTGCACT 840
c GL50803_137740 ---------- 840
b GL50803_112135 ---------- 840
GL50581_572 ---------- 840
c GL50803_33279 ---------- 840
GL50803_137613 ---------- 840
a * GL50581_4057 ---------- 840
GL50803_10330 ---------- 840
b GL50803_15250 ---------- 840
GL50581_4316 ---------- 840
a * GL50803_8687 ---------- 840
GL50803_14573 ---------- 840
b GL50803_16936 ---------- 840
GL50581_3575 ---------- 840
* GL50803_16477 ---------- 840
a * GL50581_1982 ---------- 840
GL50803_95162 ---------- 840
GL50803_114815 ---------- 840
b * GL50581_727 ---------- 900
b * GL50803_7715 ---------- 900
c * GL50581_2622 ---------- 900
c * GL50803_12063 ---------- 900
b GL50803_103454 ---------- 900
GL50581_1902 ---------- 900
GL50581_2954 ---------- 900
GL50803_101074 ---------- 900
GL50803_137617 ---------- 900
c GL50803_14586 ---------- 900
GL50581_1 ---------- 900
GL50581_3401 ---------- 900
GL50803_116477 ---------- 900
GL50581_3818 ---------- 900
GL50581_3457 ---------- 900
b GL50803_113319 TCTAATGACCAGAATATTTGTCTCCATATATCCATCTCGTTGGGAATGTGTGATAACTGT 900
c GL50803_137740 ---------- 900
b GL50803_112135 ---------- 900
GL50581_572 ---------- 900
c GL50803_33279 ---------- 900
GL50803_137613 ---------- 900
a * GL50581_4057 ---------- 900
GL50803_10330 ---------- 900
b GL50803_15250 ---------- 900
GL50581_4316 ---------- 900
a * GL50803_8687 ---------- 900
GL50803_14573 ---------- 900
b GL50803_16936 ---------- 900
GL50581_3575 ---------- 900
* GL50803_16477 ---------- 900
a * GL50581_1982 ---------- 900
GL50803_95162 ---------- 900
GL50803_114815 ---------- 900

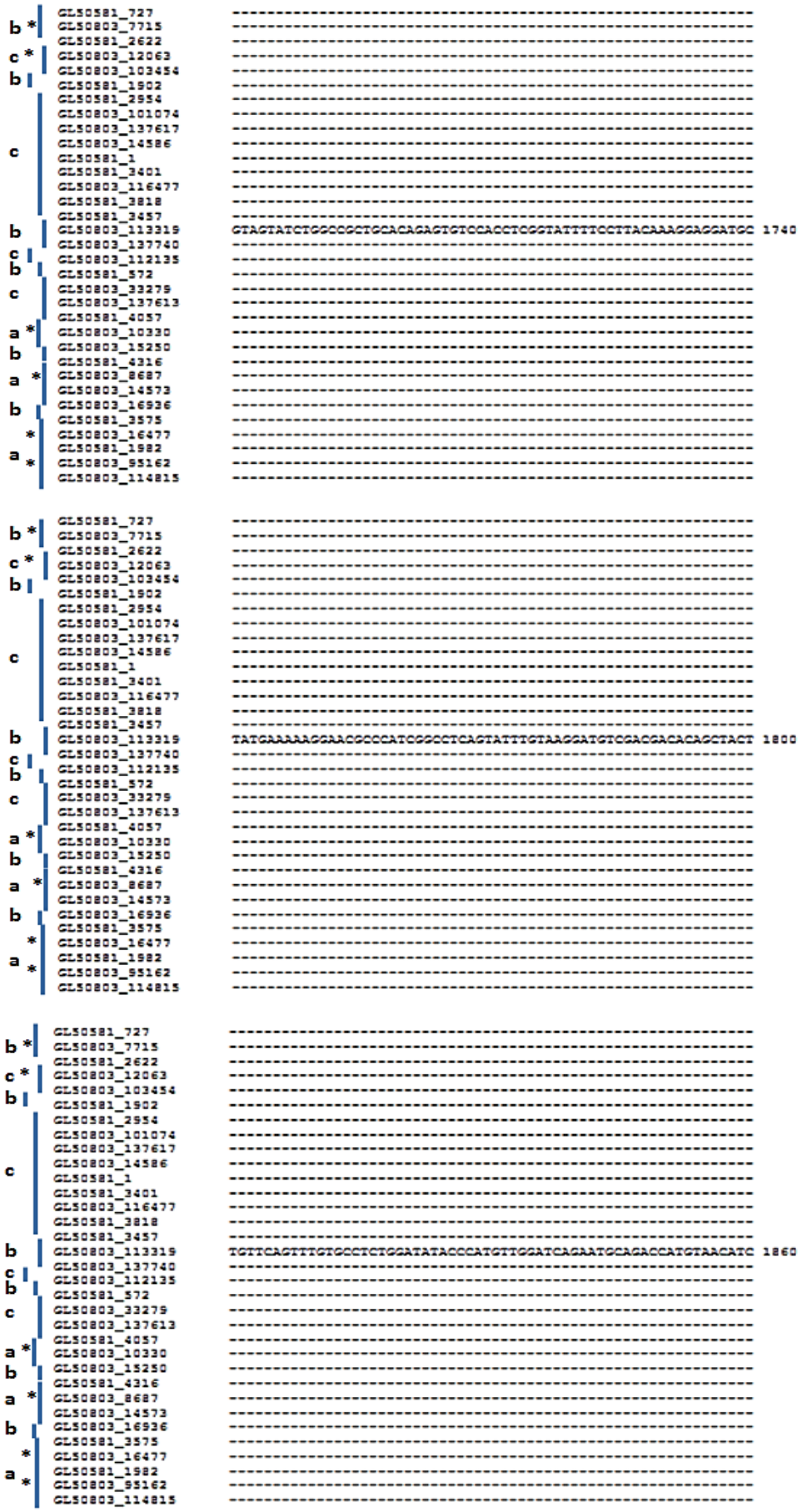

b * GL50581_727
GL50803_7715
c * GL50581_2622
GL50803_12063
b GL50803_103454
GL50581_1902
GL50581_2954
GL50803_101074
GL50803_137617
c GL50803_14586
GL50581_1
GL50581_3401
GL50803_116477
GL50581_3818
GL50581_3457
b GL50803_113319 GGAAAGGGATACATCCTTTTTCAAGGCGGATGTTATCGTGTGGATACTGAGCTAGGACTT 960
c GL50803_137740
b GL50803_112135
GL50581_572
c GL50803_33279
GL50803_137613
a * GL50581_4057
GL50803_10330
b GL50803_15250
a * GL50581_4316
GL50803_8687
GL50803_14573
b GL50803_16936
* GL50581_3575
GL50803_16477
a * GL50581_1982
GL50803_95162
GL50803_114815
b * GL50581_727
GL50803_7715
c * GL50581_2622
GL50803_12063
b GL50803_103454
GL50581_1902
GL50581_2954
GL50803_101074
GL50803_137617
c GL50803_14586
GL50581_1
GL50581_3401
GL50803_116477
GL50581_3818
GL50581_3457
b GL50803_113319 AAGGCATGTGCCTACACTACTGATATTAGGGGTGCAACCTTTTGTGCGAAATGTGGTACG 1020
c GL50803_137740
b GL50803_112135
GL50581_572
c GL50803_33279
GL50803_137613
a * GL50581_4057
GL50803_10330
b GL50803_15250
a * GL50581_4316
GL50803_8687
GL50803_14573
b GL50803_16936
* GL50581_3575
GL50803_16477
a * GL50581_1982
GL50803_95162
GL50803_114815
b * GL50581_727
GL50803_7715
c * GL50581_2622
GL50803_12063
b GL50803_103454
GL50581_1902
GL50581_2954
GL50803_101074
GL50803_137617
c GL50803_14586
GL50581_1
GL50581_3401
GL50803_116477
GL50581_3818
GL50581_3457
b GL50803_113319 CCTGGTTTCGTTCCAATCAACGGAGTATGTGTGCCATACAATAATCCGCTAGTTGTAACT 1080
c GL50803_137740
b GL50803_112135
GL50581_572
c GL50803_33279
GL50803_137613
a * GL50581_4057
GL50803_10330
b GL50803_15250
a * GL50581_4316
GL50803_8687
GL50803_14573
b GL50803_16936
* GL50581_3575
GL50803_16477
a * GL50581_1982
GL50803_95162
GL50803_114815

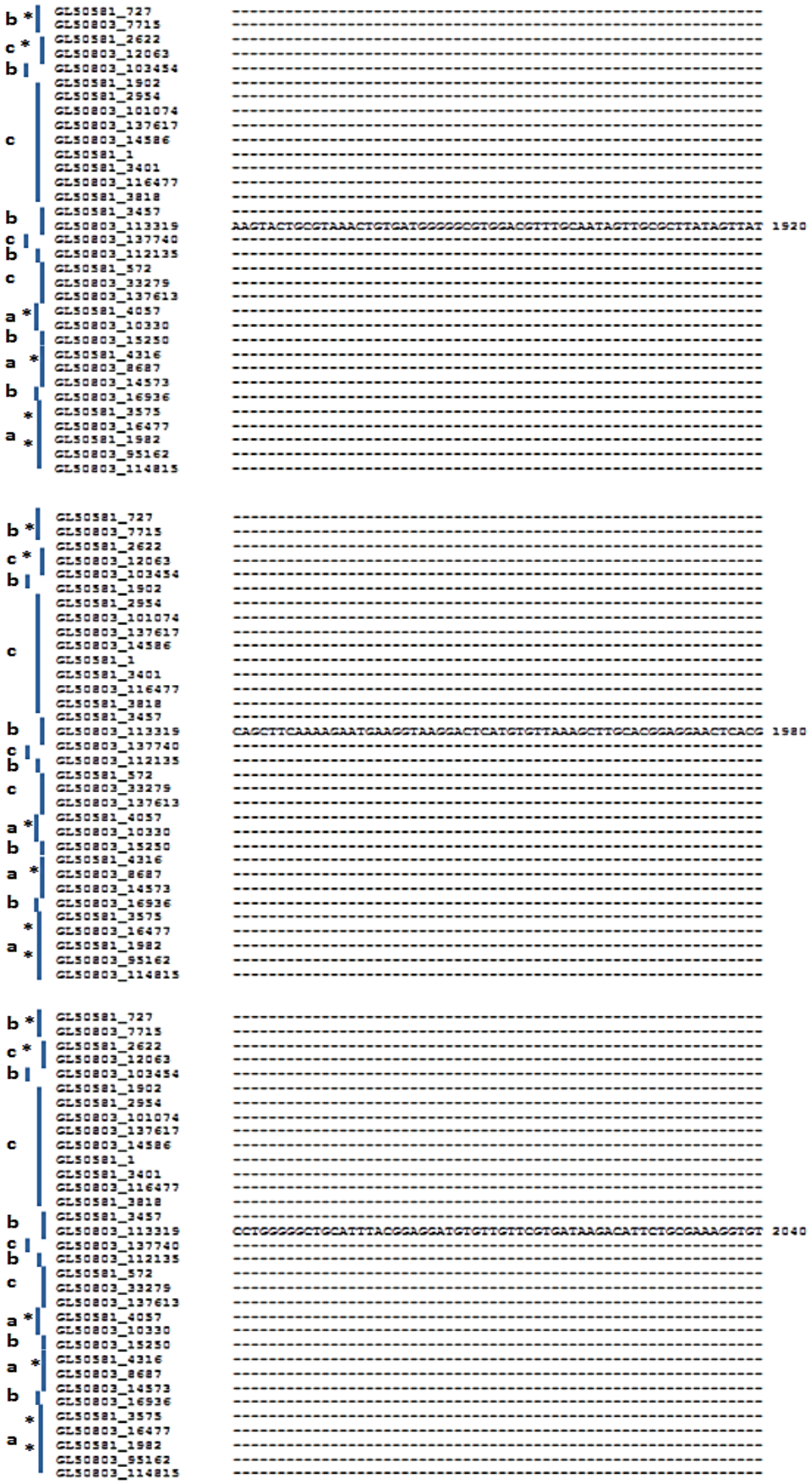

GL50581_727
GL50803_7715
GL50581_2622
GL50803_12063
GL50803_103454
GL50581_1902
GL50581_2954
GL50803_101074
GL50803_137617
GL50803_14586
GL50581_1
GL50581_3401
GL50803_116477
GL50581_3818
GL50581_3457
GL50803_113319
GL50803_137740
GL50803_112135
GL50581_572
GL50803_33279
GL50803_137613
GL50581_4057
GL50803_10330
GL50803_15250
GL50581_4316
GL50803_8687
GL50803_14573
GL50803_16936
GL50581_3575
GL50803_16477
GL50581_1982
GL50803_95162
GL50803_114815
TCTGGTTGTCGAGATTCAGCCGGCGGTCTACCAACACCTGTCTCTATTAAGTGTACCTCT 1140
TGTGGTAAGAATTACTTCATTTACCAAGGTGGATGTTATTCCCCAGAATCTGCTCCAGGA 1200
AACGTCATATGTAACGAGGTGGATAATGCTGGGAAATGCAAAGCGTGTGTAGACAACTAT 1260


b * GLS0581_727 ----------
GLS0803_7715 ----------
c * GLS0581_2622 ----------
GLS0803_12063 ----------
b GLS0803_103454 ----------
GLS0581_1902 ----------
GLS0581_2954 ----------
GLS0803_101074 ----------
GLS0803_137617 ----------
c GLS0803_14586 ----------
GLS0581_1 ----------
GLS0581_3401 ----------
GLS0803_116477 ----------
GLS0581_3818 ----------
GLS0581_3457 ----------
b GLS0803_113319 TTTCGCAACAGAGGAGATTGTATTCCATGCCATTCAAATTGTGCTTCATGCCTTGGTCCT 1320
c GLS0803_137740 ----------
b GLS0803_112135 ----------
GLS0581_572 ----------
c GLS0803_33279 ----------
GLS0803_137613 ----------
a * GLS0581_4057 ----------
GLS0803_10330 ----------
b GLS0803_15250 ----------
GLS0581_4316 ----------
a * GLS0803_8687 ----------
GLS0803_14573 ----------
b GLS0803_16936 ----------
GLS0581_3575 ----------
* GLS0803_16477 ----------
a * GLS0581_1982 ----------
GLS0803_95162 ----------
GLS0803_114815 ----------
GLS0581_727 ----------
b * GLS0803_7715 ----------
GLS0581_2622 ----------
c * GLS0803_12063 ----------
GLS0803_103454 ----------
b GLS0581_1902 ----------
GLS0581_2954 ----------
GLS0803_101074 ----------
GLS0803_137617 ----------
c GLS0803_14586 ----------
GLS0581_1 ----------
GLS0581_3401 ----------
GLS0803_116477 ----------
GLS0581_3818 ----------
GLS0581_3457 ----------
b GLS0803_113319 AACGTCATATGTAACGAGGTGGATAATGCTGGGAAATGCAAAGCGTGTGTAGACAACTAT 1260
c GLS0803_137740 ----------
b GLS0803_112135 ----------
GLS0581_572 ----------
c GLS0803_33279 ----------
GLS0803_137613 ----------
GLS0581_4057 ----------
a * GLS0803_10330 ----------
b GLS0803_15250 ----------
GLS0581_4316 ----------
a * GLS0803_8687 ----------
GLS0803_14573 ----------
b GLS0803_16936 ----------
GLS0581_3575 ----------
* GLS0803_16477 ----------
a * GLS0581_1982 ----------
GLS0803_95162 ----------
GLS0803_114815 ----------
b * GLS0581_727 ----------
GLS0803_7715 ----------
c * GLS0581_2622 ----------
GLS0803_12063 ----------
b GLS0803_103454 ----------
GLS0581_1902 ----------
GLS0581_2954 ----------
GLS0803_101074 ----------
GLS0803_137617 ----------
c GLS0803_14586 ----------
GLS0581_1 ----------
GLS0581_3401 ----------
GLS0803_116477 ----------
GLS0581_3818 ----------
GLS0581_3457 ----------
b GLS0803_113319 TTTCGCAACAGAGGAGATTGTATTCCATGCCATTCAAATTGTGCTTCATGCCTTGGTCCT 1320
c GLS0803_137740 ----------
b GLS0803_112135 ----------
GLS0581_572 ----------
c GLS0803_33279 ----------
GLS0803_137613 ----------
a * GLS0581_4057 ----------
GLS0803_10330 ----------
b GLS0803_15250 ----------
GLS0581_4316 ----------
a * GLS0803_8687 ----------
GLS0803_14573 ----------
b GLS0803_16936 ----------
GLS0581_3575 ----------
* GLS0803_16477 ----------
a * GLS0581_1982 ----------
GLS0803_95162 ----------
GLS0803_114815 ----------


GLS0581_727
GLS0803_7715
GLS0581_2622
GLS0803_12063
GLS0803_103454
GLS0581_1902
GLS0581_2954
GLS0803_101074
GLS0803_137617
GLS0803_14586
GLS0581_1
GLS0581_3401
GLS0803_116477
GLS0581_3818
GLS0581_3457
GLS0803_113319
GLS0803_137740
GLS0803_112135
GLS0581_572
GLS0803_33279
GLS0803_137613
GLS0581_4057
GLS0803_10330
GLS0803_15250
GLS0581_4316
GLS0803_8687
GLS0803_14573
GLS0803_16936
GLS0581_3575
GLS0803_16477
GLS0581_1982
GLS0803_95162
GLS0803_114815
AACGCGAATGATTGTACTTCTTGTGCTCCAGAAATATTCTTTCAAGTGTTTTCGATGGGG 138
AAAGGTATGTGCAAGTATTGTAATGATGTACTTCTTCAGTACGATAACGCTGGCATAGCT 144
CACTGTGCAAAATGCTTACCTCTTCCACGAGGCCAACGAGCGCAATGCACAGTCTGCGAA 1500

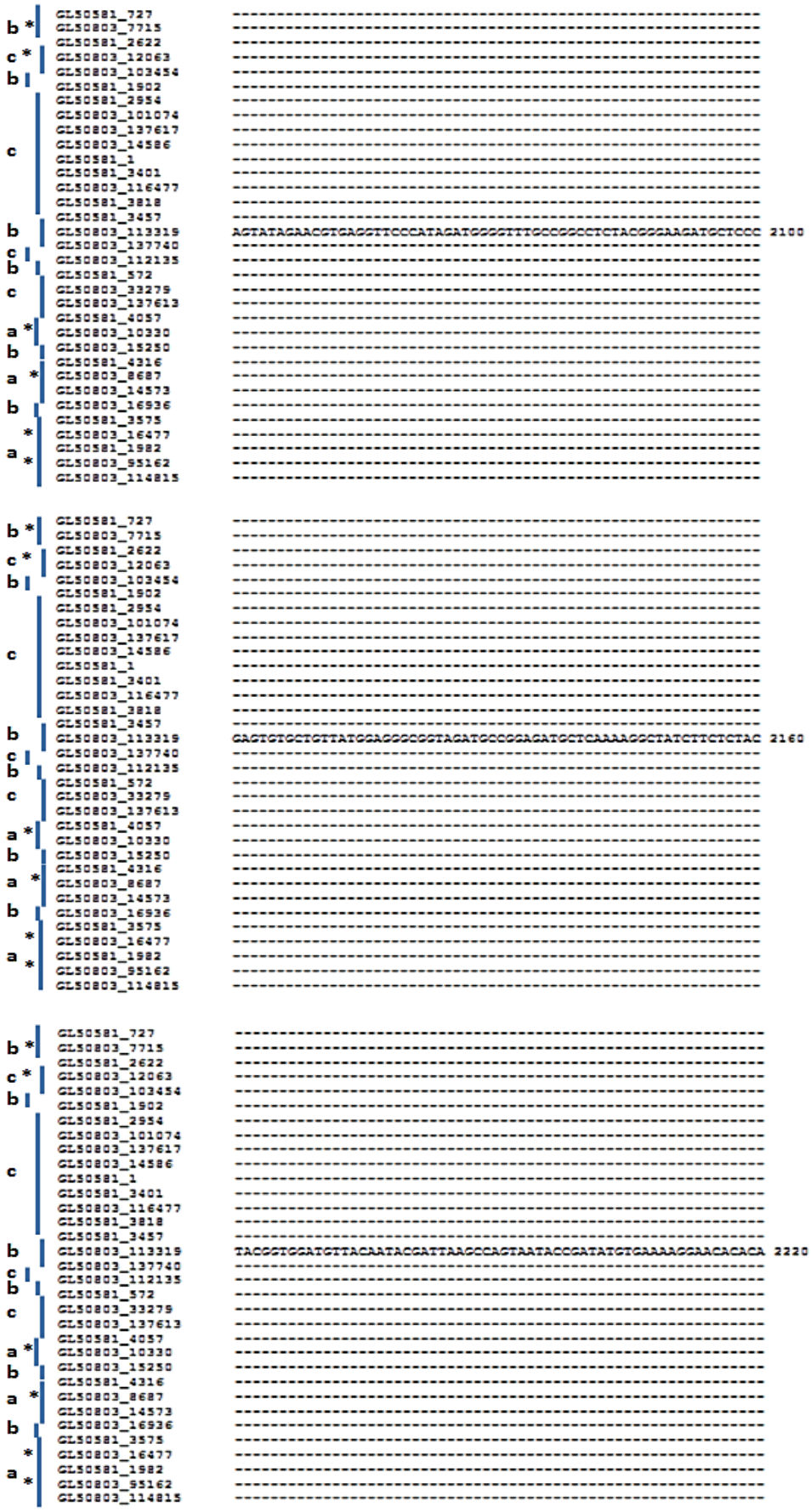

b * GL50581_727 ----------
GL50803_7715 ----------
c * GL50581_2622 ----------
GL50803_12063 ----------
b GL50803_103454 ----------
GL50581_1902 ----------
GL50581_2954 ----------
GL50803_101074 ----------
GL50803_137617 ----------
c GL50803_14586 ----------
GL50581_1 ----------
GL50581_3401 ----------
GL50803_116477 ----------
GL50581_3818 ----------
b GL50581_3457 ----------
GL50803_113319 AGGGGATATCTCCTAACAGAAAATCACTGCGAAATATCGTGCTGGACAAACGGTCTTGCT 1560
c GL50803_137740 ----------
b GL50803_112135 ----------
b GL50581_572 ----------
c GL50803_33279 ----------
GL50803_137613 ----------
a * GL50581_4057 ----------
GL50803_10330 ----------
b GL50803_15250 ----------
a * GL50581_4316 ----------
GL50803_8687 ----------
GL50803_14573 ----------
b GL50803_16936 ----------
* GL50581_3575 ----------
GL50803_16477 ----------
a * GL50581_1982 ----------
GL50803_95162 ----------
GL50803_114815 ----------
GL50581_727 ----------
b * GL50803_7715 ----------
GL50581_2622 ----------
c * GL50803_12063 ----------
b GL50803_103454 ----------
GL50581_1902 ----------
GL50581_2954 ----------
GL50803_101074 ----------
GL50803_137617 ----------
c GL50803_14586 ----------
GL50581_1 ----------
GL50581_3401 ----------
GL50803_116477 ----------
GL50581_3818 ----------
GL50581_3457 ----------
b GL50803_113319 CCGGGGTCATGTAAGTCTGATCATTGCAACATTTTCTCGTCTGCACTTTGTTCTTGTTGC 1620
c GL50803_137740 ----------
b GL50803_112135 ----------
b GL50581_572 ----------
c GL50803_33279 ----------
GL50803_137613 ----------
a * GL50581_4057 ----------
GL50803_10330 ----------
b GL50803_15250 ----------
a * GL50581_4316 ----------
GL50803_8687 ----------
GL50803_14573 ----------
b GL50803_16936 ----------
* GL50581_3575 ----------
GL50803_16477 ----------
a * GL50581_1982 ----------
GL50803_95162 ----------
GL50803_114815 ----------
b * GL50581_727 ----------
GL50803_7715 ----------
c * GL50581_2622 ----------
GL50803_12063 ----------
b GL50803_103454 ----------
GL50581_1902 ----------
GL50581_2954 ----------
GL50803_101074 ----------
GL50803_137617 ----------
c GL50803_14586 ----------
GL50581_1 ----------
GL50581_3401 ----------
GL50803_116477 ----------
GL50581_3818 ----------
b GL50581_3457 ----------
GL50803_113319 GCAAATGATAATGAGTGCCCCATAAATGGACAGTGCACATCTGCCTTAATGGGCGCCAAA 1680
c GL50803_137740 ----------
b GL50803_112135 ----------
b GL50581_572 ----------
c GL50803_33279 ----------
GL50803_137613 ----------
a * GL50581_4057 ----------
GL50803_10330 ----------
b GL50803_15250 ----------
a * GL50581_4316 ----------
GL50803_8687 ----------
GL50803_14573 ----------
b GL50803_16936 ----------
* GL50581_3575 ----------
GL50803_16477 ----------
a * GL50581_1982 ----------
GL50803_95162 ----------
GL50803_114815 ----------

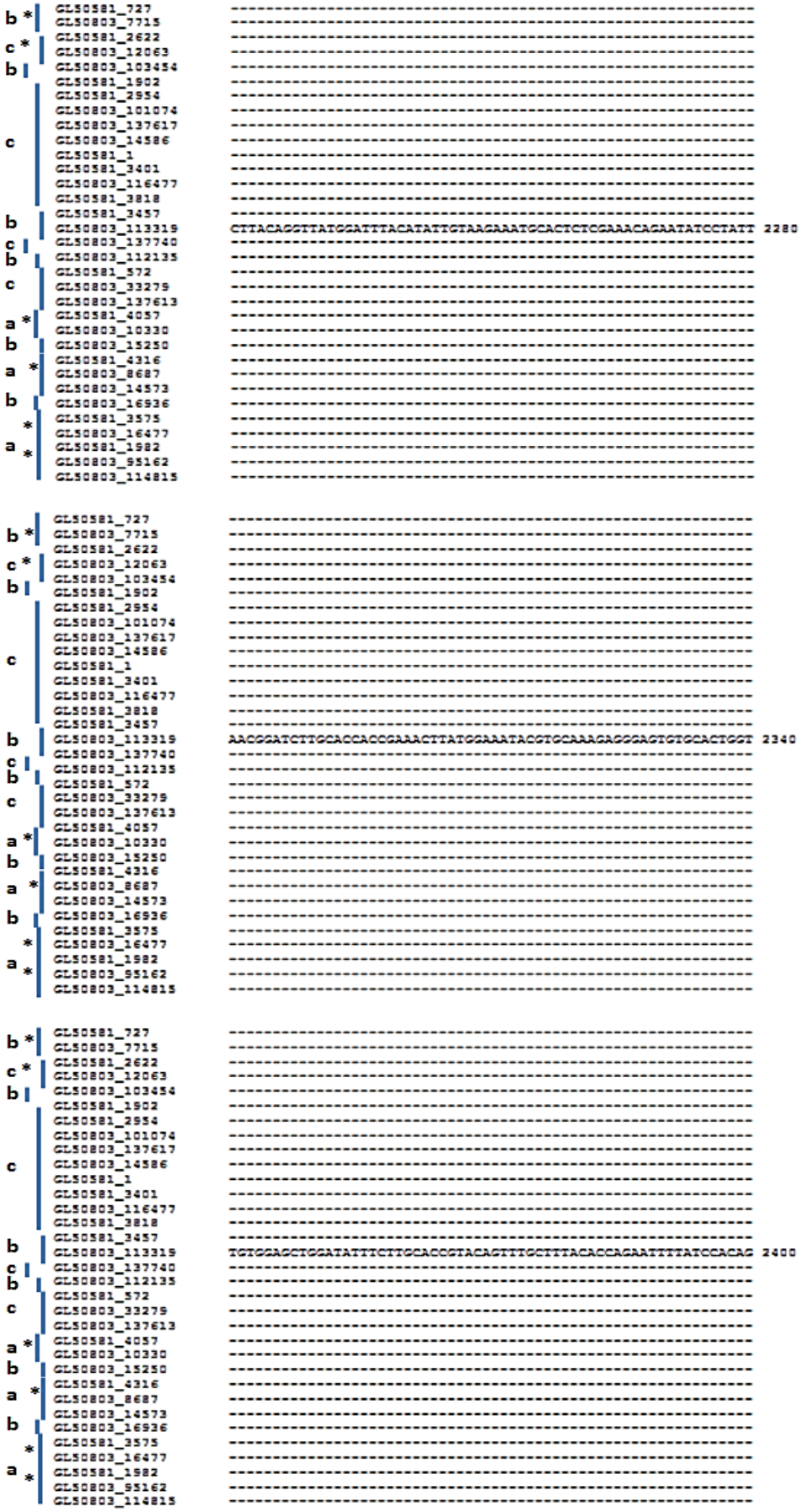

b * GLS0581_727 ---------- 1740
GLS0803_7715 ----------
GLS0581_2622 ----------
c * GLS0803_12063 ----------
GLS0803_103454 ----------
b GLS0581_1902 ----------
GLS0581_2954 ----------
GLS0803_101074 ----------
GLS0803_137617 ----------
c GLS0803_14586 ----------
GLS0581_1 ----------
GLS0581_3401 ----------
GLS0803_116477 ----------
GLS0581_3818 ----------
GLS0581_3457 ----------
b GLS0803_113319 GTAGTATCTGGCCGCTGCACAGAGTGTCCACCTCGGTATTTTCCTTACAAAGGAGGATGC 1800
c GLS0803_137740 ----------
b GLS0803_112135 ----------
GLS0581_572 ----------
c GLS0803_33279 ----------
GLS0803_137613 ----------
GLS0581_4057 ----------
a * GLS0803_10330 ----------
b GLS0803_15250 ----------
GLS0581_4316 ----------
a * GLS0803_8687 ----------
GLS0803_14573 ----------
b GLS0803_16936 ----------
GLS0581_3575 ----------
* GLS0803_16477 ----------
GLS0581_1982 ----------
a * GLS0803_95162 ----------
GLS0803_114815 ----------
b * GLS0581_727 ---------- 1800
GLS0803_7715 ----------
GLS0581_2622 ----------
c * GLS0803_12063 ----------
GLS0803_103454 ----------
b GLS0581_1902 ----------
GLS0581_2954 ----------
GLS0803_101074 ----------
GLS0803_137617 ----------
c GLS0803_14586 ----------
GLS0581_1 ----------
GLS0581_3401 ----------
GLS0803_116477 ----------
GLS0581_3818 ----------
GLS0581_3457 ----------
b GLS0803_113319 TATGAAAAAGGAACGGCCCATCGGCCTCAGTATTTGTAAGGATGTCGACGACACAGCTACT 1860
c GLS0803_137740 ----------
b GLS0803_112135 ----------
GLS0581_572 ----------
c GLS0803_33279 ----------
GLS0803_137613 ----------
GLS0581_4057 ----------
a * GLS0803_10330 ----------
b GLS0803_15250 ----------
GLS0581_4316 ----------
a * GLS0803_8687 ----------
GLS0803_14573 ----------
b GLS0803_16936 ----------
GLS0581_3575 ----------
* GLS0803_16477 ----------
a * GLS0581_1982 ----------
GLS0803_95162 ----------
GLS0803_114815 ----------
b * GLS0581_727 ---------- 1860
GLS0803_7715 ----------
GLS0581_2622 ----------
c * GLS0803_12063 ----------
GLS0803_103454 ----------
b GLS0581_1902 ----------
GLS0581_2954 ----------
GLS0803_101074 ----------
GLS0803_137617 ----------
c GLS0803_14586 ----------
GLS0581_1 ----------
GLS0581_3401 ----------
GLS0803_116477 ----------
GLS0581_3818 ----------
GLS0581_3457 ----------
b GLS0803_113319 TGTTCAGTTTGTGCCTCTGGATATACCCATGTTGGATCAGAATGCAGACCATGTAACATC 1920
c GLS0803_137740 ----------
b GLS0803_112135 ----------
GLS0581_572 ----------
c GLS0803_33279 ----------
GLS0803_137613 ----------
GLS0581_4057 ----------
a * GLS0803_10330 ----------
b GLS0803_15250 ----------
GLS0581_4316 ----------
a * GLS0803_8687 ----------
GLS0803_14573 ----------
b GLS0803_16936 ----------
GLS0581_3575 ----------
* GLS0803_16477 ----------
a * GLS0581_1982 ----------
GLS0803_95162 ----------
GLS0803_114815 ----------

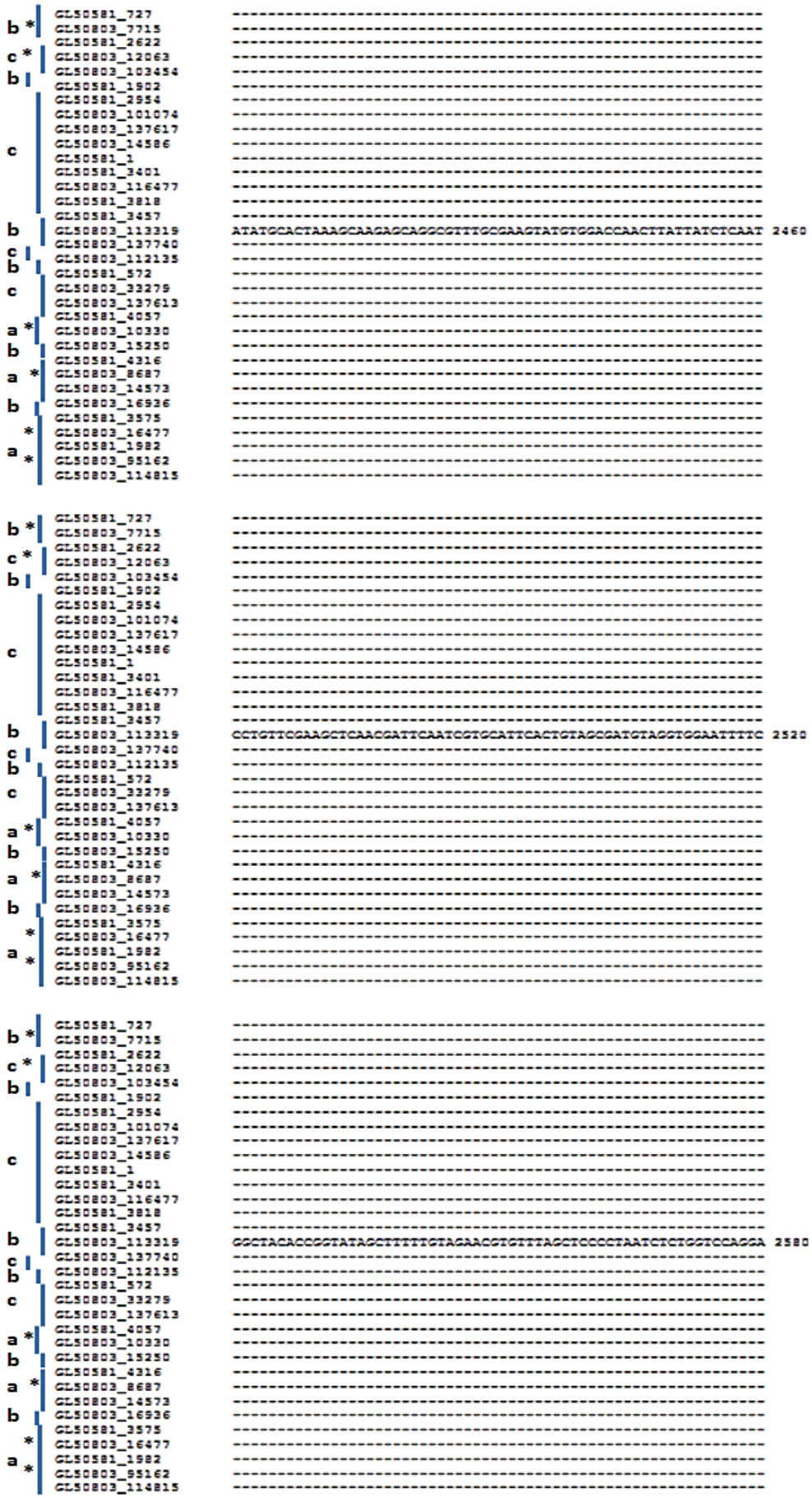

b * GL50581_727 ----------
GL50803_7715 ----------
c * GL50581_2622 ----------
GL50803_12063 ----------
b GL50803_103454 ----------
GL50581_1902 ----------
GL50581_2954 ----------
GL50803_101074 ----------
GL50803_137617 ----------
c GL50803_14586 ----------
GL50581_1 ----------
GL50581_3401 ----------
GL50803_116477 ----------
GL50581_3818 ----------
b GL50581_3457 ----------
GL50803_113319 AAGTACTGCGTAAACTGTGATGGGGGCGTGGACGTTTGCAATAGTTGCGCTTATAGTTAT 1920
c GL50803_137740 ----------
b GL50803_112135 ----------
c GL50581_572 ----------
GL50803_33279 ----------
GL50803_137613 ----------
a * GL50581_4057 ----------
GL50803_10330 ----------
b GL50803_15250 ----------
a * GL50581_4316 ----------
GL50803_8687 ----------
b GL50803_14573 ----------
GL50803_16936 ----------
* GL50581_3575 ----------
GL50803_16477 ----------
a * GL50581_1982 ----------
GL50803_95162 ----------
GL50803_114815 ----------

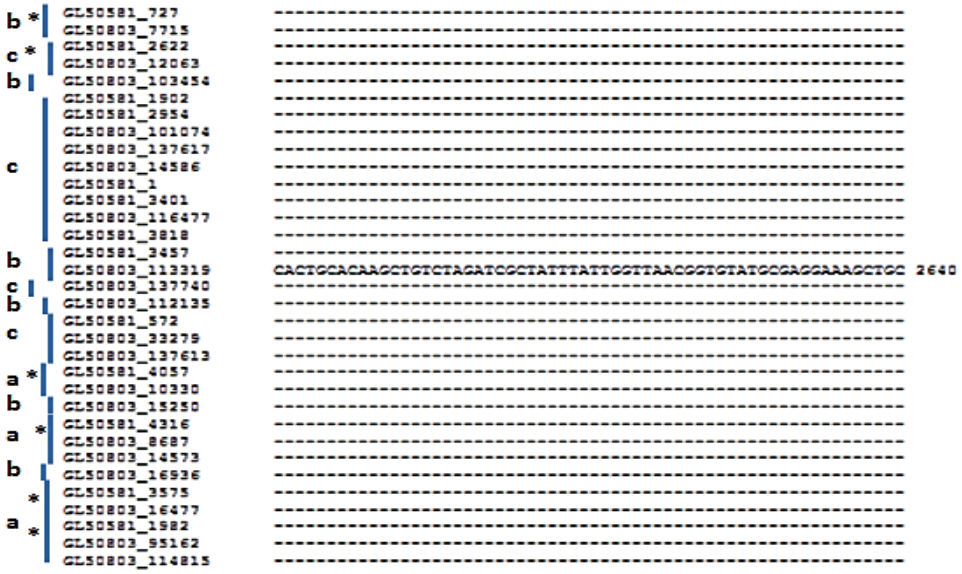
GL50581_727 ----------
b * GL50803_7715 ----------
GL50581_2622 ----------
c * GL50803_12063 ----------
b GL50803_103454 ----------
GL50581_1902 ----------
GL50581_2954 ----------
GL50803_101074 ----------
GL50803_137617 ----------
c GL50803_14586 ----------
GL50581_1 ----------
GL50581_3401 ----------
GL50803_116477 ----------
GL50581_3818 ----------
GL50581_3457 ----------
b GL50803_113319 CAGCTTCAAAAGAATGAAGGTAAGGACTCATGTGTTAAAGCTTGCACGGAGGAACTCACG 1980
c GL50803_137740 ----------
b GL50803_112135 ----------
c GL50581_572 ----------
GL50803_33279 ----------
GL50803_137613 ----------
a * GL50581_4057 ----------
GL50803_10330 ----------
b GL50803_15250 ----------
a * GL50581_4316 ----------
GL50803_8687 ----------
b GL50803_14573 ----------
GL50803_16936 ----------
* GL50581_3575 ----------
GL50803_16477 ----------
a * GL50581_1982 ----------
GL50803_95162 ----------
GL50803_114815 ----------

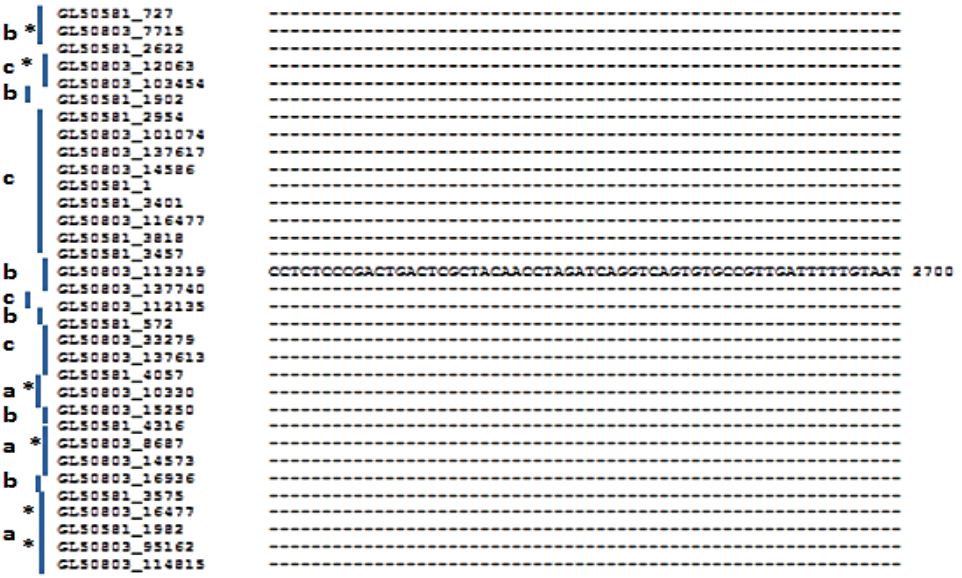
b * GL50581_727 ----------
GL50803_7715 ----------
c * GL50581_2622 ----------
GL50803_12063 ----------
b GL50803_103454 ----------
GL50581_1902 ----------
GL50581_2954 ----------
GL50803_101074 ----------
GL50803_137617 ----------
c GL50803_14586 ----------
GL50581_1 ----------
GL50581_3401 ----------
GL50803_116477 ----------
GL50581_3818 ----------
b GL50581_3457 ----------
GL50803_113319 CCTGGGGGCTGCATTTACGGAGGATGTGTTGTTCGTGATAAGACATTCTGCGAAAGGTGT 2040
c GL50803_137740 ----------
b GL50803_112135 ----------
c GL50581_572 ----------
GL50803_33279 ----------
GL50803_137613 ----------
a * GL50581_4057 ----------
GL50803_10330 ----------
b GL50803_15250 ----------
a * GL50581_4316 ----------
GL50803_8687 ----------
b GL50803_14573 ----------
GL50803_16936 ----------
* GL50581_3575 ----------
GL50803_16477 ----------
a * GL50581_1982 ----------
GL50803_95162 ----------
GL50803_114815 ----------

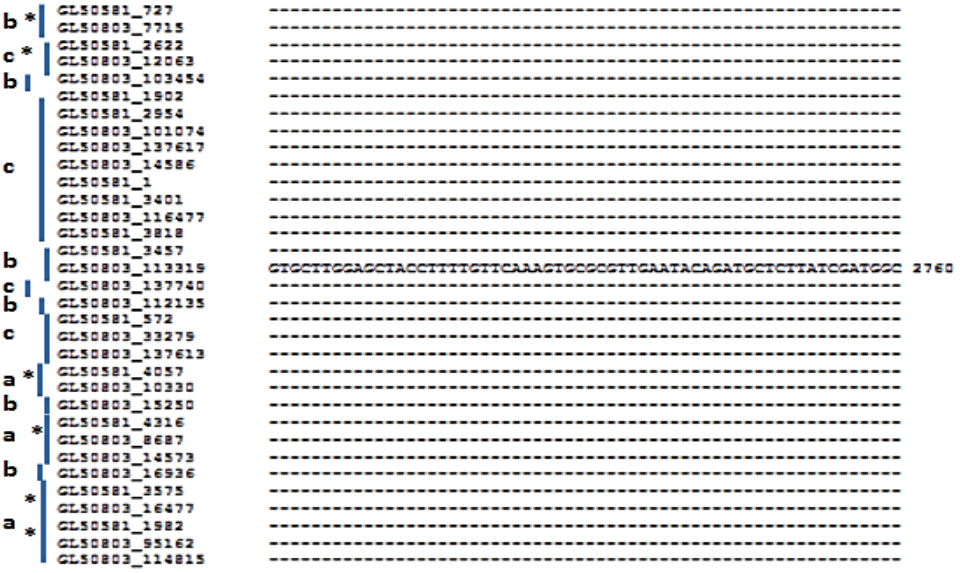

b * GL50581_727
GL50803_7715
c * GL50581_2622
GL50803_12063
b GL50803_103454
GL50581_1902
GL50581_2954
GL50803_101074
GL50803_137617
c GL50803_14586
GL50581_1
GL50581_3401
GL50803_116477
GL50581_3818
GL50581_3457
b GL50803_113319 AGTATAGAACGTGAGGTTCCCATAGATGGGGTTTGCCGGCCTCTACGGGAAGATGCTCCC 2100
c GL50803_137740
b GL50803_112135
b GL50581_572
c GL50803_33279
GL50803_137613
GL50581_4057
a * GL50803_10330
b GL50803_15250
GL50581_4316
a * GL50803_8687
GL50803_14573
b GL50803_16936
GL50581_3575
* GL50803_16477
a GL50581_1982
* GL50803_95162
GL50803_114815
b * GL50581_727
GL50803_7715
c * GL50581_2622
GL50803_12063
b GL50803_103454
GL50581_1902
GL50581_2954
GL50803_101074
GL50803_137617
c GL50803_14586
GL50581_1
GL50581_3401
GL50803_116477
GL50581_3818
GL50581_3457
b GL50803_113319 GAGTGTGCTGTTATGGAGGGCGGTAGATGCCGGAGATGCTCAAAAGGCTATCTTCTCTAC 2160
c GL50803_137740
b GL50803_112135
b GL50581_572
c GL50803_33279
GL50803_137613
GL50581_4057
a * GL50803_10330
b GL50803_15250
GL50581_4316
a * GL50803_8687
GL50803_14573
b GL50803_16936
GL50581_3575
* GL50803_16477
a GL50581_1982
* GL50803_95162
GL50803_114815
b * GL50581_727
GL50803_7715
c * GL50581_2622
GL50803_12063
b GL50803_103454
GL50581_1902
GL50581_2954
GL50803_101074
GL50803_137617
c GL50803_14586
GL50581_1
GL50581_3401
GL50803_116477
GL50581_3818
GL50581_3457
b GL50803_113319 TACGGTGGATGTTACAATACGATTAAGCCAGTAATACCGATATGTGAAAAGGAACACACA 2220
c GL50803_137740
b GL50803_112135
b GL50581_572
c GL50803_33279
GL50803_137613
GL50581_4057
a * GL50803_10330
b GL50803_15250
GL50581_4316
a * GL50803_8687
GL50803_14573
b GL50803_16936
GL50581_3575
* GL50803_16477
a GL50581_1982
* GL50803_95162
GL50803_114815

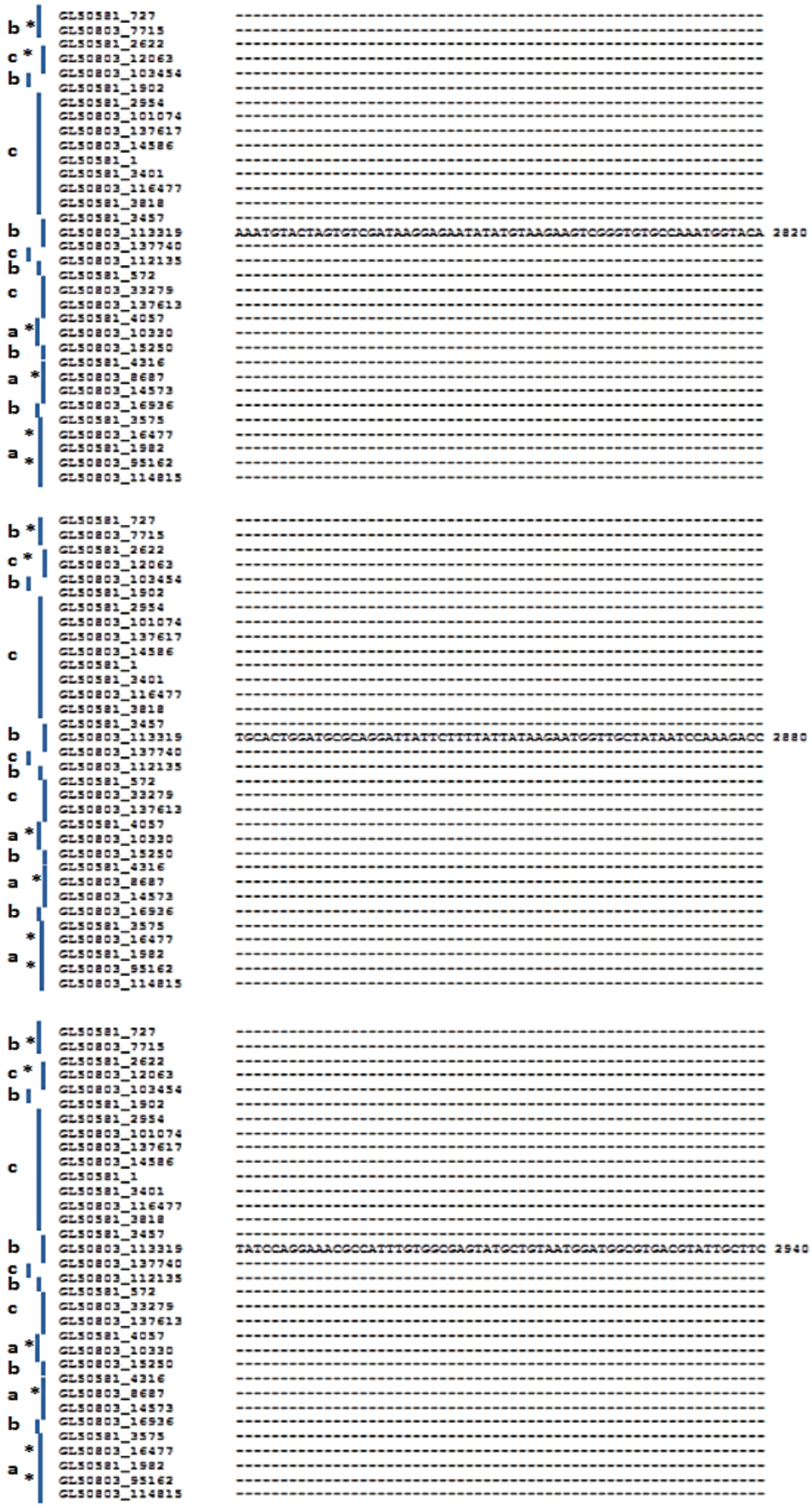

b * GLS0581_727
GLS0803_7715
c * GLS0581_2622
GLS0803_12063
b GLS0803_103454
GLS0581_1902
GLS0581_2954
GLS0803_101074
GLS0803_137617
c GLS0803_14586
GLS0581_1
GLS0581_3401
GLS0803_116477
GLS0581_3818
b GLS0581_3457
GLS0803_113319 CTTACAGGTTATGGATTTACATATTGTAAGAAATGCACTCTCGAAACAGAATATCCTATT 2280
c GLS0803_137740
b GLS0803_112135
c GLS0581_572
GLS0803_33279
GLS0803_137613
a * GLS0581_4057
GLS0803_10330
b GLS0803_15250
a * GLS0581_4316
GLS0803_8687
b GLS0803_14573
GLS0803_16936
* GLS0581_3575
a * GLS0803_16477
GLS0581_1982
GLS0803_95162
GLS0803_114815
b * GLS0581_727
GLS0803_7715
c * GLS0581_2622
GLS0803_12063
b GLS0803_103454
GLS0581_1902
GLS0581_2954
GLS0803_101074
GLS0803_137617
c GLS0803_14586
GLS0581_1
GLS0581_3401
GLS0803_116477
GLS0581_3818
b GLS0581_3457
GLS0803_113319 AACGGATCTTGCACCACCGAAACTTATGGAAATACGTGCAAAGAGGGAGTGTGCACTGGT 2340
c GLS0803_137740
b GLS0803_112135
c GLS0581_572
GLS0803_33279
GLS0803_137613
a * GLS0581_4057
GLS0803_10330
b GLS0803_15250
a * GLS0581_4316
GLS0803_8687
b GLS0803_14573
GLS0803_16936
* GLS0581_3575
a * GLS0803_16477
GLS0581_1982
GLS0803_95162
GLS0803_114815
b * GLS0581_727
GLS0803_7715
c * GLS0581_2622
GLS0803_12063
b GLS0803_103454
GLS0581_1902
GLS0581_2954
GLS0803_101074
GLS0803_137617
c GLS0803_14586
GLS0581_1
GLS0581_3401
GLS0803_116477
GLS0581_3818
b GLS0581_3457
GLS0803_113319 TGTGGAGCTGGATATTTCTTGCACCGTACAGTTTGCTTTACACCAGAATTTTATCCACAG 2400
c GLS0803_137740
b GLS0803_112135
c GLS0581_572
GLS0803_33279
GLS0803_137613
a * GLS0581_4057
GLS0803_10330
b GLS0803_15250
a * GLS0581_4316
GLS0803_8687
b GLS0803_14573
GLS0803_16936
* GLS0581_3575
a * GLS0803_16477
GLS0581_1982
GLS0803_95162
GLS0803_114815

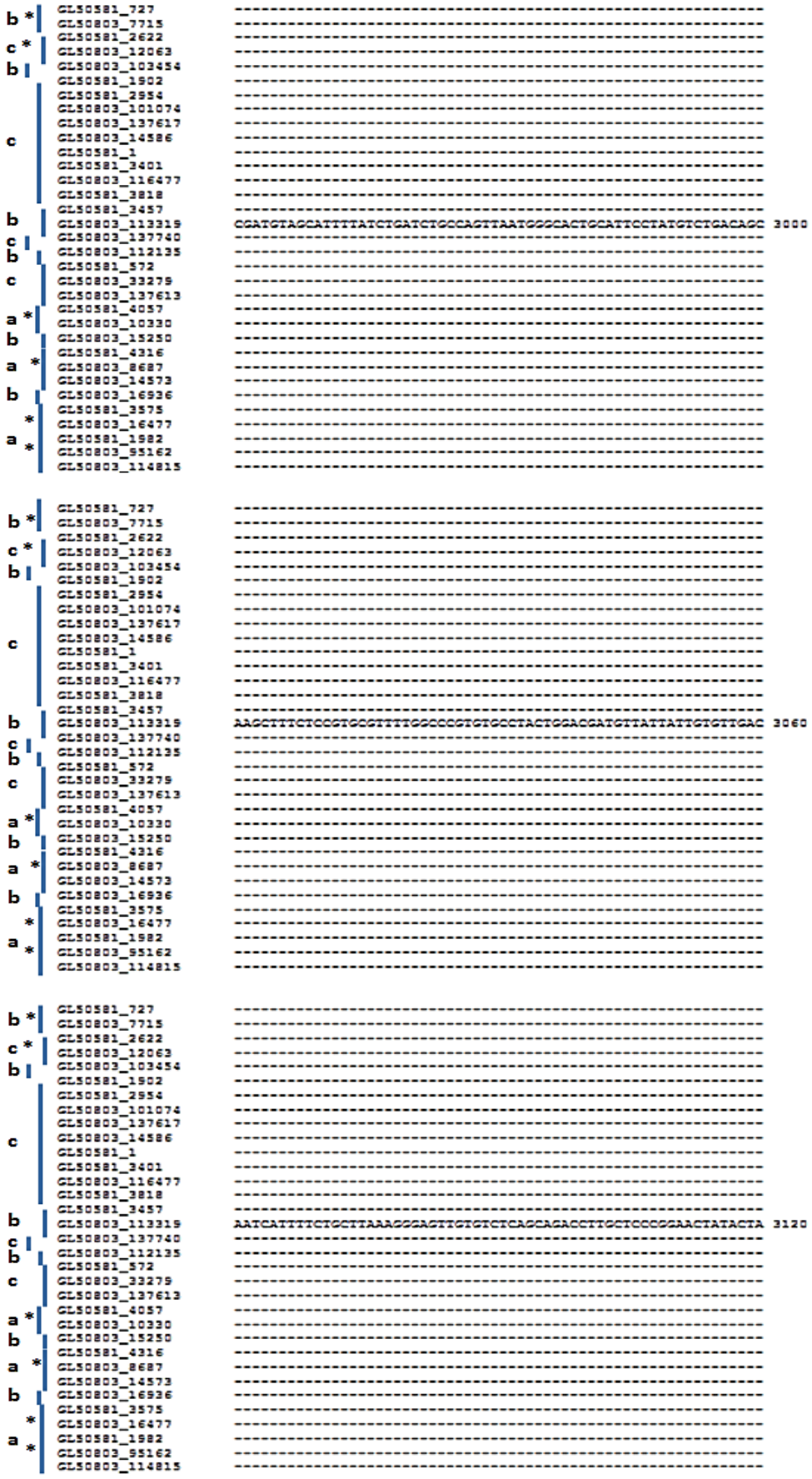

GL50581_727
GL50803_7715
GL50581_2622
GL50803_12063
GL50803_103454
GL50581_1902
GL50581_2954
GL50803_101074
GL50803_137617
GL50803_14586
GL50581_1
GL50581_3401
GL50803_116477
GL50581_3818
GL50581_3457
GL50803_113319
GL50803_137740
GL50803_112135
GL50581_572
GL50803_33279
GL50803_137613
GL50581_4057
GL50803_10330
GL50803_15250
GL50581_4316
GL50803_8687
GL50803_14573
GL50803_16936
GL50581_3575
GL50803_16477
GL50581_1982
GL50803_95162
GL50803_114815
ATATGCACTAAAGCAAGAGCAGGCGTTTGCGAAGTATGTGGACCAACTTATTATCTCAAT 2460
CCTGTTCGAAGCTCAACGATTCAATCGTGCATTCACTGTAGCGATGTAGGTGGAATTTTC 2520
GGCTACACCGGTATAGCTTTTTGTAGAACGTGTTTAGCTCCCCTAATCTCTGGTCCAGGA 2580

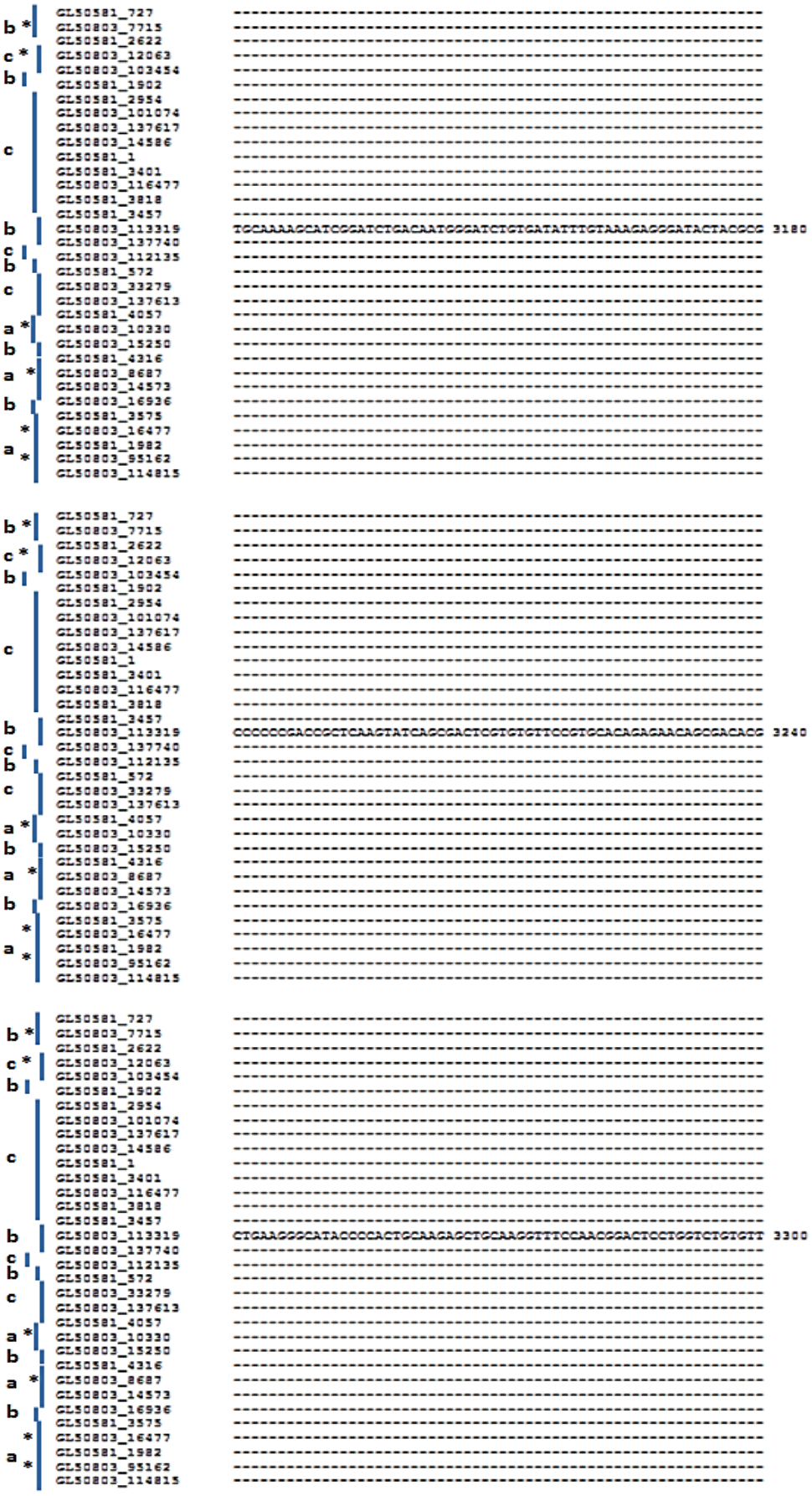

b * GL50581_727 ---------- 2640
GL50803_7715 ----------
c * GL50581_2622 ----------
GL50803_12063 ----------
b GL50803_103454 ----------
GL50581_1902 ----------
GL50581_2954 ----------
GL50803_101074 ----------
GL50803_137617 ----------
c GL50803_14586 ----------
GL50581_1 ----------
GL50581_3401 ----------
GL50803_116477 ----------
GL50581_3818 ----------
b GL50581_3457 ----------
GL50803_113319 CACTGCACAAGCTGTCTAGATCGCTATTTATTGGTTAACGGTGTATGCGAGGAAAGCTGC
c GL50803_137740 ----------
b GL50803_112135 ----------
c GL50581_572 ----------
GL50803_33279 ----------
GL50803_137613 ----------
a * GL50581_4057 ----------
GL50803_10330 ----------
b GL50803_15250 ----------
a * GL50581_4316 ----------
GL50803_8687 ----------
b GL50803_14573 ----------
GL50803_16936 ----------
* GL50581_3575 ----------
GL50803_16477 ----------
a * GL50581_1982 ----------
GL50803_95162 ----------
GL50803_114815 ----------

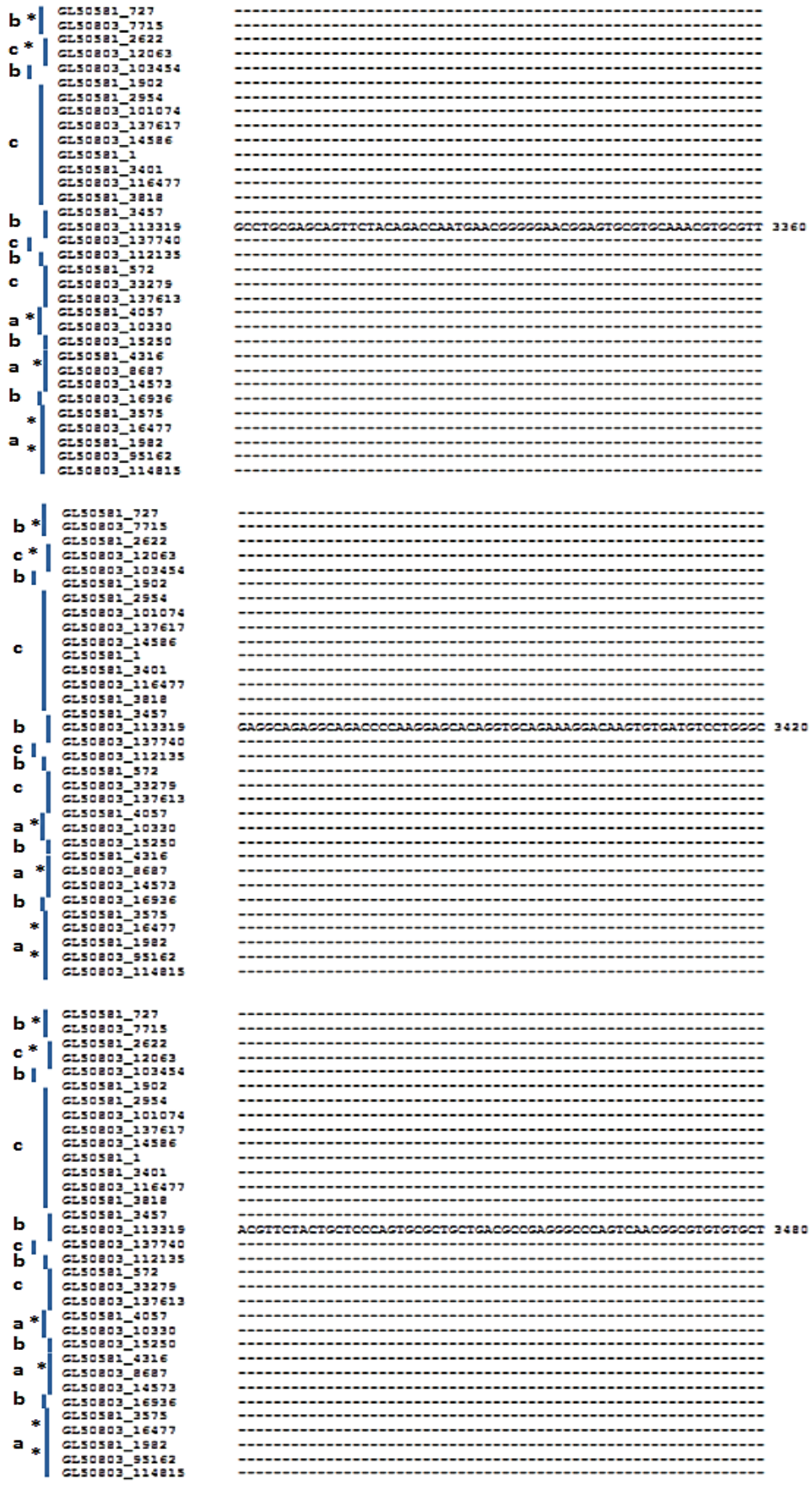

b * GL50581_727 ---------- 2700
GL50803_7715 ----------
c * GL50581_2622 ----------
GL50803_12063 ----------
b GL50803_103454 ----------
GL50581_1902 ----------
GL50581_2954 ----------
GL50803_101074 ----------
GL50803_137617 ----------
c GL50803_14586 ----------
GL50581_1 ----------
GL50581_3401 ----------
GL50803_116477 ----------
GL50581_3818 ----------
b GL50581_3457 ----------
GL50803_113319 CCTCTCCCGACTGACTCGCTACAACCTAGATCAGGTCAGTGTGCCGTTGATTTTTGTAAT
c GL50803_137740 ----------
b GL50803_112135 ----------
c GL50581_572 ----------
GL50803_33279 ----------
GL50803_137613 ----------
a * GL50581_4057 ----------
GL50803_10330 ----------
b GL50803_15250 ----------
a * GL50581_4316 ----------
GL50803_8687 ----------
b GL50803_14573 ----------
GL50803_16936 ----------
* GL50581_3575 ----------
GL50803_16477 ----------
a * GL50581_1982 ----------
GL50803_95162 ----------
GL50803_114815 ----------

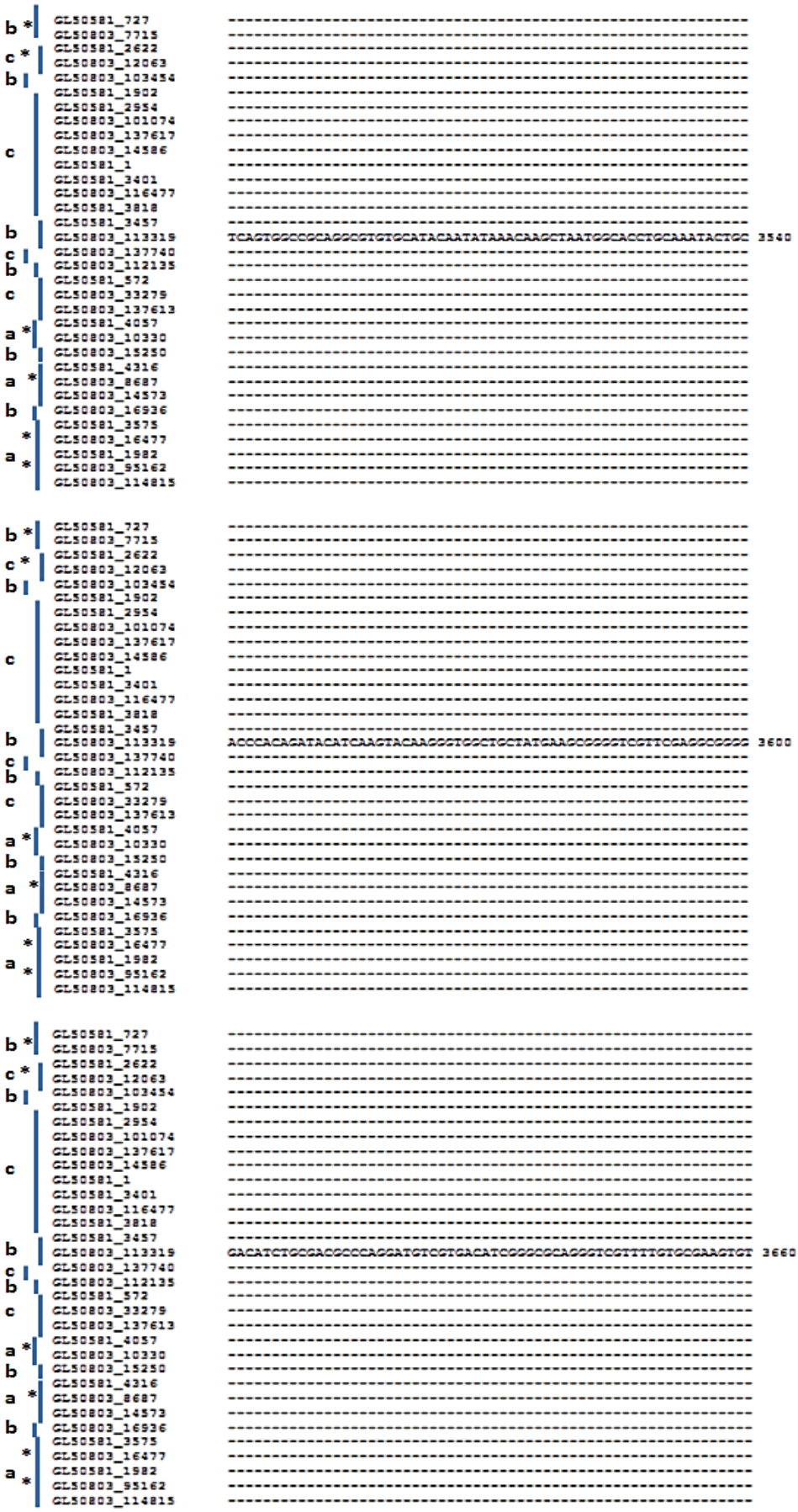

b * GL50581_727 ---------- 2760
GL50803_7715 ----------
c * GL50581_2622 ----------
GL50803_12063 ----------
b GL50803_103454 ----------
GL50581_1902 ----------
GL50581_2954 ----------
GL50803_101074 ----------
GL50803_137617 ----------
c GL50803_14586 ----------
GL50581_1 ----------
GL50581_3401 ----------
GL50803_116477 ----------
GL50581_3818 ----------
b GL50581_3457 ----------
GL50803_113319 GTGCTTGGAGCTACCTTTTGTTCAAAGTGCGCGTTGAATACAGATGCTCTTATCGATGGC
c GL50803_137740 ----------
b GL50803_112135 ----------
c GL50581_572 ----------
GL50803_33279 ----------
GL50803_137613 ----------
a * GL50581_4057 ----------
GL50803_10330 ----------
b GL50803_15250 ----------
a * GL50581_4316 ----------
GL50803_8687 ----------
b GL50803_14573 ----------
GL50803_16936 ----------
* GL50581_3575 ----------
GL50803_16477 ----------
a * GL50581_1982 ----------
GL50803_95162 ----------
GL50803_114815 ----------

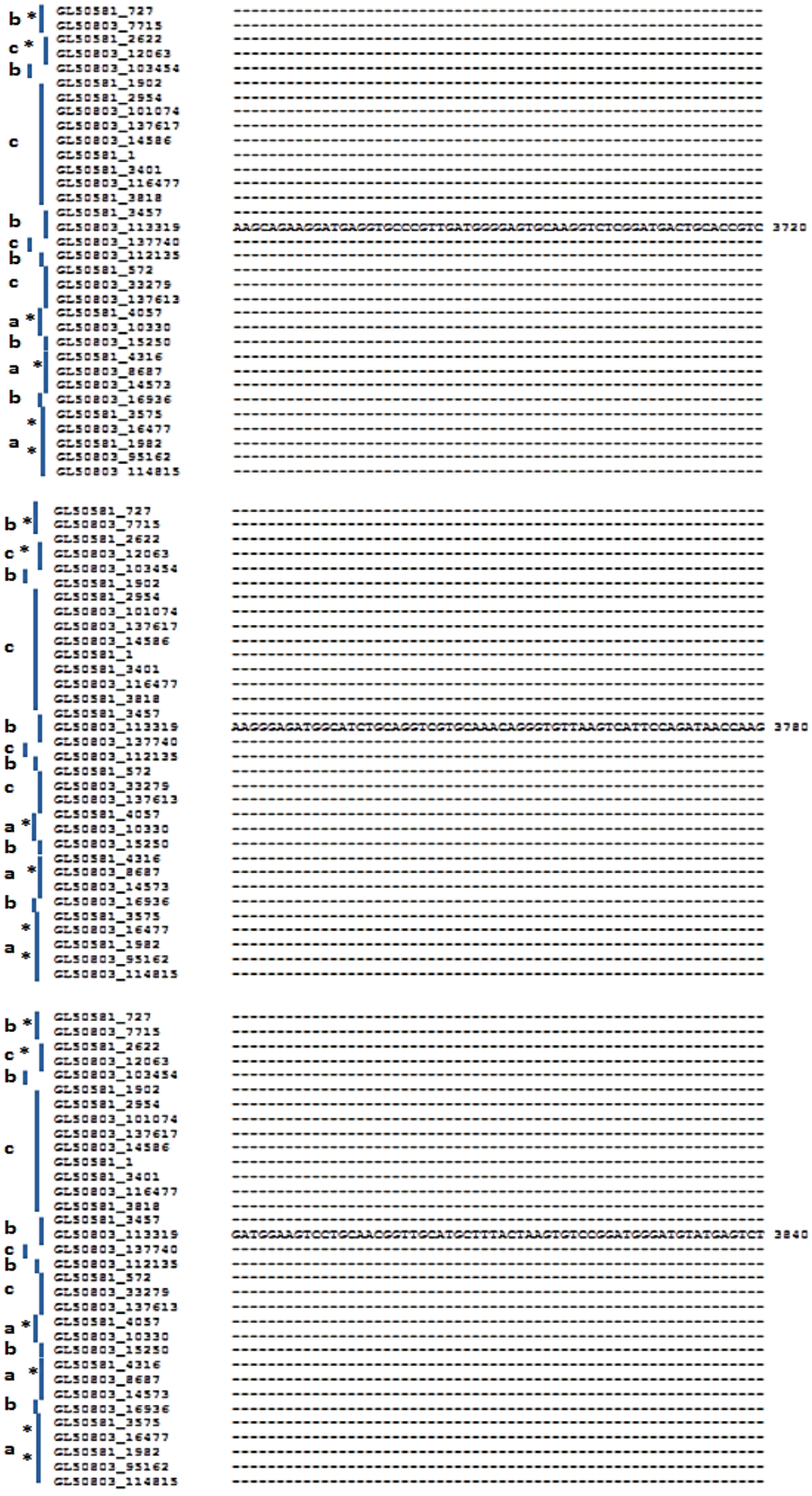

GLS0581_727
GLS0803_7715
GLS0581_2622
GLS0803_12063
GLS0803_103454
GLS0581_1902
GLS0581_2954
GLS0803_101074
GLS0803_137617
GLS0803_14586
GLS0581_1
GLS0581_3401
GLS0803_116477
GLS0581_3818
GLS0581_3457
GLS0803_113319 AAATGTACTAGTGTCGATAAGGAGAATATATGTAAGAAGTCGGGTGTGCCAAATGGTACA 2820
GLS0803_137740
GLS0803_112135
GLS0581_572
GLS0803_33279
GLS0803_137613
GLS0581_4057
GLS0803_10330
GLS0803_15250
GLS0581_4316
GLS0803_8687
GLS0803_14573
GLS0803_16936
GLS0581_3575
GLS0803_16477
GLS0581_1982
GLS0803_95162
GLS0803_114815
GLS0581_727
GLS0803_7715
GLS0581_2622
GLS0803_12063
GLS0803_103454
GLS0581_1902
GLS0581_2954
GLS0803_101074
GLS0803_137617
GLS0803_14586
GLS0581_1
GLS0581_3401
GLS0803_116477
GLS0581_3818
GLS0581_3457
GLS0803_113319 TGCACTGGATGCGCAGGATTATTCTTTTATTATAAGAATGGTTGCTATAATCCAAAGACC 2880
GLS0803_137740
GLS0803_112135
GLS0581_572
GLS0803_33279
GLS0803_137613
GLS0581_4057
GLS0803_10330
GLS0803_15250
GLS0581_4316
GLS0803_8687
GLS0803_14573
GLS0803_16936
GLS0581_3575
GLS0803_16477
GLS0581_1982
GLS0803_95162
GLS0803_114815
GLS0581_727
GLS0803_7715
GLS0581_2622
GLS0803_12063
GLS0803_103454
GLS0581_1902
GLS0581_2954
GLS0803_101074
GLS0803_137617
GLS0803_14586
GLS0581_1
GLS0581_3401
GLS0803_116477
GLS0581_3818
GLS0581_3457
GLS0803_113319 TATCCAGGAAACGCCATTTGTGGCGAGTATGCTGTAATGGATGGCGTGACGTATTGCTTC 2940
GLS0803_137740
GLS0803_112135
GLS0581_572
GLS0803_33279
GLS0803_137613
GLS0581_4057
GLS0803_10330
GLS0803_15250
GLS0581_4316
GLS0803_8687
GLS0803_14573
GLS0803_16936
GLS0581_3575
GLS0803_16477
GLS0581_1982
GLS0803_95162
GLS0803_114815

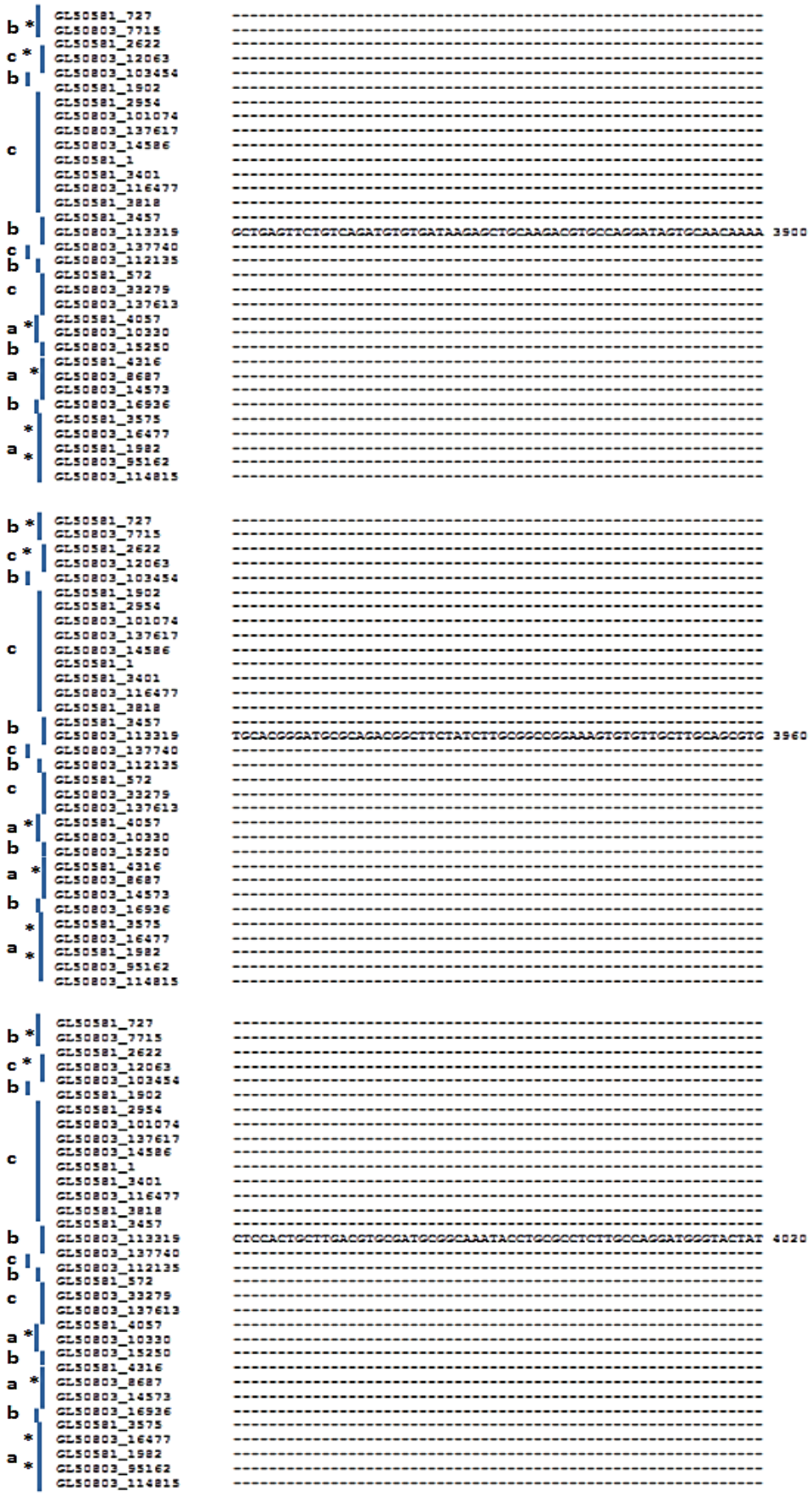

b *
GLS0581_727
GLS0803_7715
c *
GLS0581_2622
GLS0803_12063
b
GLS0803_103454
GLS0581_1902
GLS0581_2954
GLS0803_101074
GLS0803_137617
c
GLS0803_14586
GLS0581_1
GLS0581_3401
GLS0803_116477
GLS0581_3818
GLS0581_3457
b
GLS0803_113319
CGATGTAGCATTTTATCTGATCTGCCAGTTAATGGGCACTGCATTCCTATGTCTGACAGC 3000
c
GLS0803_137740
b
GLS0803_112135
GLS0581_572
c
GLS0803_33279
GLS0803_137613
GLS0581_4057
a *
GLS0803_10330
b
GLS0803_15250
GLS0581_4316
a *
GLS0803_8687
GLS0803_14573
b
GLS0803_16936
GLS0581_3575
*
GLS0803_16477
a
GLS0581_1982
*
GLS0803_95162
GLS0803_114815
b *
GLS0581_727
GLS0803_7715
c *
GLS0581_2622
GLS0803_12063
b
GLS0803_103454
GLS0581_1902
GLS0581_2954
GLS0803_101074
GLS0803_137617
c
GLS0803_14586
GLS0581_1
GLS0581_3401
GLS0803_116477
GLS0581_3818
GLS0581_3457
b
GLS0803_113319
AAGCTTTCTCCGTGCGTTTGGCCCGTGTGCCTACTGGACGATGTTATTATTGTGTGAC 3060
c
GLS0803_137740
b
GLS0803_112135
GLS0581_572
c
GLS0803_33279
GLS0803_137613
GLS0581_4057
a *
GLS0803_10330
b
GLS0803_15250
GLS0581_4316
a *
GLS0803_8687
GLS0803_14573
b
GLS0803_16936
GLS0581_3575
*
GLS0803_16477
a
GLS0581_1982
*
GLS0803_95162
GLS0803_114815
b *
GLS0581_727
GLS0803_7715
c *
GLS0581_2622
GLS0803_12063
b
GLS0803_103454
GLS0581_1902
GLS0581_2954
GLS0803_101074
GLS0803_137617
c
GLS0803_14586
GLS0581_1
GLS0581_3401
GLS0803_116477
GLS0581_3818
GLS0581_3457
b
GLS0803_113319
AATCATTTTCTGCTTAAAGGGAGTTGTGTCTCAGCAGACCTTGCTCCCGGAACTATACTA 3120
c
GLS0803_137740
b
GLS0803_112135
GLS0581_572
c
GLS0803_33279
GLS0803_137613
GLS0581_4057
a *
GLS0803_10330
b
GLS0803_15250
GLS0581_4316
a *
GLS0803_8687
GLS0803_14573
b
GLS0803_16936
GLS0581_3575
*
GLS0803_16477
a
GLS0581_1982
*
GLS0803_95162
GLS0803_114815

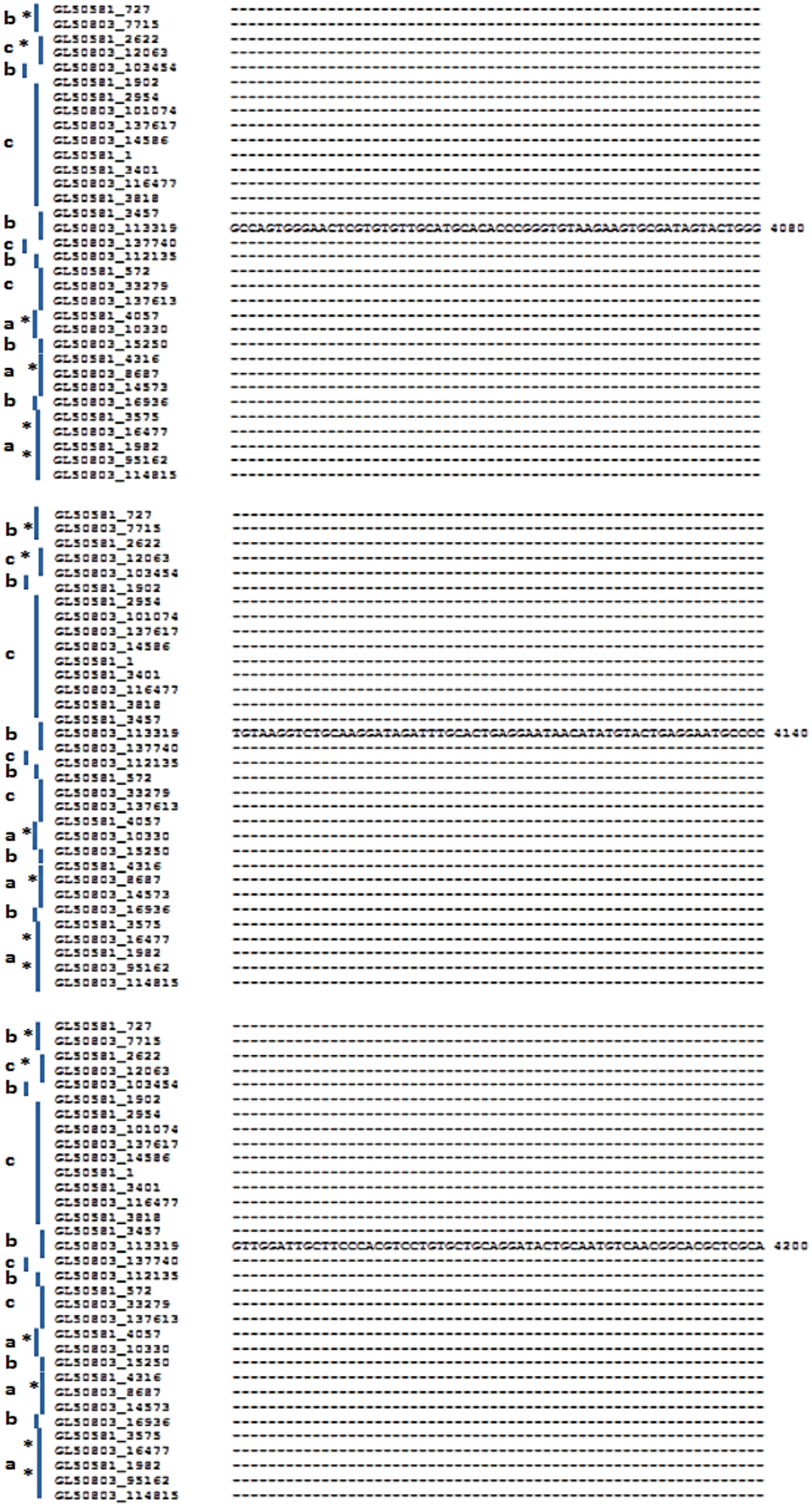

GL50581_727
GL50803_7715
GL50581_2622
GL50803_12063
GL50803_103454
GL50581_1902
GL50581_2954
GL50803_101074
GL50803_137617
GL50803_14586
GL50581_1
GL50581_3401
GL50803_116477
GL50581_3818
GL50581_3457
GL50803_113319
GL50803_137740
GL50803_112135
GL50581_572
GL50803_33279
GL50803_137613
GL50581_4057
GL50803_10330
GL50803_15250
GL50581_4316
GL50803_8687
GL50803_14573
GL50803_16936
GL50581_3575
GL50803_16477
GL50581_1982
GL50803_95162
GL50803_114815
TGCAAAAGCATCGGATCTGACAATGGGATCTGTGATATTTGTAAAGAGGGATACTACGCG 3180
GL50581_727
GL50803_7715
GL50581_2622
GL50803_12063
GL50803_103454
GL50581_1902
GL50581_2954
GL50803_101074
GL50803_137617
GL50803_14586
GL50581_1
GL50581_3401
GL50803_116477
GL50581_3818
GL50581_3457
GL50803_113319
GL50803_137740
GL50803_112135
GL50581_572
GL50803_33279
GL50803_137613
GL50581_4057
GL50803_10330
GL50803_15250
GL50581_4316
GL50803_8687
GL50803_14573
GL50803_16936
GL50581_3575
GL50803_16477
GL50581_1982
GL50803_95162
GL50803_114815
CCCCCCGACCGCTCAAGTATCAGCGACTCGTGTGTTCCGTGCACAGAGAACAGCGACACG 3240
GL50581_727
GL50803_7715
GL50581_2622
GL50803_12063
GL50803_103454
GL50581_1902
GL50581_2954
GL50803_101074
GL50803_137617
GL50803_14586
GL50581_1
GL50581_3401
GL50803_116477
GL50581_3818
GL50581_3457
GL50803_113319
GL50803_137740
GL50803_112135
GL50581_572
GL50803_33279
GL50803_137613
GL50581_4057
GL50803_10330
GL50803_15250
GL50581_4316
GL50803_8687
GL50803_14573
GL50803_16936
GL50581_3575
GL50803_16477
GL50581_1982
GL50803_95162
GL50803_114815
CTGAAGGGCATACCCCACTGCAAGAGCTGCAAGGTTTCCAACGGACTCCTGGTCTGTGTT 3300
GL50581_727
GL50803_7715
GL50581_2622
GL50803_12063
GL50803_103454
GL50581_1902
GL50581_2954
GL50803_101074
GL50803_137617
GL50803_14586
GL50581_1
GL50581_3401
GL50803_116477
GL50581_3818
GL50581_3457
GL50803_113319
GL50803_137740
GL50803_112135
GL50581_572
GL50803_33279
GL50803_137613
GL50581_4057
GL50803_10330
GL50803_15250
GL50581_4316
GL50803_8687
GL50803_14573
GL50803_16936
GL50581_3575
GL50803_16477
GL50581_1982
GL50803_95162
GL50803_114815

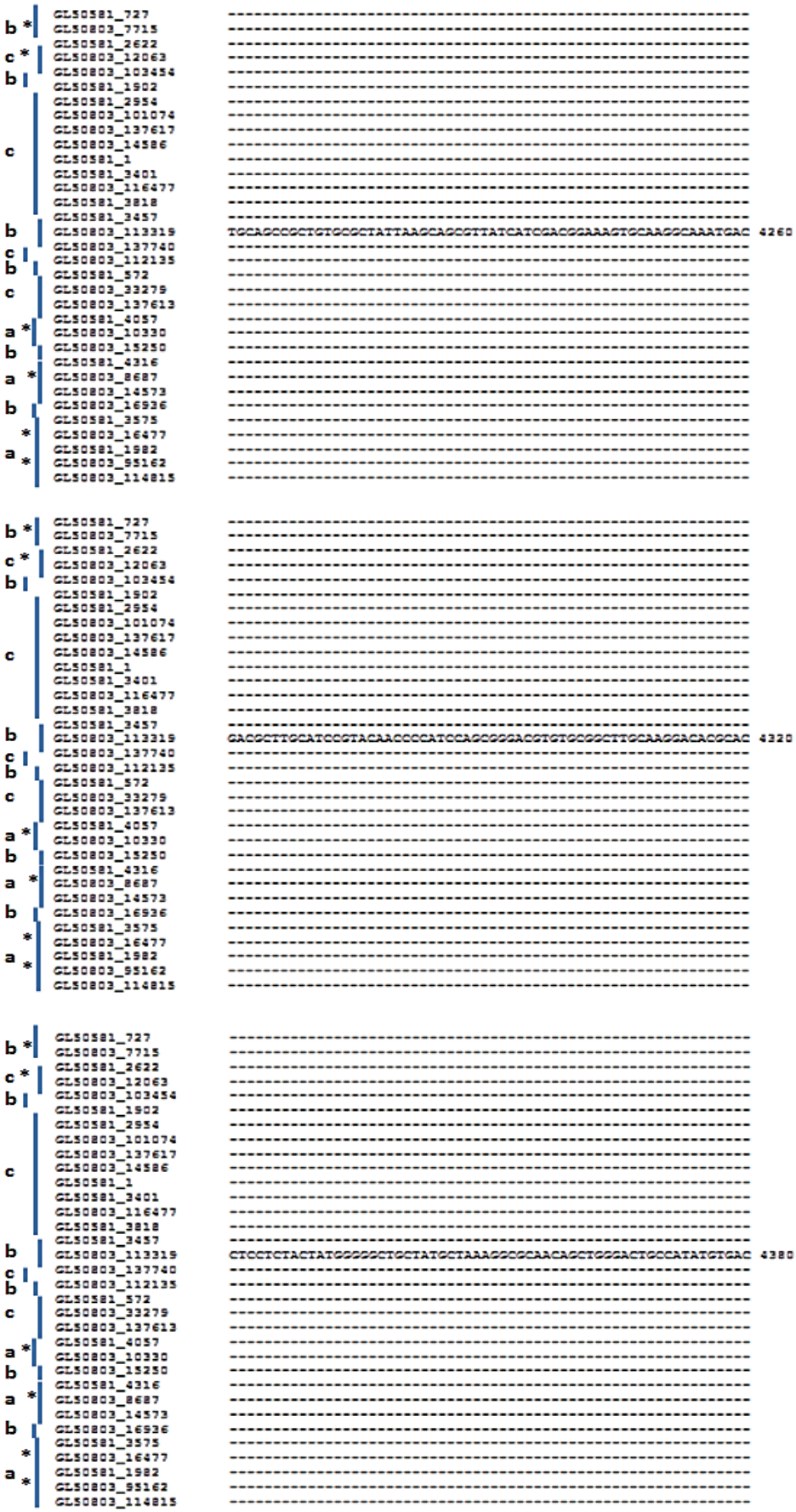

b * GL50581_727
GL50803_7715
c * GL50581_2622
GL50803_12063
b GL50803_103454
GL50581_1902
GL50581_2954
GL50803_101074
GL50803_137617
c GL50803_14586
GL50581_1
GL50581_3401
GL50803_116477
GL50581_3818
b GL50581_3457
GL50803_113319 GCCTGCGAGCAGTTCTACAGACCAATGAACGGGGGAACGGAGTGCGTGCAAACGTGCGTT 3360
c GL50803_137740
b GL50803_112135
GL50581_572
c GL50803_33279
GL50803_137613
a * GL50581_4057
GL50803_10330
b GL50803_15250
a * GL50581_4316
GL50803_8687
GL50803_14573
b GL50803_16936
* GL50581_3575
GL50803_16477
a * GL50581_1982
GL50803_95162
GL50803_114815
b * GL50581_727
GL50803_7715
c * GL50581_2622
GL50803_12063
b GL50803_103454
GL50581_1902
GL50581_2954
GL50803_101074
GL50803_137617
c GL50803_14586
GL50581_1
GL50581_3401
GL50803_116477
GL50581_3818
b GL50581_3457
GL50803_113319 GAGGCAGAGGCAGACCCCAAGGAGCACAGGTGCAGAAAGCACAAGTGTGATGTCCTGGGC 3420
c GL50803_137740
b GL50803_112135
GL50581_572
c GL50803_33279
GL50803_137613
a * GL50581_4057
GL50803_10330
b GL50803_15250
a * GL50581_4316
GL50803_8687
GL50803_14573
b GL50803_16936
* GL50581_3575
GL50803_16477
a * GL50581_1982
GL50803_95162
GL50803_114815
b * GL50581_727
GL50803_7715
c * GL50581_2622
GL50803_12063
b GL50803_103454
GL50581_1902
GL50581_2954
GL50803_101074
GL50803_137617
c GL50803_14586
GL50581_1
GL50581_3401
GL50803_116477
GL50581_3818
b GL50581_3457
GL50803_113319 ACGTTCTACTGCTCCCAGTGCGCTGCTGACGCCGAGGGCCCAGTCAACGGCGTGTGTGCT 3480
c GL50803_137740
b GL50803_112135
GL50581_572
c GL50803_33279
GL50803_137613
a * GL50581_4057
GL50803_10330
b GL50803_15250
a * GL50581_4316
GL50803_8687
GL50803_14573
b GL50803_16936
* GL50581_3575
GL50803_16477
a * GL50581_1982
GL50803_95162
GL50803_114815

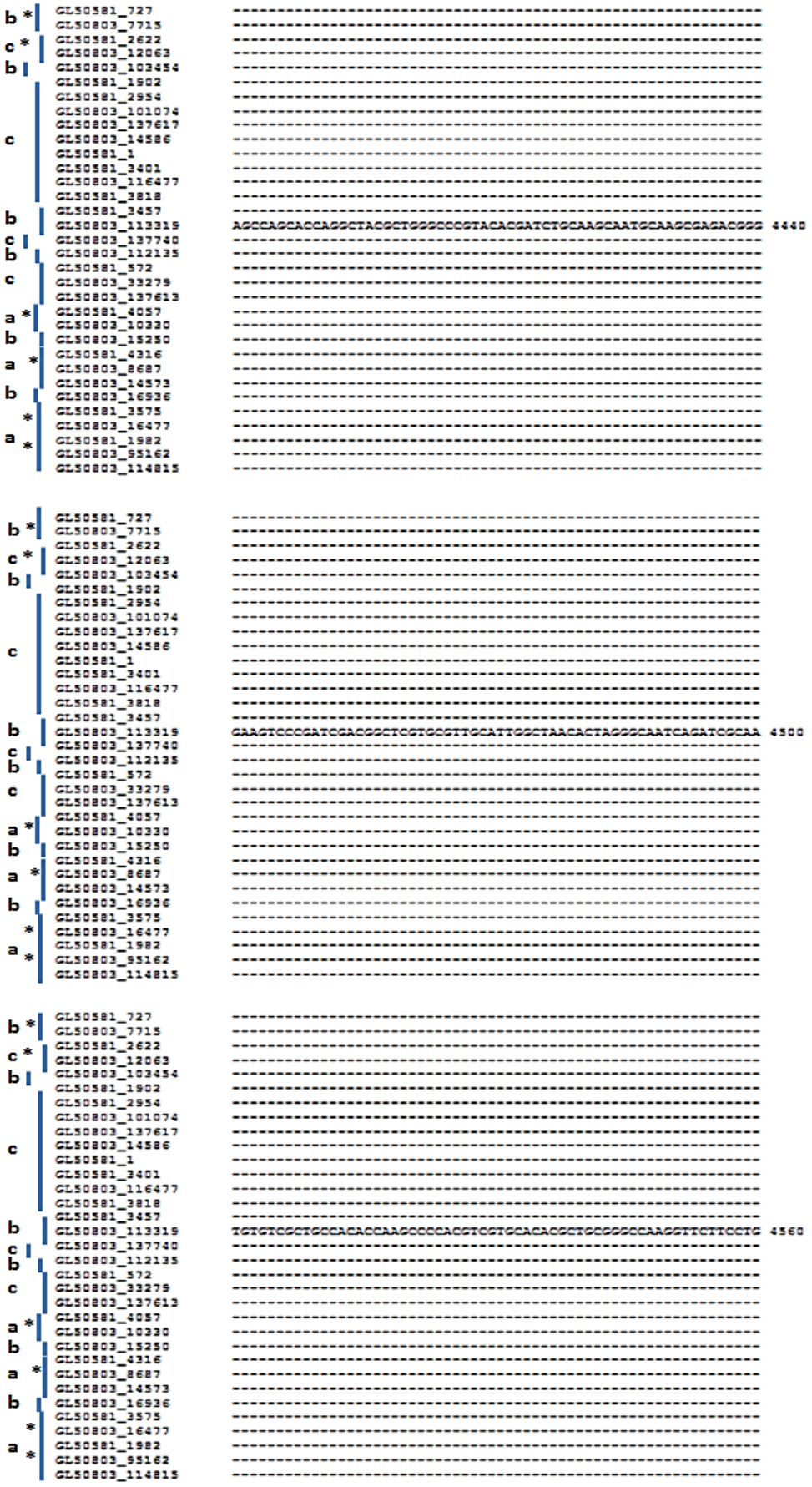

b * GL50581_727 ----------
GL50803_7715 ----------
c * GL50581_2622 ----------
GL50803_12063 ----------
b GL50803_103454 ----------
GL50581_1902 ----------
GL50581_2954 ----------
GL50803_101074 ----------
GL50803_137617 ----------
c GL50803_14586 ----------
GL50581_1 ----------
GL50581_3401 ----------
GL50803_116477 ----------
GL50581_3818 ----------
GL50581_3457 ----------
b GL50803_113319 TCAGTGGCCGCAGGCGTGTGCATACAATATAAACAAGCTAATGGCACCTGCAAATACTGC 3540
c GL50803_137740 ----------
b GL50803_112135 ----------
GL50581_572 ----------
c GL50803_33279 ----------
GL50803_137613 ----------
GL50581_4057 ----------
a * GL50803_10330 ----------
b GL50803_15250 ----------
GL50581_4316 ----------
a * GL50803_8687 ----------
GL50803_14573 ----------
b GL50803_16936 ----------
GL50581_3575 ----------
* GL50803_16477 ----------
a GL50581_1982 ----------
* GL50803_95162 ----------
GL50803_114815 ----------
b * GL50581_727 ----------
GL50803_7715 ----------
c * GL50581_2622 ----------
GL50803_12063 ----------
b GL50803_103454 ----------
GL50581_1902 ----------
GL50581_2954 ----------
GL50803_101074 ----------
GL50803_137617 ----------
c GL50803_14586 ----------
GL50581_1 ----------
GL50581_3401 ----------
GL50803_116477 ----------
GL50581_3818 ----------
GL50581_3457 ----------
b GL50803_113319 ACCCACAGATACATCAAGTACAAGGGTGGCTGCTATGAAGCGGGGTCGTTCGAGGCGGGG 3600
c GL50803_137740 ----------
b GL50803_112135 ----------
GL50581_572 ----------
c GL50803_33279 ----------
GL50803_137613 ----------
GL50581_4057 ----------
a * GL50803_10330 ----------
b GL50803_15250 ----------
GL50581_4316 ----------
a * GL50803_8687 ----------
GL50803_14573 ----------
b GL50803_16936 ----------
GL50581_3575 ----------
* GL50803_16477 ----------
a GL50581_1982 ----------
* GL50803_95162 ----------
GL50803_114815 ----------
b * GL50581_727 ----------
GL50803_7715 ----------
c * GL50581_2622 ----------
GL50803_12063 ----------
b GL50803_103454 ----------
GL50581_1902 ----------
GL50581_2954 ----------
GL50803_101074 ----------
GL50803_137617 ----------
c GL50803_14586 ----------
GL50581_1 ----------
GL50581_3401 ----------
GL50803_116477 ----------
GL50581_3818 ----------
GL50581_3457 ----------
b GL50803_113319 GACATCTGCGACGCCCAGGATGTCGTGACATCGGGCGCAGGGTCGTTTTGTGCGAAGTGT 3660
c GL50803_137740 ----------
b GL50803_112135 ----------
GL50581_572 ----------
c GL50803_33279 ----------
GL50803_137613 ----------
GL50581_4057 ----------
a * GL50803_10330 ----------
b GL50803_15250 ----------
GL50581_4316 ----------
a * GL50803_8687 ----------
GL50803_14573 ----------
b GL50803_16936 ----------
GL50581_3575 ----------
* GL50803_16477 ----------
a GL50581_1982 ----------
* GL50803_95162 ----------
GL50803_114815 ----------

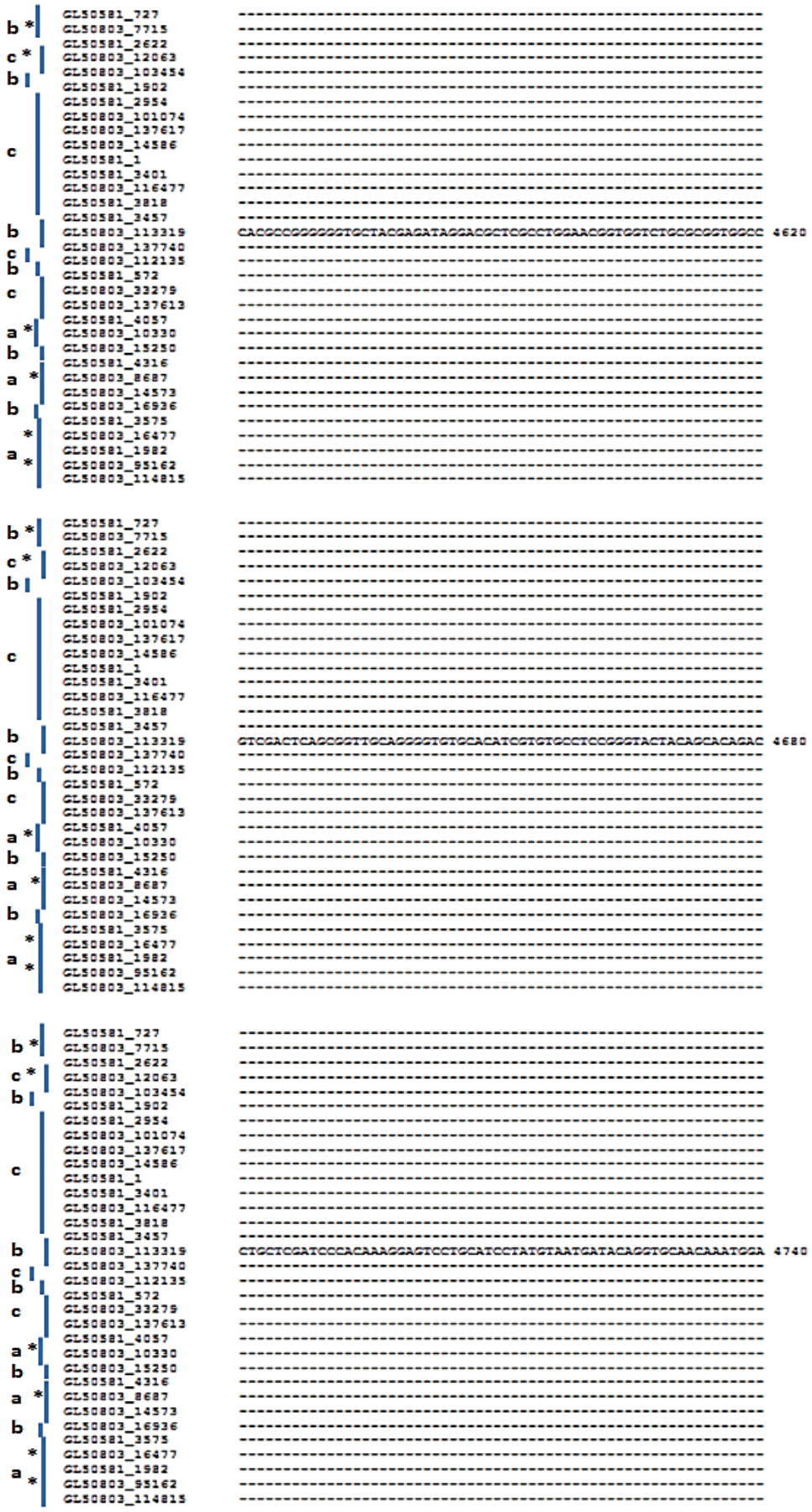

b * GL50581_727
GL50803_7715
c * GL50581_2622
GL50803_12063
b GL50803_103454
GL50581_1902
GL50581_2954
GL50803_101074
GL50803_137617
c GL50803_14586
GL50581_1
GL50581_3401
GL50803_116477
GL50581_3818
GL50581_3457
b GL50803_113319 AAGCAGAAGGATGAGGTGCCCGTTGATGGGGAGTGCAAGGTCTCGGATGACTGCACCGTC 3720
c GL50803_137740
b GL50803_112135
GL50581_572
c GL50803_33279
GL50803_137613
a * GL50581_4057
GL50803_10330
b GL50803_15250
GL50581_4316
a * GL50803_8687
GL50803_14573
b GL50803_16936
GL50581_3575
* GL50803_16477
a * GL50581_1982
GL50803_95162
GL50803_114815
b * GL50581_727
GL50803_7715
c * GL50581_2622
GL50803_12063
b GL50803_103454
GL50581_1902
GL50581_2954
GL50803_101074
GL50803_137617
c GL50803_14586
GL50581_1
GL50581_3401
GL50803_116477
GL50581_3818
GL50581_3457
b GL50803_113319 AAGGGAGATGGCATCTGCAGGTCGTGCAAACAGGGTGTTAAGTCATTCCAGATAACCAAG 3780
c GL50803_137740
b GL50803_112135
GL50581_572
c GL50803_33279
GL50803_137613
a * GL50581_4057
GL50803_10330
b GL50803_15250
GL50581_4316
a * GL50803_8687
GL50803_14573
b GL50803_16936
GL50581_3575
* GL50803_16477
a * GL50581_1982
GL50803_95162
GL50803_114815
b * GL50581_727
GL50803_7715
c * GL50581_2622
GL50803_12063
b GL50803_103454
GL50581_1902
GL50581_2954
GL50803_101074
GL50803_137617
c GL50803_14586
GL50581_1
GL50581_3401
GL50803_116477
GL50581_3818
GL50581_3457
b GL50803_113319 GATGGAAGTCCTGCAACGGTTGCATGCTTTACTAAGTGTCCGGATGGGATGTATGAGTCT 3840
c GL50803_137740
b GL50803_112135
GL50581_572
c GL50803_33279
GL50803_137613
a * GL50581_4057
GL50803_10330
b GL50803_15250
GL50581_4316
a * GL50803_8687
GL50803_14573
b GL50803_16936
GL50581_3575
* GL50803_16477
a * GL50581_1982
GL50803_95162
GL50803_114815

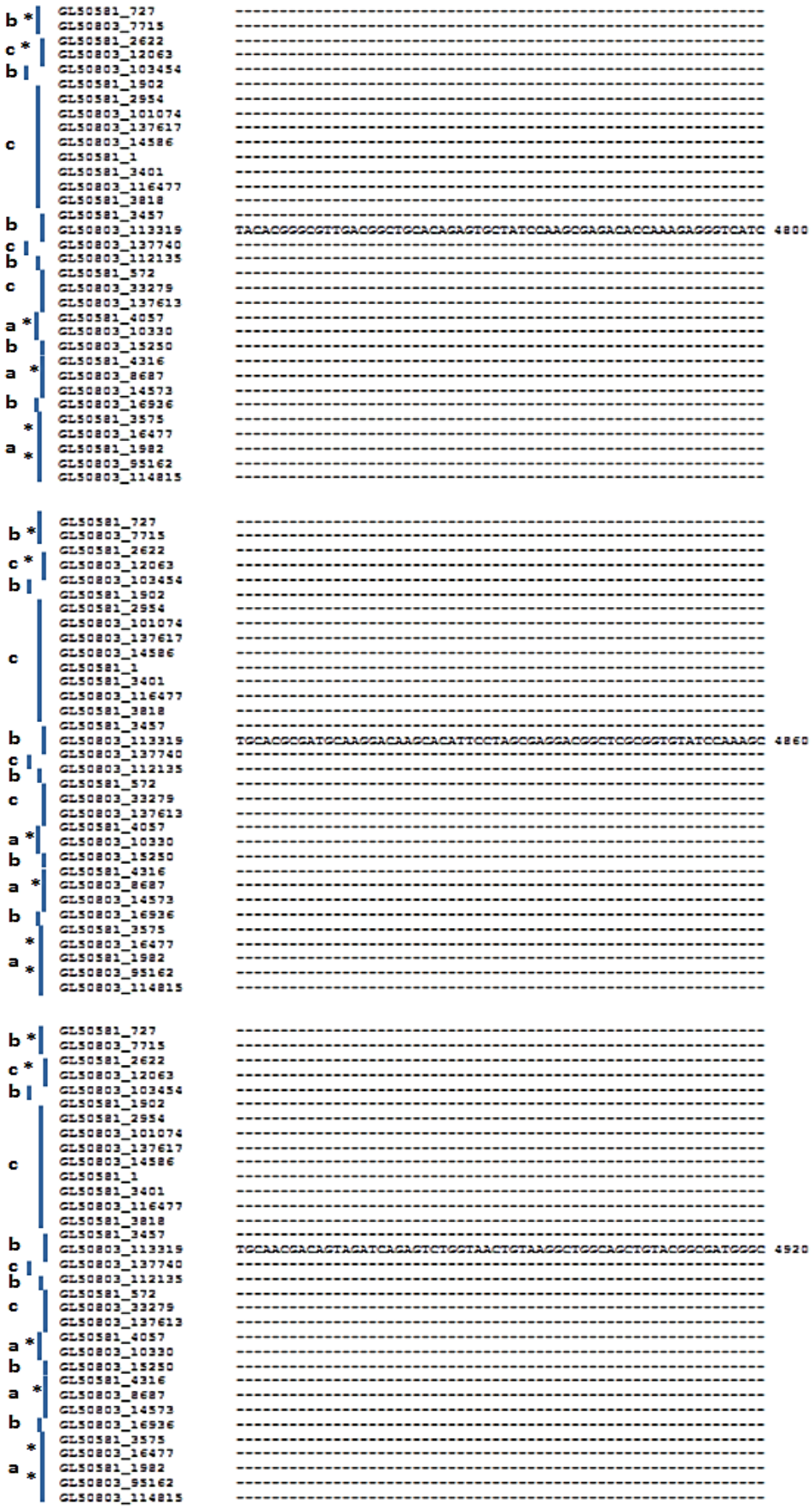

GLS0581_727
GLS0803_7715
GLS0581_2622
GLS0803_12063
GLS0803_103454
GLS0581_1902
GLS0581_2954
GLS0803_101074
GLS0803_137617
GLS0803_14586
GLS0581_1
GLS0581_3401
GLS0803_116477
GLS0581_3818
GLS0581_3457
GLS0803_113319
GLS0803_137740
GLS0803_112135
GLS0581_572
GLS0803_33279
GLS0803_137613
GLS0581_4057
GLS0803_10330
GLS0803_15250
GLS0581_4316
GLS0803_8687
GLS0803_14573
GLS0803_16936
GLS0581_3575
GLS0803_16477
GLS0581_1982
GLS0803_95162
GLS0803_114815
3900
3960
4020

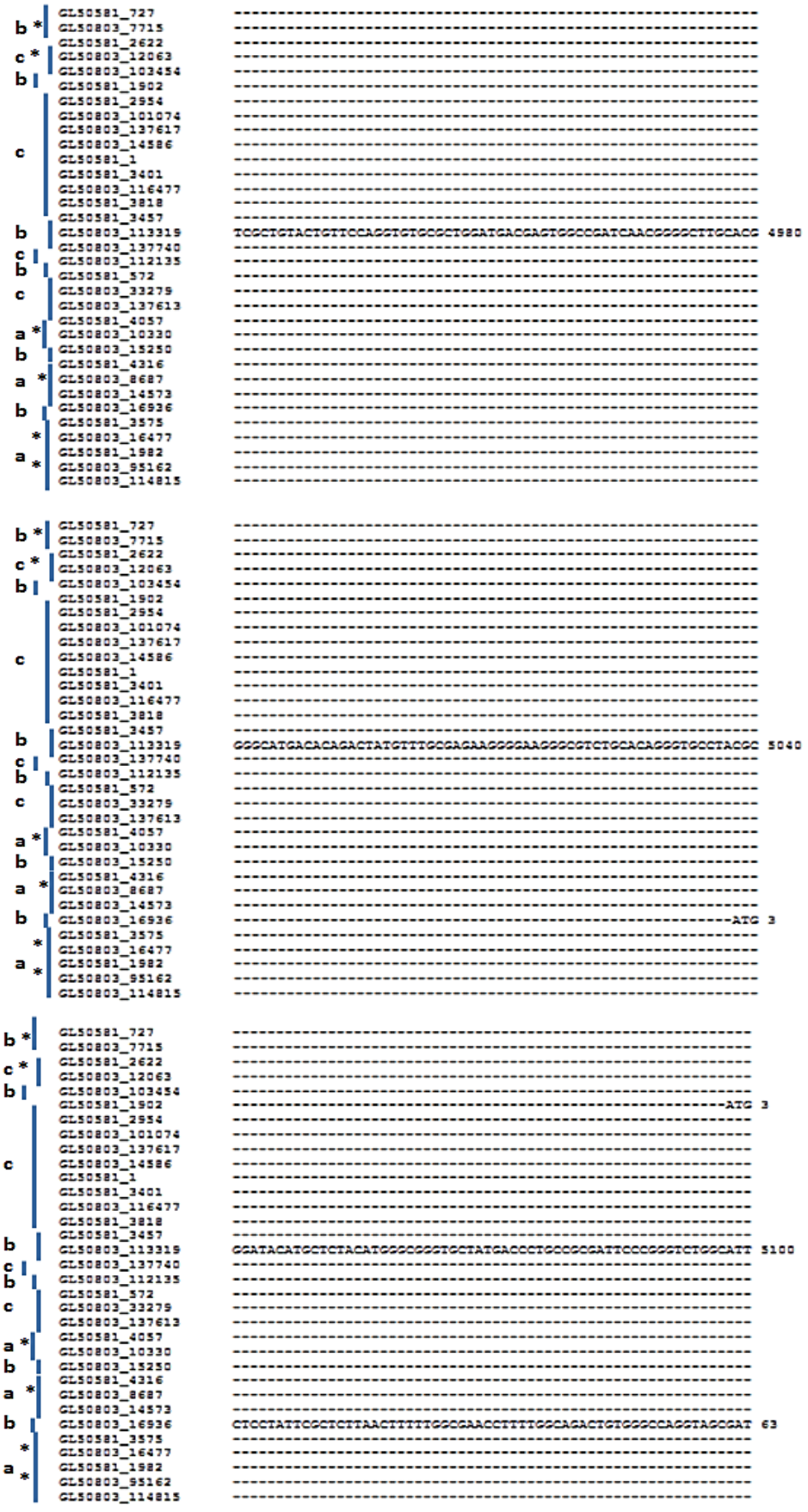

b * GLS0581_727
GLS0803_7715
c * GLS0581_2622
GLS0803_12063
b GLS0803_103454
GLS0581_1902
GLS0581_2954
GLS0803_101074
GLS0803_137617
c GLS0803_14586
GLS0581_1
GLS0581_3401
GLS0803_116477
GLS0581_3818
b GLS0581_3457
GLS0803_113319 GCCAGTGGGAACTCGTGTGTTGCATGCACACCCGGGTGTAAGAAGTGCGATAGTACTGGG 4080
c GLS0803_137740
b GLS0803_112135
GLS0581_572
c GLS0803_33279
GLS0803_137613
a * GLS0581_4057
GLS0803_10330
b GLS0803_15250
GLS0581_4316
a * GLS0803_8687
GLS0803_14573
b GLS0803_16936
GLS0581_3575
* GLS0803_16477
a * GLS0581_1982
GLS0803_95162
GLS0803_114815
b * GLS0581_727
GLS0803_7715
GLS0581_2622
c * GLS0803_12063
GLS0803_103454
b GLS0581_1902
GLS0581_2954
GLS0803_101074
GLS0803_137617
c GLS0803_14586
GLS0581_1
GLS0581_3401
GLS0803_116477
GLS0581_3818
GLS0581_3457
b GLS0803_113319 TGTAAGGTCTGCAAGGATAGATTTGCACTGAGGAATAACATATGTACTGAGGAATGCCCC 4140
c GLS0803_137740
b GLS0803_112135
GLS0581_572
c GLS0803_33279
GLS0803_137613
GLS0581_4057
a * GLS0803_10330
b GLS0803_15250
GLS0581_4316
a * GLS0803_8687
GLS0803_14573
b GLS0803_16936
GLS0581_3575
* GLS0803_16477
a * GLS0581_1982
GLS0803_95162
GLS0803_114815
b * GLS0581_727
GLS0803_7715
c * GLS0581_2622
GLS0803_12063
b GLS0803_103454
GLS0581_1902
GLS0581_2954
GLS0803_101074
GLS0803_137617
c GLS0803_14586
GLS0581_1
GLS0581_3401
GLS0803_116477
GLS0581_3818
GLS0581_3457
b GLS0803_113319 GTTGGATTGCTTCCCACGTCCTGTGCTGCAGGATACTGCAATGTCAACGGCACGCTCGCA 4200
c GLS0803_137740
b GLS0803_112135
GLS0581_572
c GLS0803_33279
GLS0803_137613
a * GLS0581_4057
GLS0803_10330
b GLS0803_15250
GLS0581_4316
a * GLS0803_8687
GLS0803_14573
b GLS0803_16936
GLS0581_3575
* GLS0803_16477
a * GLS0581_1982
GLS0803_95162
GLS0803_114815

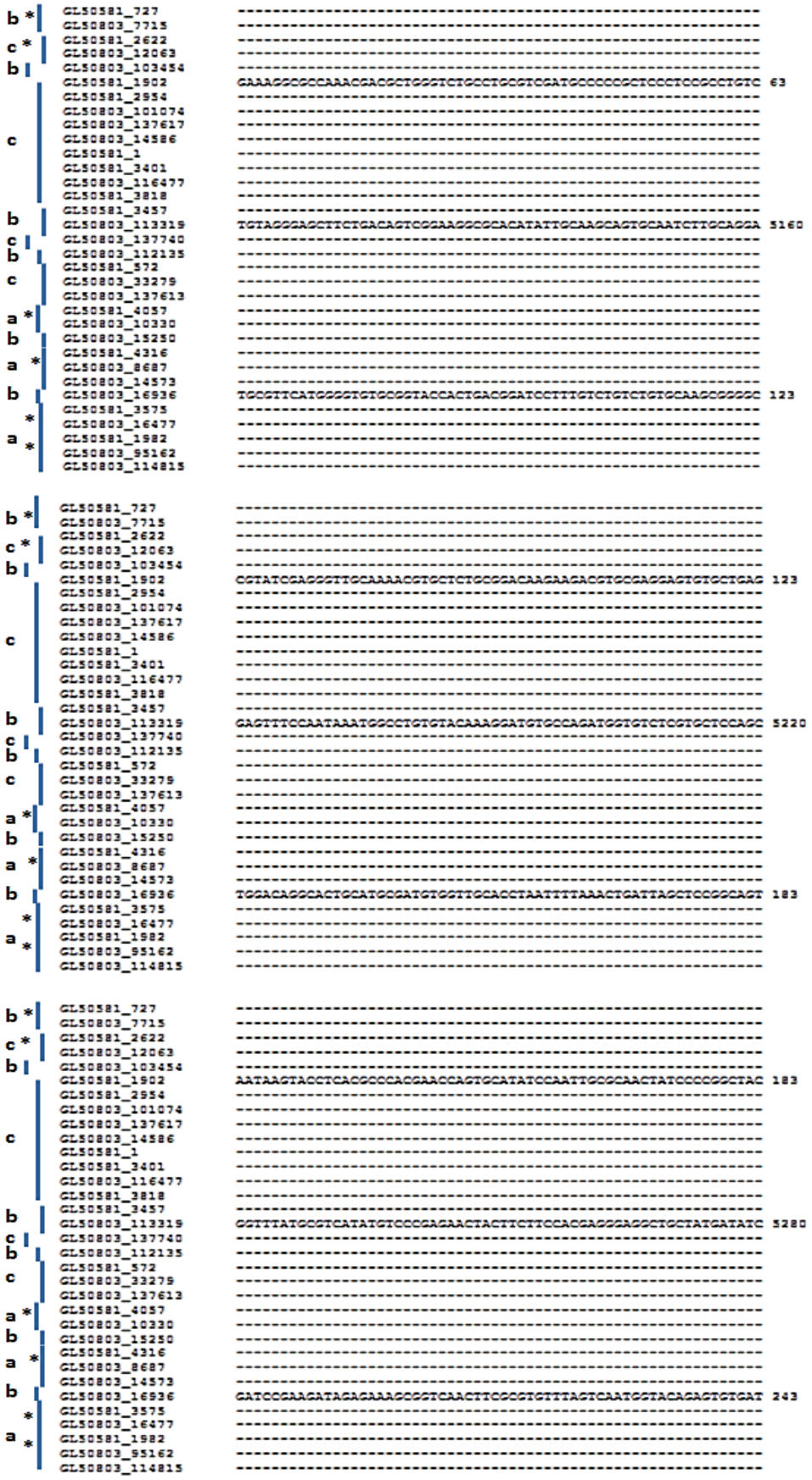

b * GL50581_727
GL50803_7715
c * GL50581_2622
GL50803_12063
b GL50803_103454
GL50581_1902
GL50581_2954
GL50803_101074
GL50803_137617
c GL50803_14586
GL50581_1
GL50581_3401
GL50803_116477
GL50581_3818
GL50581_3457
b GL50803_113319 TGCAGCCGCTGTGCGCTATTAAGCAGCGTTATCATCGACGGAAAGTGCAAGGCAAATGAC 4260
c GL50803_137740
b GL50803_112135
b GL50581_572
c GL50803_33279
GL50803_137613
GL50581_4057
a * GL50803_10330
b GL50803_15250
GL50581_4316
a * GL50803_8687
GL50803_14573
b GL50803_16936
GL50581_3575
* GL50803_16477
a * GL50581_1982
GL50803_95162
GL50803_114815
b * GL50581_727
GL50803_7715
c * GL50581_2622
GL50803_12063
b GL50803_103454
GL50581_1902
GL50581_2954
GL50803_101074
GL50803_137617
c GL50803_14586
GL50581_1
GL50581_3401
GL50803_116477
GL50581_3818
GL50581_3457
b GL50803_113319 GACGCTTGCATCCGTACAACCCCATCCAGCGGGACGTGTGCGGCTTGCAAGGACACGCAC 4320
c GL50803_137740
b GL50803_112135
b GL50581_572
c GL50803_33279
GL50803_137613
GL50581_4057
a * GL50803_10330
b GL50803_15250
GL50581_4316
a * GL50803_8687
GL50803_14573
b GL50803_16936
GL50581_3575
* GL50803_16477
a * GL50581_1982
GL50803_95162
GL50803_114815
b * GL50581_727
GL50803_7715
c * GL50581_2622
GL50803_12063
b GL50803_103454
GL50581_1902
GL50581_2954
GL50803_101074
GL50803_137617
c GL50803_14586
GL50581_1
GL50581_3401
GL50803_116477
GL50581_3818
GL50581_3457
b GL50803_113319 CTCCTCTACTATGGGGGCTGCTATGCTAAAGGCGCAACAGCTGGGACTGCCATATGTGAC 4380
c GL50803_137740
b GL50803_112135
b GL50581_572
c GL50803_33279
GL50803_137613
GL50581_4057
a * GL50803_10330
b GL50803_15250
GL50581_4316
a * GL50803_8687
GL50803_14573
b GL50803_16936
GL50581_3575
* GL50803_16477
a * GL50581_1982
GL50803_95162
GL50803_114815


b * GLS0581_727
GLS0803_7715
c * GLS0581_2622
GLS0803_12063
b GLS0803_103454
GLS0581_1902
GLS0581_2954
GLS0803_101074
GLS0803_137617
c GLS0803_14586
GLS0581_1
GLS0581_3401
GLS0803_116477
GLS0581_3818
GLS0581_3457
b GLS0803_113319 AGCCAGCACCAGGCTACGCTGGGCCCGTACACGATCTGCAAGCAATGCAAGCGAGACGGG 4440
c GLS0803_137740
b GLS0803_112135
GLS0581_572
c GLS0803_33279
GLS0803_137613
a * GLS0581_4057
GLS0803_10330
b GLS0803_15250
GLS0581_4316
a * GLS0803_8687
GLS0803_14573
b GLS0803_16936
GLS0581_3575
* GLS0803_16477
a * GLS0581_1982
GLS0803_95162
GLS0803_114815
b * GLS0581_727
GLS0803_7715
GLS0581_2622
c * GLS0803_12063
b GLS0803_103454
GLS0581_1902
GLS0581_2954
GLS0803_101074
GLS0803_137617
c GLS0803_14586
GLS0581_1
GLS0581_3401
GLS0803_116477
GLS0581_3818
GLS0581_3457
b GLS0803_113319 GAAGTCCCGATCGACGGCTCGTGCGTTGCATTGGCTAACACTAGGGCAATCAGATCGCAA 4500
c GLS0803_137740
b GLS0803_112135
GLS0581_572
c GLS0803_33279
GLS0803_137613
GLS0581_4057
a * GLS0803_10330
b GLS0803_15250
GLS0581_4316
a * GLS0803_8687
GLS0803_14573
b GLS0803_16936
GLS0581_3575
* GLS0803_16477
a * GLS0581_1982
GLS0803_95162
GLS0803_114815
b * GLS0581_727
GLS0803_7715
GLS0581_2622
c * GLS0803_12063
b GLS0803_103454
GLS0581_1902
GLS0581_2954
GLS0803_101074
GLS0803_137617
c GLS0803_14586
GLS0581_1
GLS0581_3401
GLS0803_116477
GLS0581_3818
GLS0581_3457
b GLS0803_113319 TGTGTCGCTGCCACACCAAGCCCCACGTCGTGCACACGCTGCGGGCCAAGGTTCTTCCTG 4560
c GLS0803_137740
b GLS0803_112135
GLS0581_572
c GLS0803_33279
GLS0803_137613
GLS0581_4057
a * GLS0803_10330
b GLS0803_15250
GLS0581_4316
a * GLS0803_8687
GLS0803_14573
b GLS0803_16936
GLS0581_3575
* GLS0803_16477
a * GLS0581_1982
GLS0803_95162
GLS0803_114815


b *
GL50581_727
GL50803_7715
c *
GL50581_2622
GL50803_12063
b
GL50803_103454
GL50581_1902
GL50581_2954
GL50803_101074
GL50803_137617
c
GL50803_14586
GL50581_1
GL50581_3401
GL50803_116477
GL50581_3818
GL50581_3457
b
GL50803_113319
CACGCCGGGGGGTGCTACGAGATAGGACGCTCGCCTGGAACGGTGGTCTGCGCGGTGGCC 4620
c
GL50803_137740
b
GL50803_112135
GL50581_572
c
GL50803_33279
GL50803_137613
GL50581_4057
a *
GL50803_10330
b
GL50803_15250
GL50581_4316
a *
GL50803_8687
GL50803_14573
b
GL50803_16936
GL50581_3575
*
GL50803_16477
a
GL50581_1982
*
GL50803_95162
GL50803_114815
b *
GL50581_727
GL50803_7715
c *
GL50581_2622
GL50803_12063
b
GL50803_103454
GL50581_1902
GL50581_2954
GL50803_101074
GL50803_137617
c
GL50803_14586
GL50581_1
GL50581_3401
GL50803_116477
GL50581_3818
GL50581_3457
b
GL50803_113319
GTCGACTCAGCGGTTGCAGGGGTGTGCACATCGTGTGCCTCCGGGTACTACAGCACAGAC 4680
c
GL50803_137740
b
GL50803_112135
GL50581_572
c
GL50803_33279
GL50803_137613
GL50581_4057
a *
GL50803_10330
b
GL50803_15250
GL50581_4316
a *
GL50803_8687
GL50803_14573
b
GL50803_16936
GL50581_3575
*
GL50803_16477
a
GL50581_1982
*
GL50803_95162
GL50803_114815
b *
GL50581_727
GL50803_7715
c *
GL50581_2622
GL50803_12063
b
GL50803_103454
GL50581_1902
GL50581_2954
GL50803_101074
GL50803_137617
c
GL50803_14586
GL50581_1
GL50581_3401
GL50803_116477
GL50581_3818
GL50581_3457
b
GL50803_113319
CTGCTCGATCCCACAAAGGAGTCCTGCATCCTATGTAATGATACAGGTGCAACAAATGGA 4740
c
GL50803_137740
b
GL50803_112135
GL50581_572
c
GL50803_33279
GL50803_137613
GL50581_4057
a *
GL50803_10330
b
GL50803_15250
GL50581_4316
a *
GL50803_8687
GL50803_14573
b
GL50803_16936
GL50581_3575
*
GL50803_16477
a
GL50581_1982
*
GL50803_95162
GL50803_114815

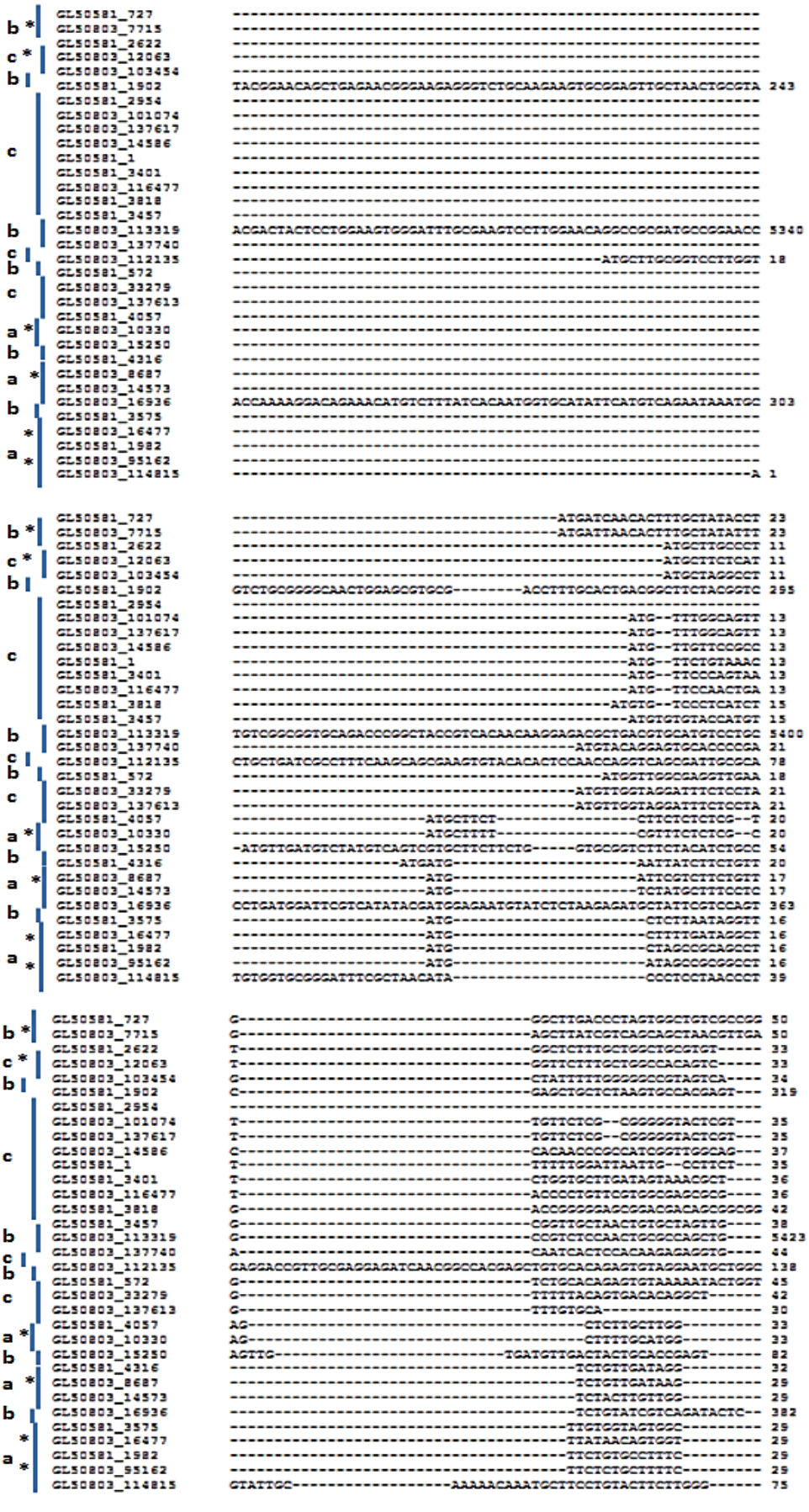

b * GL50581_727 ---------- 4800
b * GL50803_7715 ---------- 4860
c * GL50581_2622 ---------- 4920
c * GL50803_12063 ----------
b GL50803_103454 ----------
GL50581_1902 ----------
GL50581_2954 ----------
GL50803_101074 ----------
GL50803_137617 ----------
c GL50803_14586 ----------
GL50581_1 ----------
GL50581_3401 ----------
GL50803_116477 ----------
GL50581_3818 ----------
b GL50581_3457 ----------
b GL50803_113319 TACACGGGGGTTGACGGCTGCACAGAGTGCTATCCAAGCGAGACACCAAAGAGGGTCATC
c GL50803_137740 ----------
b GL50803_112135 ----------
b GL50581_572 ----------
c GL50803_33279 ----------
GL50803_137613 ----------
a * GL50581_4057 ----------
GL50803_10330 ----------
b GL50803_15250 ----------
GL50581_4316 ----------
a * GL50803_8687 ----------
GL50803_14573 ----------
b GL50803_16936 ----------
GL50581_3575 ----------
* GL50803_16477 ----------
a * GL50581_1982 ----------
GL50803_95162 ----------
GL50803_114815 ----------
b * GL50581_727 ----------
b * GL50803_7715 ----------
c * GL50581_2622 ----------
c * GL50803_12063 ----------
b GL50803_103454 ----------
GL50581_1902 ----------
GL50581_2954 ----------
GL50803_101074 ----------
GL50803_137617 ----------
c GL50803_14586 ----------
GL50581_1 ----------
GL50581_3401 ----------
GL50803_116477 ----------
GL50581_3818 ----------
b GL50581_3457 ----------
b GL50803_113319 TGCACGCGATGCAAGGACAAGCACATTCCTAGCGAGGACGGCTCGCGGTGTATCCAAAGC
c GL50803_137740 ----------
b GL50803_112135 ----------
b GL50581_572 ----------
c GL50803_33279 ----------
GL50803_137613 ----------
a * GL50581_4057 ----------
GL50803_10330 ----------
b GL50803_15250 ----------
GL50581_4316 ----------
a * GL50803_8687 ----------
GL50803_14573 ----------
b GL50803_16936 ----------
GL50581_3575 ----------
* GL50803_16477 ----------
a * GL50581_1982 ----------
GL50803_95162 ----------
GL50803_114815 ----------
b * GL50581_727 ----------
b * GL50803_7715 ----------
c * GL50581_2622 ----------
c * GL50803_12063 ----------
b GL50803_103454 ----------
GL50581_1902 ----------
GL50581_2954 ----------
GL50803_101074 ----------
GL50803_137617 ----------
c GL50803_14586 ----------
GL50581_1 ----------
GL50581_3401 ----------
GL50803_116477 ----------
GL50581_3818 ----------
b GL50581_3457 ----------
b GL50803_113319 TGCAACGACAGTAGATCAGAGTCTGGTAACTGTAAGGCTGGCAGCTGTACGGCGATGGGC
c GL50803_137740 ----------
b GL50803_112135 ----------
b GL50581_572 ----------
c GL50803_33279 ----------
GL50803_137613 ----------
a * GL50581_4057 ----------
GL50803_10330 ----------
b GL50803_15250 ----------
GL50581_4316 ----------
a * GL50803_8687 ----------
GL50803_14573 ----------
b GL50803_16936 ----------
GL50581_3575 ----------
* GL50803_16477 ----------
a * GL50581_1982 ----------
GL50803_95162 ----------
GL50803_114815 ----------


b * GL50581_727
GL50803_7715
c * GL50581_2622
GL50803_12063
b GL50803_103454
GL50581_1902
GL50581_2954
GL50803_101074
GL50803_137617
c GL50803_14586
GL50581_1
GL50581_3401
GL50803_116477
GL50581_3818
GL50581_3457
b GL50803_113319 TCGCTGTACTGTTCCAGGTGTGCGCTGGATGACGAGTGGCCGATCAACGGGGCTTGCACG 4980
c GL50803_137740
b GL50803_112135
GL50581_572
c GL50803_33279
GL50803_137613
GL50581_4057
a * GL50803_10330
b GL50803_15250
GL50581_4316
a * GL50803_8687
GL50803_14573
b GL50803_16936
GL50581_3575
* GL50803_16477
a * GL50581_1982
GL50803_95162
GL50803_114815
b * GL50581_727
GL50803_7715
c * GL50581_2622
GL50803_12063
b GL50803_103454
GL50581_1902
GL50581_2954
GL50803_101074
GL50803_137617
c GL50803_14586
GL50581_1
GL50581_3401
GL50803_116477
GL50581_3818
GL50581_3457
b GL50803_113319 GGGCATGACACAGACTATGTTTGCGAGAAGGGGAAGGGCGTCTGCACAGGGTGCCTACGC 5040
c GL50803_137740
b GL50803_112135
GL50581_572
c GL50803_33279
GL50803_137613
GL50581_4057
a * GL50803_10330
b GL50803_15250
GL50581_4316
a * GL50803_8687
GL50803_14573
b GL50803_16936 ATG 3
GL50581_3575
* GL50803_16477
a * GL50581_1982
GL50803_95162
GL50803_114815
b * GL50581_727
GL50803_7715
c * GL50581_2622
GL50803_12063
b GL50803_103454
GL50581_1902 ATG 3
GL50581_2954
GL50803_101074
GL50803_137617
c GL50803_14586
GL50581_1
GL50581_3401
GL50803_116477
GL50581_3818
GL50581_3457
b GL50803_113319 GGATACATGCTCTACATGGGCGGGTGCTATGACCCTGCCGCGATTCCCGGGTCTGGCATT 5100
c GL50803_137740
b GL50803_112135
GL50581_572
c GL50803_33279
GL50803_137613
GL50581_4057
a * GL50803_10330
b GL50803_15250
GL50581_4316
a * GL50803_8687
GL50803_14573
b GL50803_16936 CTCCTATTCGCTCTTAACTTTTTGGCGAACCTTTTGGCAGACTGTGGGCCAGGTAGCGAT 63
GL50581_3575
* GL50803_16477
a * GL50581_1982
GL50803_95162
GL50803_114815


b * GL50581_727 ---------- 63
GL50803_7715 ----------
c * GL50581_2622 ----------
GL50803_12063 ----------
b GL50803_103454 ----------
GL50581_1902 GAAAGGCGCCAAACGACGCTGGGTCTGCCTGCGTCGATGCCCCCGCTCCCTCCGCCTGTC
GL50581_2954 ----------
GL50803_101074 ----------
GL50803_137617 ----------
c GL50803_14586 ----------
GL50581_1 ----------
GL50581_3401 ----------
GL50803_116477 ----------
GL50581_3818 ----------
b GL50581_3457 ---------- 5160
GL50803_113319 TGTAGGGAGCTTCTGACAGTCGGAAGGCGCACATATTGCAAGCAGTGCAATCTTGCAGGA
c GL50803_137740 ----------
b GL50803_112135 ----------
GL50581_572 ----------
c GL50803_33279 ----------
GL50803_137613 ----------
a * GL50581_4057 ----------
GL50803_10330 ----------
b GL50803_15250 ----------
a * GL50581_4316 ----------
GL50803_8687 ----------
GL50803_14573 ----------
b GL50803_16936 TGCGTTCATGGGGTGTGCGGTACCACTGACGGATCCTTTGTCTGTCTGTGCAAGCGGGC 123
GL50581_3575 ----------
* GL50803_16477 ----------
a * GL50581_1982 ----------
GL50803_95162 ----------
GL50803_114815 ----------

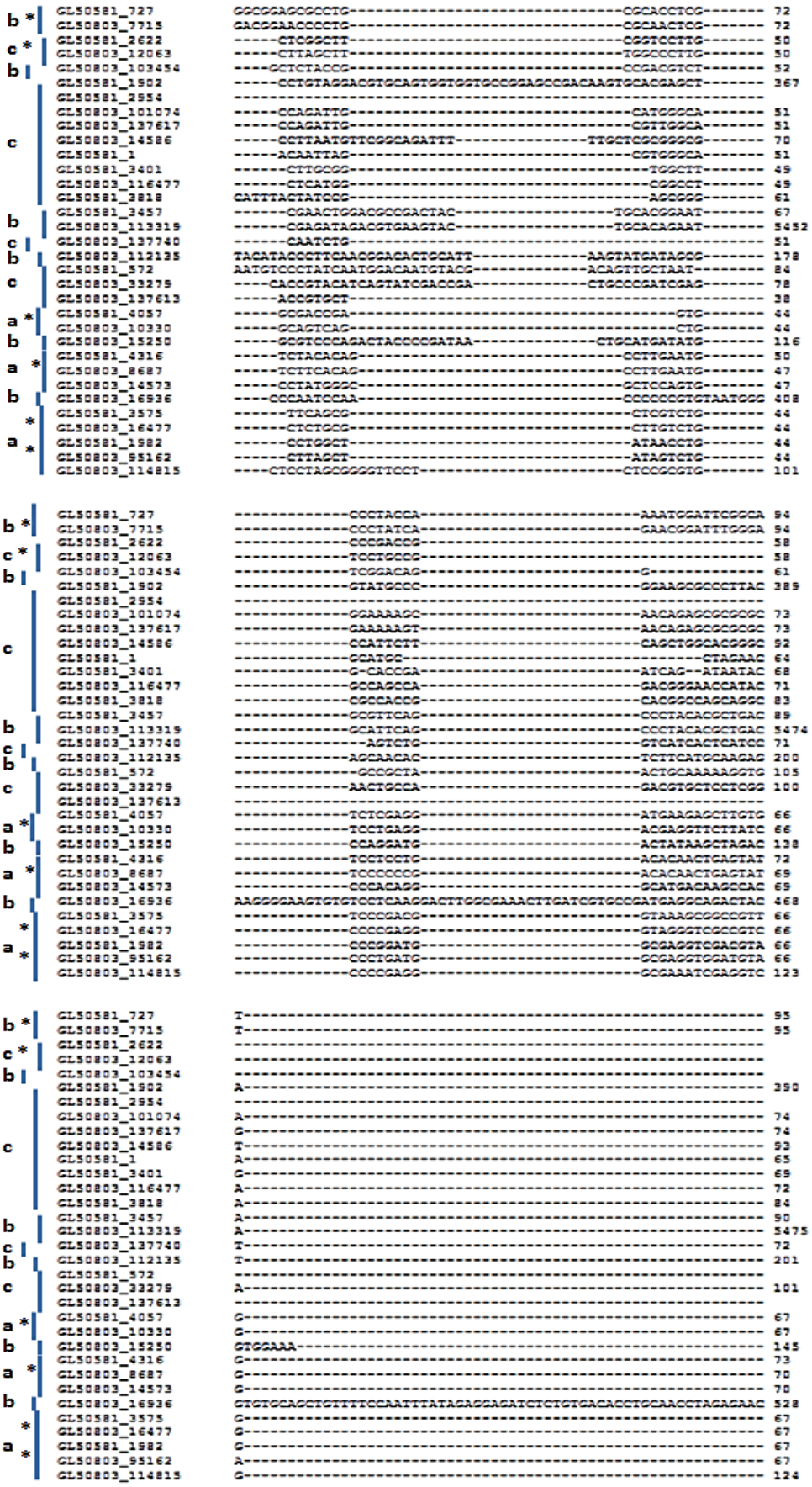
b * GL50581_727 ----------
GL50803_7715 ----------
c * GL50581_2622 ----------
GL50803_12063 ----------
b GL50803_103454 ----------
GL50581_1902 CGTATCGAGGGTTGCAAAACGTGCTCTGCGGACAAGAAGACGTGCGAGGAGTGTGCTGAG 123
GL50581_2954 ----------
GL50803_101074 ----------
GL50803_137617 ----------
c GL50803_14586 ----------
GL50581_1 ----------
GL50581_3401 ----------
GL50803_116477 ----------
GL50581_3818 ----------
b GL50581_3457 ----------
GL50803_113319 GAGTTTCCAATAAATGGCCTGTGTACAAAGGATGTGCCAGATGGTGTCTCGTGCTCCAGC 5220
c GL50803_137740 ----------
b GL50803_112135 ----------
GL50581_572 ----------
c GL50803_33279 ----------
GL50803_137613 ----------
a * GL50581_4057 ----------
GL50803_10330 ----------
b GL50803_15250 ----------
a * GL50581_4316 ----------
GL50803_8687 ----------
GL50803_14573 ----------
b GL50803_16936 TGGACAGGCACTGCATGCGATGTGGTTGCACCTAATTTTAAACTGATTAGCTCCGGCAGT 183
GL50581_3575 ----------
* GL50803_16477 ----------
a * GL50581_1982 ----------
GL50803_95162 ----------
GL50803_114815 ----------

b * GL50581_727 ----------
GL50803_7715 ----------
c * GL50581_2622 ----------
GL50803_12063 ----------
b GL50803_103454 ----------
GL50581_1902 AATAAGTACCTCACGCCCACGAACCAGTGCATATCCAATTGCGCAACTATCCCCGGCTAC 183
GL50581_2954 ----------
GL50803_101074 ----------
GL50803_137617 ----------
c GL50803_14586 ----------
GL50581_1 ----------
GL50581_3401 ----------
GL50803_116477 ----------
GL50581_3818 ----------
b GL50581_3457 ----------
GL50803_113319 GGTTTATGCGTCATATGTCCCGAGAACTACTTCTTCCACGAGGGAGGCTGCTATGATATC 5280
c GL50803_137740 ----------
b GL50803_112135 ----------
GL50581_572 ----------
c GL50803_33279 ----------
GL50803_137613 ----------
a * GL50581_4057 ----------
GL50803_10330 ----------
b GL50803_15250 ----------
a * GL50581_4316 ----------
GL50803_8687 ----------
GL50803_14573 ----------
b GL50803_16936 GATCCGAAGATAGAGAAAGCGGTCAACTTCGCGGTGTTTAGTCAATGGTACAGAGTGTGAT 243
GL50581_3575 ----------
* GL50803_16477 ----------
a * GL50581_1982 ----------
GL50803_95162 ----------
GL50803_114815 ----------


b * GL50581_727 ---------- 0
GL50803_7715 ---------- 0
GL50581_2622 ---------- 0
c * GL50803_12063 ---------- 0
GL50803_103454 ---------- 0
b GL50581_1902 TACGGAACAGCTGAGAACGGGAAGAGGGTCTGCAAGAAGTGCGGAGTTGCTAACTGCGTA 243
GL50581_2954 ---------- 0
GL50803_101074 ---------- 0
GL50803_137617 ---------- 0
c GL50803_14586 ---------- 0
GL50581_1 ---------- 0
GL50581_3401 ---------- 0
GL50803_116477 ---------- 0
GL50581_3818 ---------- 0
GL50581_3457 ---------- 0
b GL50803_113319 ACGACTACTCCTGGAAGTGGGATTTGCGAAGTCCTTGGAACAGGCCGCGATGCCGGAACC 5340
GL50803_137740 ---------- 0
c GL50803_112135 ----------ATGCTTGCGGTCCTTGGT 18
b GL50581_572 ---------- 0
c GL50803_33279 ---------- 0
GL50803_137613 ---------- 0
GL50581_4057 ---------- 0
a * GL50803_10330 ---------- 0
b GL50803_15250 ---------- 0
GL50581_4316 ---------- 0
a * GL50803_8687 ---------- 0
GL50803_14573 ---------- 0
b GL50803_16936 ACCAAAAGGACAGAAACATGTCTTTATCACAATGGTGCATATTCATGTCAGAATAAATGC 303
GL50581_3575 ---------- 0
* GL50803_16477 ---------- 0
GL50581_1982 ---------- 0
a * GL50803_95162 ---------- 0
GL50803_114815 ----------A 1

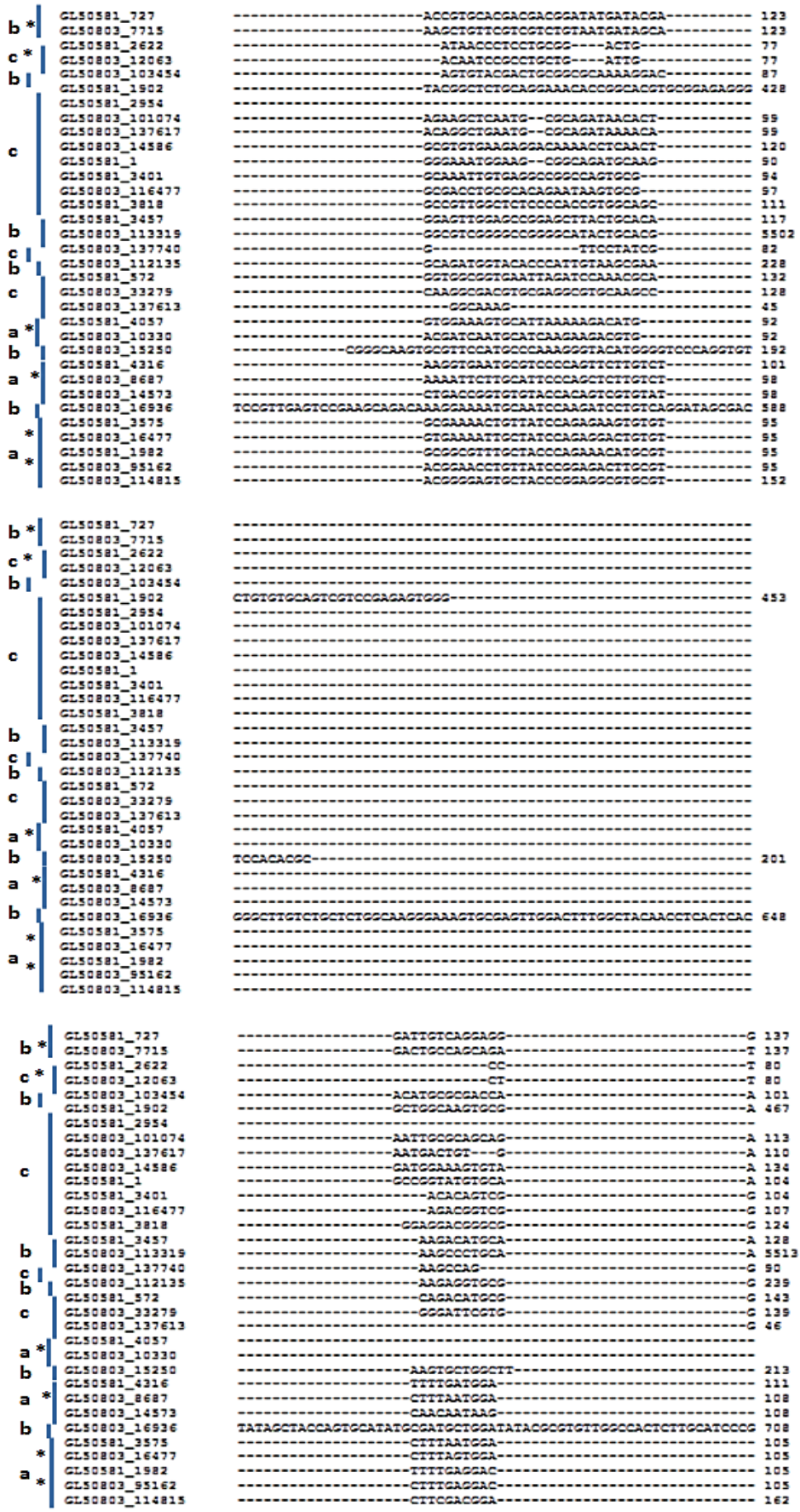
GL50581_727 ----------ATGATCAACACTTTGCTATACCT 23
b * GL50803_7715 ----------ATGATTAACACTTTGCTATATTT 23
GL50581_2622 ----------ATGCTTGCCCT 11
c * GL50803_12063 ----------ATGCTTCTCAT 11
b GL50803_103454 ----------ATGCTAGGCCT 11
GL50581_1902 GTCTGCGGGGGCAACTGGAGGGTGCG--------ACCTTTGCACTGACGGCTTCTACGGTC 295
GL50581_2954 ---------- 0
GL50803_101074 ----------ATG--TTTGGCAGTT 13
GL50803_137617 ----------ATG--TTTGGCAGTT 13
c GL50803_14586 ----------ATG--TTGTTCCGCC 13
GL50581_1 ----------ATG--TTCTGTAAAC 13
GL50581_3401 ----------ATG--TTCCCAGTAA 13
GL50803_116477 ----------ATG--TTCCAACTGA 13
GL50581_3818 ----------ATGTG--TCCCTCATCT 15
GL50581_3457 ----------ATGTGTGTACCATGT 15
b GL50803_113319 TGTCGGCGGTGCAGACCCGGCTACCGTCACAACAAGGAGACGCTGACGTGCATGTCCTGC 5400
GL50803_137740 ----------ATGTACAGGAGTGCACCCCGA 21
c GL50803_112135 CTGCTGATCGCCTTTCAAGCAGCGAAGTGTACACACTCCAACCAGGTCAGCGATTGCGCA 78
b GL50581_572 ----------ATGGTTGGCGAGGTTGAA 18
c GL50803_33279 ----------ATGTTGGTAGGATTTCTCCTA 21
GL50803_137613 ----------ATGTTGGTAGGATTTCTCCTA 21
GL50581_4057 ----------ATGCTTCT----------CTTCTCTCTCG--T 20
a * GL50803_10330 ----------ATGCTTTT----------CGTTTCTCTCG--C 20
b GL50803_15250 -ATGTTGATGTCTATGTCAGTCGTGCTTCTTCTG-----GTGCGGTCTTCTACATCTGCC 54
GL50581_4316 ----------ATGATG----------AATTATCTTCTGTT 20
a * GL50803_8687 ----------ATG----------ATTGGTCTTCTGTT 17
GL50803_14573 ----------ATG----------TCTATGCTTTCCTC 17
b GL50803_16936 CCTGATGGATTCGTCATATACGATGGAGAATGTATCTCTAAGAGATGCTATTCGTCCAGT 363
GL50581_3575 ----------ATG----------CTCTTAATAGGTT 16
* GL50803_16477 ----------ATG----------CTTTTGATAGGCT 16
GL50581_1982 ----------ATG----------CTAGCCGCAGCCT 16
a * GL50803_95162 ----------ATG----------ATAGCCGCGGCCT 16
GL50803_114815 TGTGGTGCGGGATTTCGCTAACATA----------CCCTCCTAACCCT 39

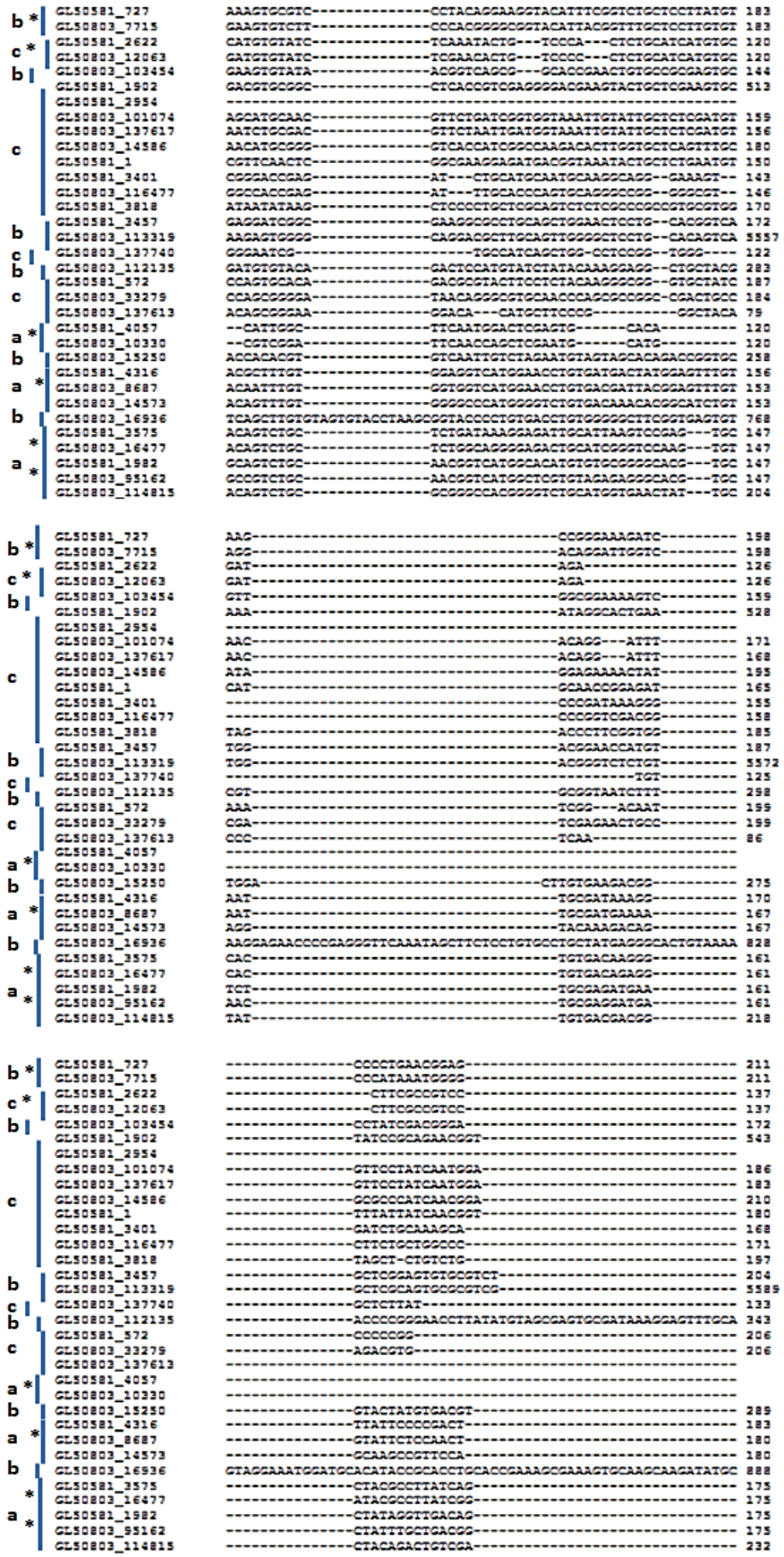
GL50581_727 G----------GGCTTGACCCTAGTGGCTGTCGCCGG 50
b * GL50803_7715 G----------AGCTTATCGTCAGCAGCTAACGTTGA 50
GL50581_2622 T----------GGCTCTTTGCTGGCTGCGTGT----- 33
c * GL50803_12063 T----------GGTTCTTTGCTGGCCACAGTC----- 33
b GL50803_103454 G----------CTATTTTTGGGGGCCGTAGTCA---- 34
GL50581_1902 C----------GAGCTGCTCTAAGTGCCACGAGT--- 319
GL50581_2954 ---------- 0
GL50803_101074 T----------TGTTCTCG--CGGGGGTACTCGT--- 35
GL50803_137617 T----------TGTTCTCG--CGGGGGTACTCGT--- 35
c GL50803_14586 C----------CACAACCCGCCATCGGTTGGCAG--- 37
GL50581_1 T----------TTTTTGGATTAATTG--CCTTCT--- 35
GL50581_3401 T----------CTGGTGCTTGATAGTAAACGCT---- 36
GL50803_116477 T----------ACCCCTGTTCGTGGCGAGCGCG---- 36
GL50581_3818 G----------ACCGGGGGAGCGGCACGACAGCGGCGG 42
GL50581_3457 G----------CGGTTGCTAACTGTGCTAGTTG---- 38
b GL50803_113319 G----------CCGTCTCCAACTGCGCCAGCTG---- 5423
c GL50803_137740 A----------CAATCACTCCACAAGAGAGGTG---- 44
b GL50803_112135 GAGGACCGTTGCGAGGAGATCAACGGCCACGAGCTGTGCACAGAGTGTAGGAATGCTGGC 138
GL50581_572 G----------TCTGCACAGAGTGTAAAAATACTGGT 45
c GL50803_33279 G----------TTTTTACAGTGACACAGGCT------ 42
GL50803_137613 G----------TTTGTGCA---------- 30
GL50581_4057 AG----------CTCTTGCTTGG---------- 33
a * GL50803_10330 AG----------CTTTTGCATGG---------- 33
b GL50803_15250 AGTTG----------TGATGTTGACTACTGCACCGAGT------ 82
GL50581_4316 ----------TCTGTTGATAGG---------- 32
a * GL50803_8687 ----------TCTGTTGATAAG---------- 29
GL50803_14573 ----------TCTACTTGTTGG---------- 29
b GL50803_16936 ----------TCTGTATCGTCAGATACTC-- 382
GL50581_3575 ----------TTGTGGTAGTGGC---------- 29
* GL50803_16477 ----------TTATAACAGTGGT---------- 29
GL50581_1982 ----------TTCTGTGCCTTTC---------- 29
a * GL50803_95162 ----------TTCTCTGCTTTTC---------- 29
GL50803_114815 GTATTGC----------AAAAACAAATGCTTCCTGTACTTCTTGGG------ 75


GL50581_727 GGCGGAGCGGCTG------------------------------CGCACCTCG------- 72
GL50803_7715 GACGGAACCCCTG------------------------------CGCAACTCG------- 72
GL50581_2622 -----CTGGGCTT------------------------------CGGTCCTTG------- 50
GL50803_12063 -----CTTAGCTT------------------------------TGGCCCTTG------- 50
GL50803_103454 ----GCTCTACCG------------------------------CCGACGTCT------- 52
GL50581_1902 -----CCTGTAGGACGTGCAGTGGTGGTGCCGGAGCCGACAAGTGCACGAGCT------- 367
GL50581_2954 ----------
GL50803_101074 -----CCAGATTG------------------------------CATGGGCA------- 51
GL50803_137617 -----CCAGATTG------------------------------CGTTGGCA------- 51
GL50803_14586 -----CCTTAATGTTCGGCAGATTT--------------TTGCTCGGGGGCG------- 70
GL50581_1 -----ACAATTAG------------------------------CGTGGGCA------- 51
GL50581_3401 ------CTTGCGG------------------------------TGGCTT------- 49
GL50803_116477 ------CTCATGG------------------------------CGGCCT------- 49
GL50581_3818 CATTTACTATCCG------------------------------AGCGGG------- 61
GL50581_3457 ------CGAACTGGACGCCGACTAC----------------TGCACGGAAT------- 67
GL50803_113319 ------CGAGATAGACGTGAAGTAC----------------TGCACAGAAT------- 5452
GL50803_137740 ------CAATCTG--- 51
GL50803_112135 TACATACCCTTCAACGGACACTGCATT-------------AAGTATGATAGCG------- 178
GL50581_572 AATGTCCCTATCAATGGACAATGTACG-------------ACAGTTGCTAAT-------- 84
GL50803_33279 ----CACCGTACATCAGTATCGACCGA-------------CTGCCCGATCGAG------- 78
GL50803_137613 -----ACCGTGCT--- 38
GL50581_4057 -----GCGACCGA------------------------------GTG------- 44
GL50803_10330 -----CCAGTCAG------------------------------CTG------- 44
GL50803_15250 -----GCGTCCCAGACTACCCCGATAA-------------CTGCATGATATG------- 116
GL50581_4316 -----TCTACACAG------------------------------CCTTGAATG------- 50
GL50803_8687 -----TCTTCACAG------------------------------CCTTGAATG------- 47
GL50803_14573 -----CCTATGGGC------------------------------GCTCCAGTG------- 47
GL50803_16936 ----CCCAATCCAA------------------------------CCCCCCGTGTAATGGG 408
GL50581_3575 ------TTCAGCG------------------------------CTCGTCTG------- 44
GL50803_16477 ------CTCTGCG------------------------------CTTGTCTG------- 44
GL50581_1982 ------CCTGGCT------------------------------ATAACCTG------- 44
GL50803_95162 ------CTTAGCT------------------------------ATAGTCTG------- 44
GL50803_114815 ----CTCCTAGCGGGGTTCCT----------------------CTCCGCGTG------- 101

GL50581_727 --------------CCCTACCA----------------------AAATGCATTCGGCA 94
GL50803_7715 --------------CCCTATCA----------------------GAACGGATTTGGGA 94
GL50581_2622 --------------CCCGACCG---------------------------------- 58
GL50803_12063 --------------TCCTGCCG---------------------------------- 58
GL50803_103454 --------------TCGGACAG----------------------G---------- 61
GL50581_1902 --------------GTATGCCC----------------------GGAAGCGCCCTTAC 389
GL50581_2954 ----------
GL50803_101074 --------------GGAAAAGC----------------------AACAGAGCGCGCGC 73
GL50803_137617 --------------GAAAAAGT----------------------AACAGAGCGCGCGC 73
GL50803_14586 --------------CCATTCTT----------------------CAGCTGGCACGGGC 92
GL50581_1 --------------GCATGC----------------------CTAGAAC 64
GL50581_3401 --------------G-CACCGA----------------------ATCAG--ATAATAC 68
GL50803_116477 --------------GCCAGCCA----------------------GACGGGAACCATAC 71
GL50581_3818 --------------CGCCACCG----------------------CACGGCCAGCAGGC 83
GL50581_3457 --------------GCGTTCAG----------------------CCCTACACGCTGAC 89
GL50803_113319 --------------GCATTCAG----------------------CCCTACACGCTGAC 5474
GL50803_137740 ----------------AGTCTG----------------------GTCATCACTCATCC 71
GL50803_112135 --------------AGCAACAC----------------------TCTTCATGCAAGAG 200
GL50581_572 --------------GCCGCTA----------------------ACTGCAAAAAGGTG 105
GL50803_33279 --------------AACTGCCA----------------------GACGTGCTCCTCGG 100
GL50803_137613 ----------
GL50581_4057 --------------TCTCGAGG----------------------ATGAAGAGCTTGTG 66
GL50803_10330 --------------TCCTGAGG----------------------ACGAGGTTCTTATC 66
GL50803_15250 --------------CCAGGATG----------------------ACTATAAGCTAGAC 138
GL50581_4316 --------------TCCTCCTG----------------------ACACAACTGAGTAT 72
GL50803_8687 --------------TCCCCCCG----------------------ACACAACTGAGTAT 69
GL50803_14573 --------------CCCACAGG----------------------GCATGACAAGCCAC 69
GL50803_16936 AAGGGGAAGTGTGTCCTCAAGGACTTGGCGAAACTTGATCGTGCCGATGAGCCAGACTAC 468
GL50581_3575 --------------TCCCGACG----------------------GTAAAGCGGCCGTT 66
GL50803_16477 --------------CCCCGAGG----------------------GTAGGGTCGCCGTC 66
GL50581_1982 --------------CCCGGATG----------------------GCGAGGTCGACGTA 66
GL50803_95162 --------------CCCTGATG----------------------GCGAGGTGGATGTA 66
GL50803_114815 --------------CCCCGAGG----------------------GCGAAATCGAGGTC 123

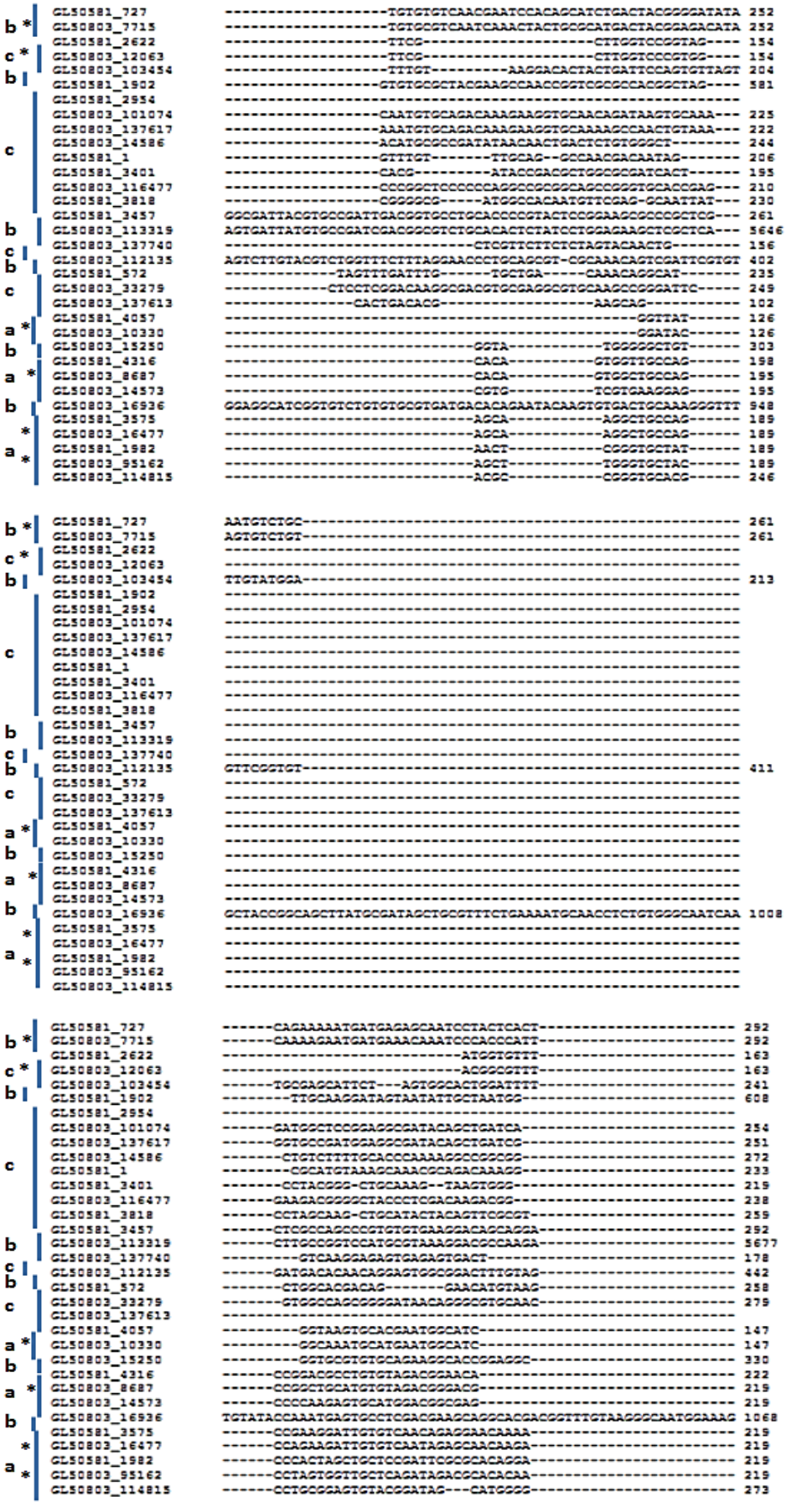
GL50581_727 T---------- 95
GL50803_7715 T---------- 95
GL50581_2622 ----------
GL50803_12063 ----------
GL50803_103454 ----------
GL50581_1902 A---------- 390
GL50581_2954 ----------
GL50803_101074 A---------- 74
GL50803_137617 G---------- 74
GL50803_14586 T---------- 93
GL50581_1 A---------- 65
GL50581_3401 G---------- 69
GL50803_116477 A---------- 72
GL50581_3818 A---------- 84
GL50581_3457 A---------- 90
GL50803_113319 A---------- 5475
GL50803_137740 T---------- 72
GL50803_112135 T---------- 201
GL50581_572 ----------
GL50803_33279 A---------- 101
GL50803_137613 ----------
GL50581_4057 G---------- 67
GL50803_10330 G---------- 67
GL50803_15250 GTGGAAA---------- 145
GL50581_4316 G---------- 73
GL50803_8687 G---------- 70
GL50803_14573 G---------- 70
GL50803_16936 GTGTGCAGCTGTTTTCCAATTTATAGAGGAGATCTCTGTGACACCTGCAACCTAGAGAAC 528
GL50581_3575 G---------- 67
GL50803_16477 G---------- 67
GL50581_1982 G---------- 67
GL50803_95162 A---------- 67
GL50803_114815 G---------- 124


b *
c *
b
c
b
c
b
a *
b
a *
b
*
a *
GL50581_727 ----------ACCGTGCACGACGACGGATATGATACGA---------- 123
GL50803_7715 ----------AAGCTGTTCGTCGTCTGTAATGATAGCA---------- 123
GL50581_2622 ----------ATAACCCTCCTGCGG----ACTG---------- 77
GL50803_12063 ----------ACAATCCGCCTGCTG----ATTG---------- 77
GL50803_103454 ----------AGTGTACGACTGCGGCGCAAAAGGAC---------- 87
GL50581_1902 ----------TACGGCTCTGCAGGAAACACCGGCACGTGCGGAGAGGG 428
GL50581_2954 ----------
GL50803_101074 ----------AGAAGCTCAATG--CGCAGATAACACT---------- 99
GL50803_137617 ----------ACAGGCTGAATG--CGCAGATAAAACA---------- 99
GL50803_14586 ----------GCGTGTGAAGAGGACAAAACCTCAACT---------- 120
GL50581_1 ----------GGGAAATGGAAG--CGGCAGATGCAAG---------- 90
GL50581_3401 ----------GCAAATTGTGAGGCCGGCCAGTGCG---------- 94
GL50803_116477 ----------GCGACCTGCGCACAGAATAAGTGCG---------- 97
GL50581_3818 ----------GCGGTTGGCTCTCCCCACCGTGGCAGC---------- 111
GL50581_3457 ----------GGAGTTGGAGCCGGAGCTTACTGCACA---------- 117
GL50803_113319 ----------GGCGTCGGGGGCGGGGCATACTGCACG---------- 5502
GL50803_137740 ----------G----------TTCCTATCG---------- 82
GL50803_112135 ----------GCAGATGGTACACCCATTGTAAGCGAA---------- 228
GL50581_572 ----------GGTGGCGGTGAATTAGATCCAAACGCA---------- 132
GL50803_33279 ----------CAAGGCGACGTGCGAGGCGTGCAAGCC---------- 128
GL50803_137613 ----------GGCAAAG---------- 45
GL50581_4057 ----------GTGGAAAGTGCATTAAAAAGACATG---------- 92
GL50803_10330 ----------ACGATCAATGCATCAAGAAGACGTG---------- 92
GL50803_15250 ----------CGGGCAAGTGCGTTCCATGCCCAAAGGGTACATGGGGTCCCAGGTGT 192
GL50581_4316 ----------AAGGTGAATGCGTCCCCAGTTCTTGTCT---------- 101
GL50803_8687 ----------AAAATTCTTGCATTCCCAGCTCTTGTCT---------- 98
GL50803_14573 ----------CTGACCGGTGTGTACCACAGTCGTGTAT---------- 98
GL50803_16936 TCCGTTGAGTCCGAAGCAGACAAAGGAAAATGCAATCCAAGATCCTGTCAGGATAGCGAC 588
GL50581_3575 ----------GCGAAAACTGTTATCCAGAGAAGTGTGT---------- 95
GL50803_16477 ----------GTGAAAATTGCTATCCAGAGCACTGTGT---------- 95
GL50581_1982 ----------GCGGCGTTTGCTACCCAGAAACATGCGT---------- 95
GL50803_95162 ----------ACGGAACCTGTTATCCGGAGACTTGCGT---------- 95
GL50803_114815 ----------ACGGGGAGTGCTACCCGGAGGCGTGCGT---------- 152
b *
c *
b
c
b
c
b
a *
b
a *
b
*
a *
GL50581_727 ----------
GL50803_7715 ----------
GL50581_2622 ----------
GL50803_12063 ----------
GL50803_103454 ----------
GL50581_1902 CTGTGTGGCAGTCGTCCGAGAGTGGG---------- 453
GL50581_2954 ----------
GL50803_101074 ----------
GL50803_137617 ----------
GL50803_14586 ----------
GL50581_1 ----------
GL50581_3401 ----------
GL50803_116477 ----------
GL50581_3818 ----------
GL50581_3457 ----------
GL50803_113319 ----------
GL50803_137740 ----------
GL50803_112135 ----------
GL50581_572 ----------
GL50803_33279 ----------
GL50803_137613 ----------
GL50581_4057 ----------
GL50803_10330 ----------
GL50803_15250 TCCACACGC---------- 201
GL50581_4316 ----------
GL50803_8687 ----------
GL50803_14573 ----------
GL50803_16936 GGGCTTGTCTGCTCTGGCAAGGGAAAGTGCGAGTTGGACTTTGGCTACAACCTCACTCAC 648
GL50581_3575 ----------
GL50803_16477 ----------
GL50581_1982 ----------
GL50803_95162 ----------
GL50803_114815 ----------
b *
c *
b
c
b
c
b
a *
b
a *
b
*
a *
GL50581_727 ----------GATTGTCAGGAGG----------G 137
GL50803_7715 ----------GACTGCCAGCAGA----------T 137
GL50581_2622 ----------CC----------T 80
GL50803_12063 ----------CT----------T 80
GL50803_103454 ----------ACATGCGCGACCA----------A 101
GL50581_1902 ----------GCTGGCAAGTGCG----------A 467
GL50581_2954 ----------
GL50803_101074 ----------AATTGCGCAGCAG----------A 113
GL50803_137617 ----------AATGACTGT---G----------A 110
GL50803_14586 ----------GATGGAAAGTGTA----------A 134
GL50581_1 ----------GCCGGTATGTGCA----------A 104
GL50581_3401 ----------ACACAGTCG----------G 104
GL50803_116477 ----------AGACGGTCG----------G 107
GL50581_3818 ----------GGAGGACGGGCG----------G 124
GL50581_3457 ----------AAGACATGCA----------A 128
GL50803_113319 ----------AAGCCCTGCA----------A 5513
GL50803_137740 ----------AAGCCAG----------G 90
GL50803_112135 ----------AAGAGGTGCG----------G 239
GL50581_572 ----------CAGACATGCG----------G 143
GL50803_33279 ----------GGGATTCGTG----------G 139
GL50803_137613 ----------G 46
GL50581_4057 ----------
GL50803_10330 ----------
GL50803_15250 ----------AAGTGCTGGCTT---------- 213
GL50581_4316 ----------TTTTGATGGA---------- 111
GL50803_8687 ----------CTTTAATGGA---------- 108
GL50803_14573 ----------CAACAATAAG---------- 108
GL50803_16936 TATAGCTACCAGTGCATATGCGATGCTGGATATACGCGTGTTGGCCACTCTTGCATCCCG 708
GL50581_3575 ----------CTTTAATGGA---------- 105
GL50803_16477 ----------CTTTAGTGGA---------- 105
GL50581_1982 ----------TTTTGAGGAC---------- 105
GL50803_95162 ----------CTTTGAGGAC---------- 105
GL50803_114815 ----------CTTCGACGGA---------- 162


b * GL50581_727 AAAGTGCGTC----------CCTACAGGAAGGTACATTTCGGTCTGCTCCTTATGT 183
GL50803_7715 GAAGTGTCTT----------CCCACGGGGCGGTACATTACGGTTTGCTCCTTGTGT 183
c * GL50581_2622 CATGTGTATC----------TCAAATACTG---TCCCA---CTCTGCATCATGTGC 120
GL50803_12063 GATGTGTATC----------TCGAACACTG---TCCCC---CTCTGCATCATGTGC 120
b GL50803_103454 GAAGTGTATA----------ACGGTCAGCG---GCACCGAACTGTGCCGCGAGTGC 144
GL50581_1902 GACGTGCGGC----------CTCACCGTCGAGGGGACGAAGTACTGCTCGAAGTGC 513
GL50581_2954 ----------
GL50803_101074 AGCATGCAAC----------GTTCTGATCGGTGGTAAATTGTATTGCTCTCGATGT 159
GL50803_137617 AATCTGCGAC----------GTTCTAATTGATGGTAAATTGTATTGCTCTCGATGT 156
c GL50803_14586 AACATGCGGG----------GTCACCATCGGCCAAGACACTTGGTGCTCAGTTTGC 180
GL50581_1 CGTTCAACTC----------GGCGAAGGAGATGACGGTAAATACTGCTCTGAATGT 150
GL50581_3401 CGGGACCGAC----------AT---CTGCATGCAATGCAAGGCAGG--GAAAGT-- 143
GL50803_116477 GGCCACCGAC----------AT---TTGCACCCAGTGCAGGGCCGG--GGGCGT-- 146
GL50581_3818 ATAATATAAG----------CTCCCCTGCTCGCAGTCTCTCGCCCGCCGTGGGTGG 170
GL50581_3457 GAGGATCGGC----------GAAGGCGCCTGCAGCTGGAACTCCTG--CACGGTCA 172
b GL50803_113319 AAGAGTGGGG----------CAGGACGCTTGCAGTTGGGGCTCCTG--CACAGTCA 5557
c GL50803_137740 GGGAATCG----------TGCCATCAGCTGG-CCTCCGG--TGGG---- 122
b GL50803_112135 GATGTGTACA----------GACTCCATGTATCTATACAAAGGAGG--CTGCTACG 283
GL50581_572 CCAGTGCACA----------GACGCGTACTTCCTCTACAAGGGCGG--GTGCTATC 187
c GL50803_33279 CCAGCGGGGA----------TAACAGGGCGTGCAACCCAGCGCCGGC-CGACTGCC 184
GL50803_137613 ACAGCGGGAA----------GGACA---CATGCTTCCCG----------GGCTACA 79
GL50581_4057 --CATTGGC----------TTCAATGGACTCGAGTG------CACA---------- 120
a * GL50803_10330 --CGTCGGA----------TTCAACCAGCTCGAATG------CATG---------- 120
b GL50803_15250 ACCACACGT----------GTCAATTGTCTAGAATGTAGTAGCACAGACCGGTGC 258
GL50581_4316 ACGCTTTGT----------GGAGGTCATGGAACCTGTGATGACTATGGAGTTTGT 156
a * GL50803_8687 ACAATTTGT----------GGTGGTCATGGAACCTGTGACGATTACGGAGTTTGT 153
GL50803_14573 ACAGTTTGT----------GGGGCCCATGGGGTCTGTGACAAACACGGCATCTGT 153
b GL50803_16936 TCAGCTTGTGTAGTGTACCTAAGCGGTACCCCTGTGACCTGTGGGGGCTTCGGTGAGTGT 768
GL50581_3575 ACAGTCTGC----------TCTGATAAAGGAGATTGCATTAAGTCCGAG---TGC 147
* GL50803_16477 ACAGTCTGC----------TCTGGCAGGGGAGACTGCATCGGGTCCAAG---TGT 147
a * GL50581_1982 GCAGTCTGC----------AACGGTCATGGCACATGTGTGCGGGGCACG---TGC 147
GL50803_95162 GCCGTCTGC----------AACGGTCATGGCTCGTGTAGAGAGGGCACG---TGC 147
GL50803_114815 ACAGTCTGC----------GCGGGGCACGGGGTCTGCATGGTGAACTAT---TGC 204

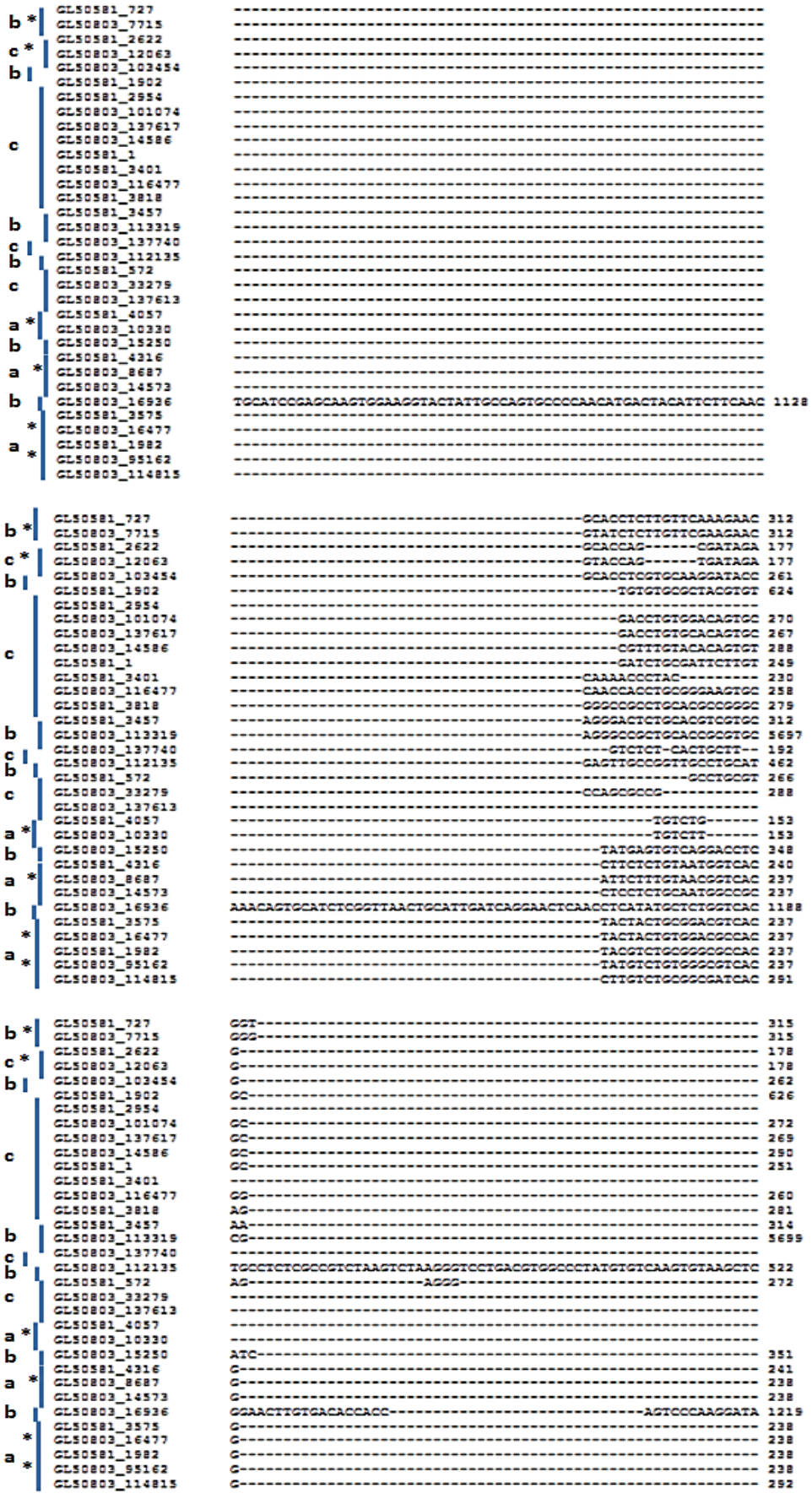
GL50581_727 AAG----------CCGGGAAAGATC---------- 198
b * GL50803_7715 AGG----------ACAGGATTGGTC---------- 198
GL50581_2622 GAT----------AGA---------- 126
c * GL50803_12063 GAT----------AGA---------- 126
b GL50803_103454 GTT----------GGCGGAAAAGTC---------- 159
GL50581_1902 AAA----------ATAGGCACTGAA---------- 528
GL50581_2954 ----------
GL50803_101074 AAC----------ACAGG---ATTT---------- 171
GL50803_137617 AAC----------ACAGG---ATTT---------- 168
c GL50803_14586 ATA----------GGAGAAAACTAT---------- 195
GL50581_1 CAT----------GCAACCGGAGAT---------- 165
GL50581_3401 ----------CCCGATAAAGGG---------- 155
GL50803_116477 ----------CCCGGTCGACGG---------- 158
GL50581_3818 TAG----------ACCCTTCGGTGG---------- 185
GL50581_3457 TGG----------ACGGAACCATGT---------- 187
b GL50803_113319 TGG----------ACGGGTCTCTGT---------- 5572
GL50803_137740 ----------TGT---------- 125
c GL50803_112135 CGT----------GCGGTAATCTTT---------- 298
b GL50581_572 AAA----------TCGG---ACAAT---------- 199
c GL50803_33279 CGA----------TCGAGAACTGCC---------- 199
GL50803_137613 CCC----------TCAA---------- 86
GL50581_4057 ----------
a * GL50803_10330 ----------
b GL50803_15250 TGGA----------CTTGTGAAGACGG---------- 275
GL50581_4316 AAT----------TGCGATAAAGG---------- 170
a * GL50803_8687 AAT----------TGCGATGAAAA---------- 167
GL50803_14573 AGG----------TACAAAGACAG---------- 167
b GL50803_16936 AAGGAGAACCCCGAGGGTTCAAATAGCTTCTCCTGTGCCTGCTATGAGGGCACTGTAAAA 828
GL50581_3575 CAC----------TGTGACAAGGG---------- 161
* GL50803_16477 CAC----------TGTGACAGAGG---------- 161
a * GL50581_1982 TCT----------TGCGAGATGAA---------- 161
GL50803_95162 AAC----------TGCGAGGATGA---------- 161
GL50803_114815 TAT----------TGTGACGACGG---------- 218

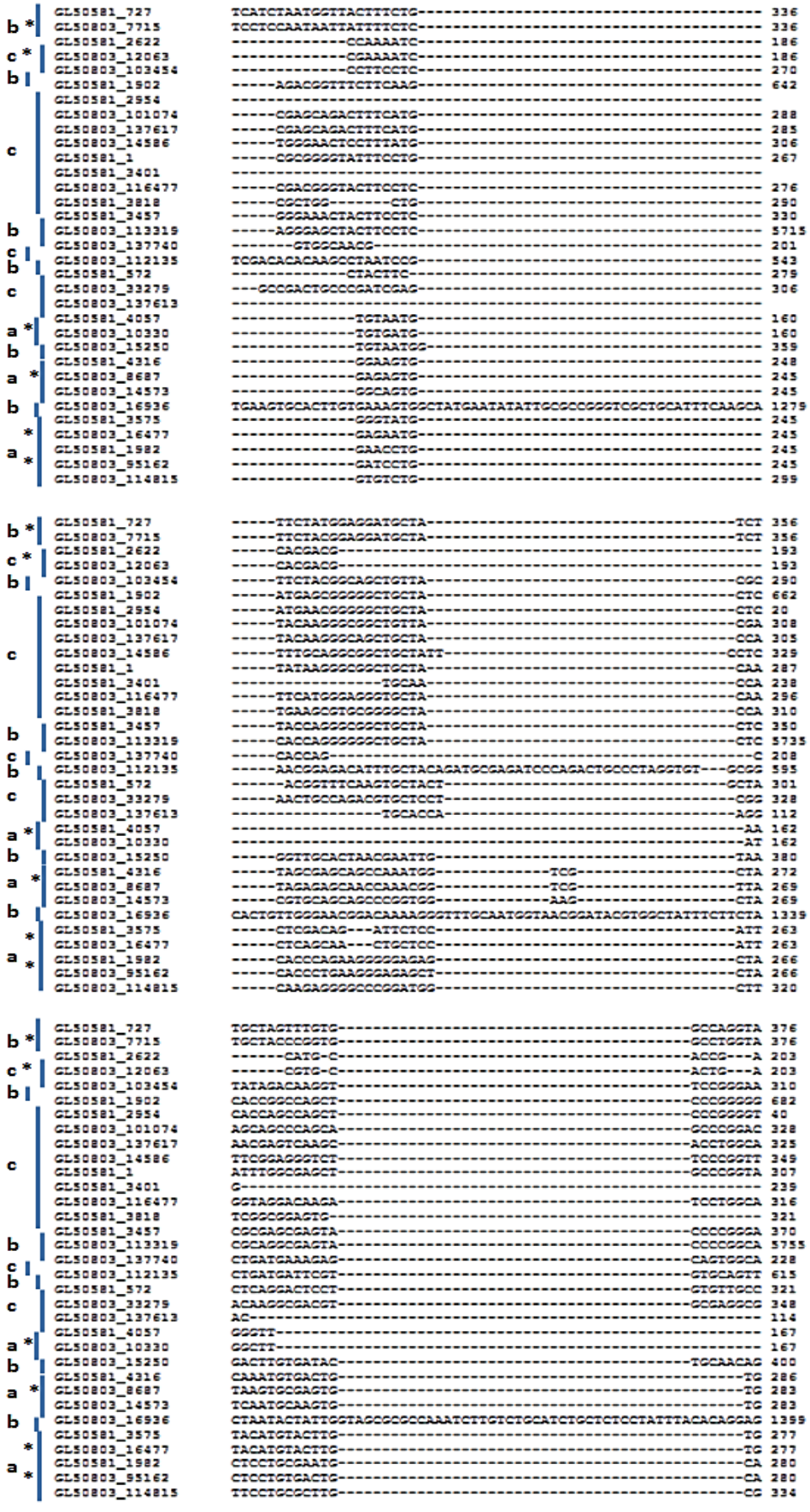
b * GL50581_727 ----------CCCCTGAACGGAG---------- 211
GL50803_7715 ----------CCCATAAATGGGG---------- 211
c * GL50581_2622 ----------CTTCGCCGTCC---------- 137
GL50803_12063 ----------CTTCGCCGTCC---------- 137
b GL50803_103454 ----------CCTATCGACGGGA---------- 172
GL50581_1902 ----------TATCGGCAGAACGGT---------- 543
GL50581_2954 ----------
GL50803_101074 ----------GTTCCTATCAATGGA---------- 186
GL50803_137617 ----------GTTCCTATCAATGGA---------- 183
c GL50803_14586 ----------GCGCCCATCAACGGA---------- 210
GL50581_1 ----------TTTATTATCAACGGT---------- 180
GL50581_3401 ----------GATCTGCAAAGCA---------- 168
GL50803_116477 ----------CTTCTGCTGGCCC---------- 171
GL50581_3818 ----------TAGCT-CTGTCTG---------- 197
GL50581_3457 ----------GCTCGGAGTGTGCGTCT---------- 204
b GL50803_113319 ----------GCTCGCAGTGCGCGTCG---------- 5589
c GL50803_137740 ----------GCTCTTAT---------- 133
b GL50803_112135 ----------ACCCCGGGGAACCTTATATGTAGCGAGTGCGATAAAGGAGTTTGCA 343
GL50581_572 ----------CCCCCGG---------- 206
c GL50803_33279 ----------AGACGTG---------- 206
GL50803_137613 ----------
GL50581_4057 ----------
a * GL50803_10330 ----------
b GL50803_15250 ----------GTACTATGTGACGT---------- 289
GL50581_4316 ----------TTATTCCCCGACT---------- 183
a * GL50803_8687 ----------GTATTCTCCAACT---------- 180
GL50803_14573 ----------GCAAGCCGTTCCA---------- 180
b GL50803_16936 GTAGGAAATGGATGCACATACCGCACCTGCACCGAAAGCGAAAGTGCAAGCAAGATATGC 888
GL50581_3575 ----------CTACGCCTTATCAG---------- 175
* GL50803_16477 ----------ATACGCCTTATCGG---------- 175
a * GL50581_1982 ----------CTATAGGTTGACAG---------- 175
GL50803_95162 ----------CTATTTGCTGACGG---------- 175
GL50803_114815 ----------CTACAGACTGTCGA---------- 232


GL50581_727 ----------TGTGTGTCAACGAATCCACAGCATCTGACTACGGGGATATA 252
b * GL50803_7715 ----------TGTGCGTCAATCAAACTACTGCGCATGACTACGGAGACATA 252
GL50581_2622 ----------TTCG----------CTTGGTCCGGTAG---- 154
c * GL50803_12063 ----------TTCG----------CTTGGTCCCGTGG---- 154
b GL50803_103454 ----------TTTGT---------AAGGACACTACTGATTCCAGTGTTAGT 204
GL50581_1902 ----------GTGTGCGCTACGAAGCCAACCGGTCGCGCCACGGCTAG---- 581
GL50581_2954 ----------
GL50803_101074 ----------CAATGTGCAGACAAAGAAGGTGCAACAGATAAGTGCAAA--- 225
GL50803_137617 ----------AAATGTGCAGACAAAGAAGGTGCAAAAGCCAACTGTAAA--- 222
GL50803_14586 ----------ACATGCGCCGATATAACAACTGACTCTGTGGGCT--------- 244
c GL50581_1 ----------GTTTGT-------TTGCAG--GCCAACGACAATAG-------- 206
GL50581_3401 ----------CACG---------ATACCGACGCTGGCGCGATCACT------ 195
GL50803_116477 ----------CCCGGCTCCCCCCAGGCCGCGGCAGCCGGGTGCACCGAG--- 210
GL50581_3818 ----------CGGGGCG-----ATGGCCACAATGTTCGAG-GCAATTAT--- 230
GL50581_3457 GGCGATTACGTGCCGATTGACGGTGCCTGCACCCCGTACTCCGGAAGGCCCCGCTCG--- 261
b GL50803_113319 AGTGATTATGTGCCGATCGACGGCGTCTGCACACTCTATCCTGGAGAAGCTCGCTCA--- 5646
c GL50803_137740 ----------CTCGTTCTTCTCTAGTACAACTG-------- 156
b GL50803_112135 AGTCTTGTACGTCTGGTTTCTTTAGGAACCCTGCAGCGT-CGCAAACAGTCGATTCGTGT 402
GL50581_572 ----------TAGTTTGATTTG------TGCTGA-----CAAACAGGCAT-------- 235
c GL50803_33279 ----------CTCCTCGGACAAGGCGACGTGCGAGGCGTGCAAGCCCGGATTC----- 249
GL50803_137613 ----------CACTGACACG----------AAGCAG---------- 102
GL50581_4057 ----------GGTTAT------ 126
a * GL50803_10330 ----------GGATAC------ 126
b GL50803_15250 ----------GGTA----------TGGGGGCTGT------ 303
GL50581_4316 ----------CACA----------GTGGTTGCCAG------ 198
a * GL50803_8687 ----------CACA----------GTGGCTGCCAG------ 195
GL50803_14573 ----------CGTG----------TCGTGAAGGAG------ 195
b GL50803_16936 GGAGGCATCGGTGTCTGTGTGCGTGATGACACAGAATACAAGTGTGACTGCAAAGGGTTT 948
GL50581_3575 ----------AGCA----------AGGCTGCCAG------ 189
* GL50803_16477 ----------AGCA----------AGGCTGCCAG------ 189
a * GL50581_1982 ----------AACT----------CGGGTGCTAT------ 189
GL50803_95162 ----------AGCT----------TGGGTGCTAC------ 189
GL50803_114815 ----------ACGC----------CGGGTGCACG------ 246

b * GL50581_727 AATGTCTGC---------- 261
GL50803_7715 AGTGTCTGT---------- 261
c * GL50581_2622 ----------
GL50803_12063 ----------
b GL50803_103454 TTGTATGGA---------- 213
GL50581_1902 ----------
GL50581_2954 ----------
GL50803_101074 ----------
GL50803_137617 ----------
c GL50803_14586 ----------
GL50581_1 ----------
GL50581_3401 ----------
GL50803_116477 ----------
GL50581_3818 ----------
b GL50581_3457 ----------
GL50803_113319 ----------
c GL50803_137740 ----------
b GL50803_112135 GTTCGGTGT---------- 411
GL50581_572 ----------
c GL50803_33279 ----------
GL50803_137613 ----------
a * GL50581_4057 ----------
GL50803_10330 ----------
b GL50803_15250 ----------
GL50581_4316 ----------
a * GL50803_8687 ----------
GL50803_14573 ----------
b GL50803_16936 GCTACCGGCAGCTTATGCGATAGCTGCGTTTCTGAAAATGCAACCTCTGTGGGCAATCAA 1008
GL50581_3575 ----------
* GL50803_16477 ----------
a * GL50581_1982 ----------
GL50803_95162 ----------
GL50803_114815 ----------

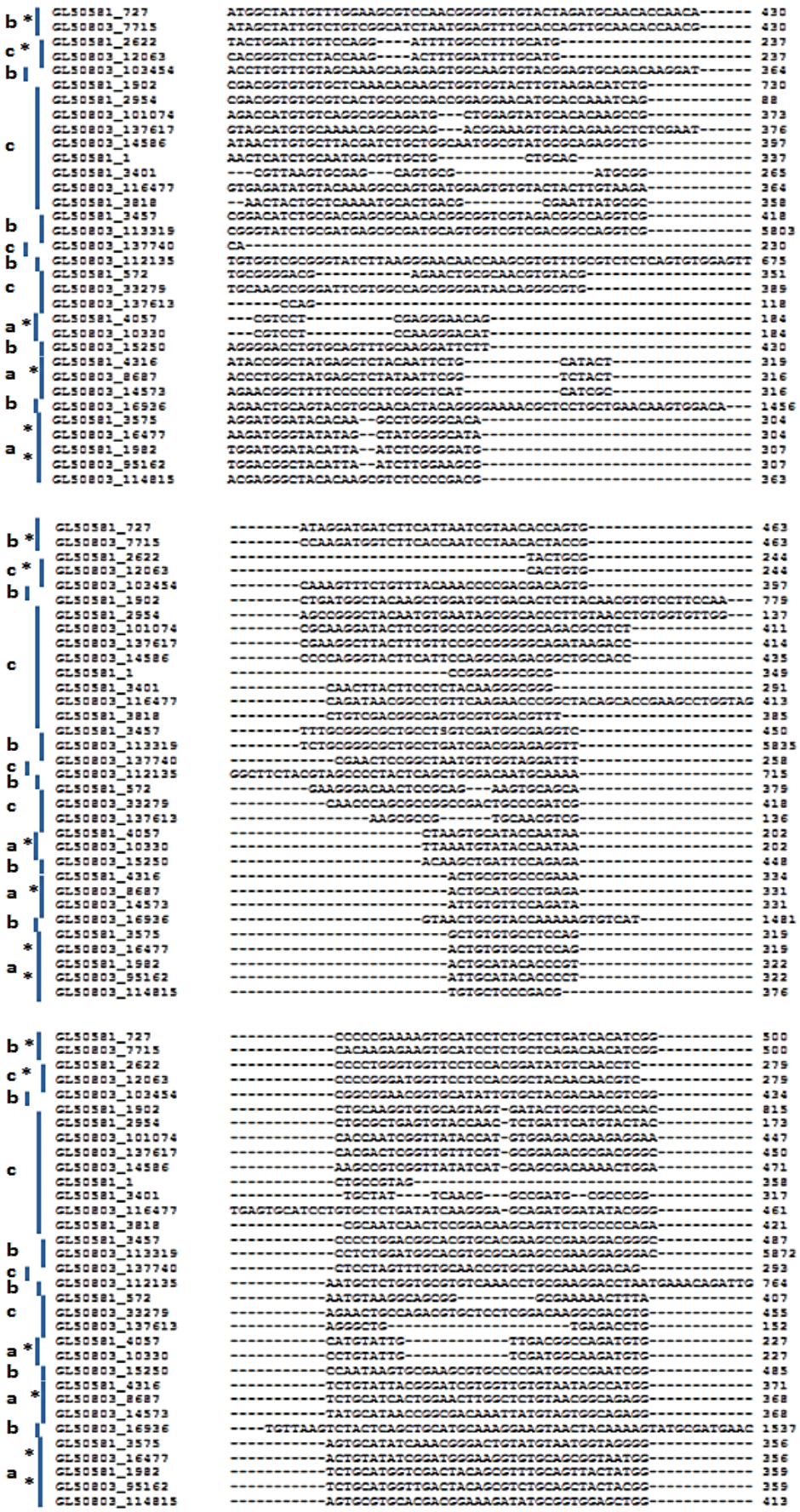
b * GL50581_727 ------CAGAAAAATGATGAGAGCAATCCTACTCACT---------- 292
GL50803_7715 ------CAAAAGAATGATGAAACAAATCCCACCCATT---------- 292
c * GL50581_2622 ----------ATGGTGTTT---------- 163
GL50803_12063 ----------ACGGCGTTT---------- 163
b GL50803_103454 ------TGCGAGCATTCT---AGTGGCACTGGATTTT---------- 241
GL50581_1902 --------TTGCAAGGATAGTAATATTGCTAATGG---------- 608
GL50581_2954 ----------
GL50803_101074 ------GATGGCTCCGGAGGCGATACAGCTGATCA---------- 254
GL50803_137617 ------GGTGCCGATGGAGGCGATACAGCTGATCG---------- 251
c GL50803_14586 -------CTGTCTTTTGCACCCAAAAGCCCGCCG---------- 272
GL50581_1 --------CGCATGTAAAGCAAACGCAGACAAAGG---------- 233
GL50581_3401 -------CCTACGGG-CTGCAAAG--TAAGTGGG---------- 219
GL50803_116477 ------GAAGACGGGGCTACCCTCGACAAGACGG---------- 238
GL50581_3818 ------CCTAGCAAG-CTGCATACTACAGTTCGCGT---------- 259
GL50581_3457 ------CTCGCCAGCCCGTGTGTGAAGGACAGCAGGA---------- 292
b GL50803_113319 ------CTTGCCGGTCCATGCGTAAAGGACGCCAAGA---------- 5677
GL50803_137740 ----------GTCAAGGAGAGTGAGAGTGACT---------- 178
c GL50803_112135 ------GATGACACAACAGGAGTGGCGGACTTTGTAG---------- 442
b GL50581_572 -------CTGGCACGACAG--------GAACATGTAAG---------- 258
c GL50803_33279 --------GTGGCCAGCGGGGATAACAGGCGTGCAAC---------- 279
GL50803_137613 ----------
GL50581_4057 ----------GGTAAGTGCACGAATGGCATC---------- 147
a * GL50803_10330 ----------GGCAAATGCATGAATGGCATC---------- 147
b GL50803_15250 ----------GGTGCGTGTGCAGAAGGCACCGGAGGC---------- 330
GL50581_4316 ------CCGGACGCCTGTGTAGACGGAACA---------- 222
a * GL50803_8687 ------CCGGCTGCATGTGTAGACGGGACG---------- 219
GL50803_14573 ------CCCCAAGAGTGCATGGACGGCGAG---------- 219
b GL50803_16936 TGTATACCAAATGAGTGCCTCGACGAAGCAGGCACGACGGTTTGTAAGGGCAATGGAAAG 1068
GL50581_3575 ------CCGAAGGATTGTGTCAACAGAGGAACAAAA---------- 219
* GL50803_16477 ------CCAGAAGATTGTGTCAATAGAGCAACAAGA---------- 219
a * GL50581_1982 ------CCCACTAGCTGCTCCGATTCGCGCACAGGA---------- 219
GL50803_95162 ------CCTAGTGGTTGCTCAGATAGACGCACACAA---------- 219
GL50803_114815 ------CCTGCGGAGTGTACGGATAG---CATGGGG---------- 273

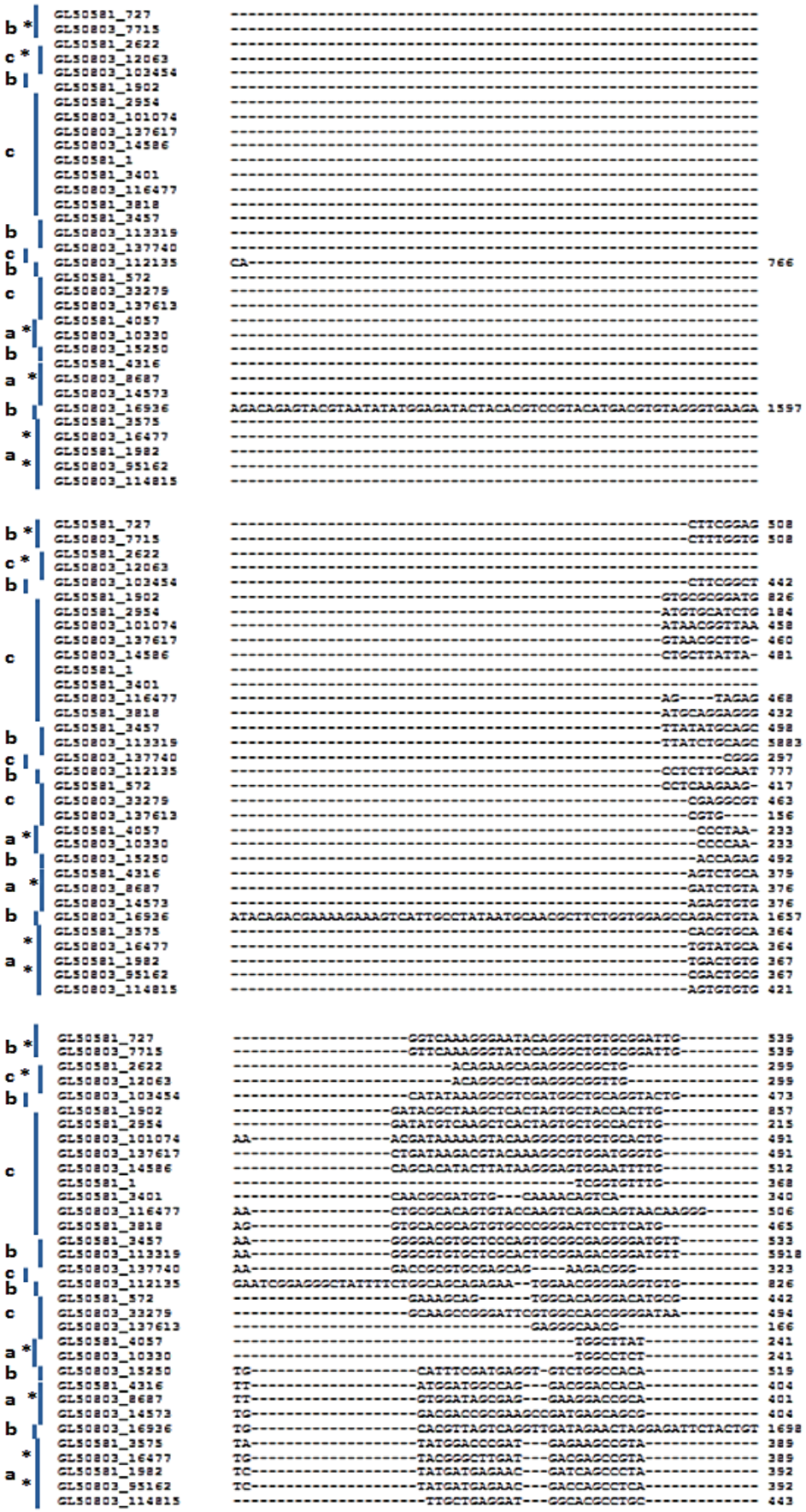

b * GL50581_727
GL50803_7715
c * GL50581_2622
GL50803_12063
b GL50803_103454
GL50581_1902
GL50581_2954
GL50803_101074
GL50803_137617
c GL50803_14586
GL50581_1
GL50581_3401
GL50803_116477
GL50581_3818
GL50581_3457
b GL50803_113319
c GL50803_137740
b GL50803_112135
GL50581_572
c GL50803_33279
GL50803_137613
GL50581_4057
a * GL50803_10330
b GL50803_15250
GL50581_4316
a * GL50803_8687
GL50803_14573
b GL50803_16936
GL50581_3575
* GL50803_16477
a * GL50581_1982
GL50803_95162
GL50803_114815

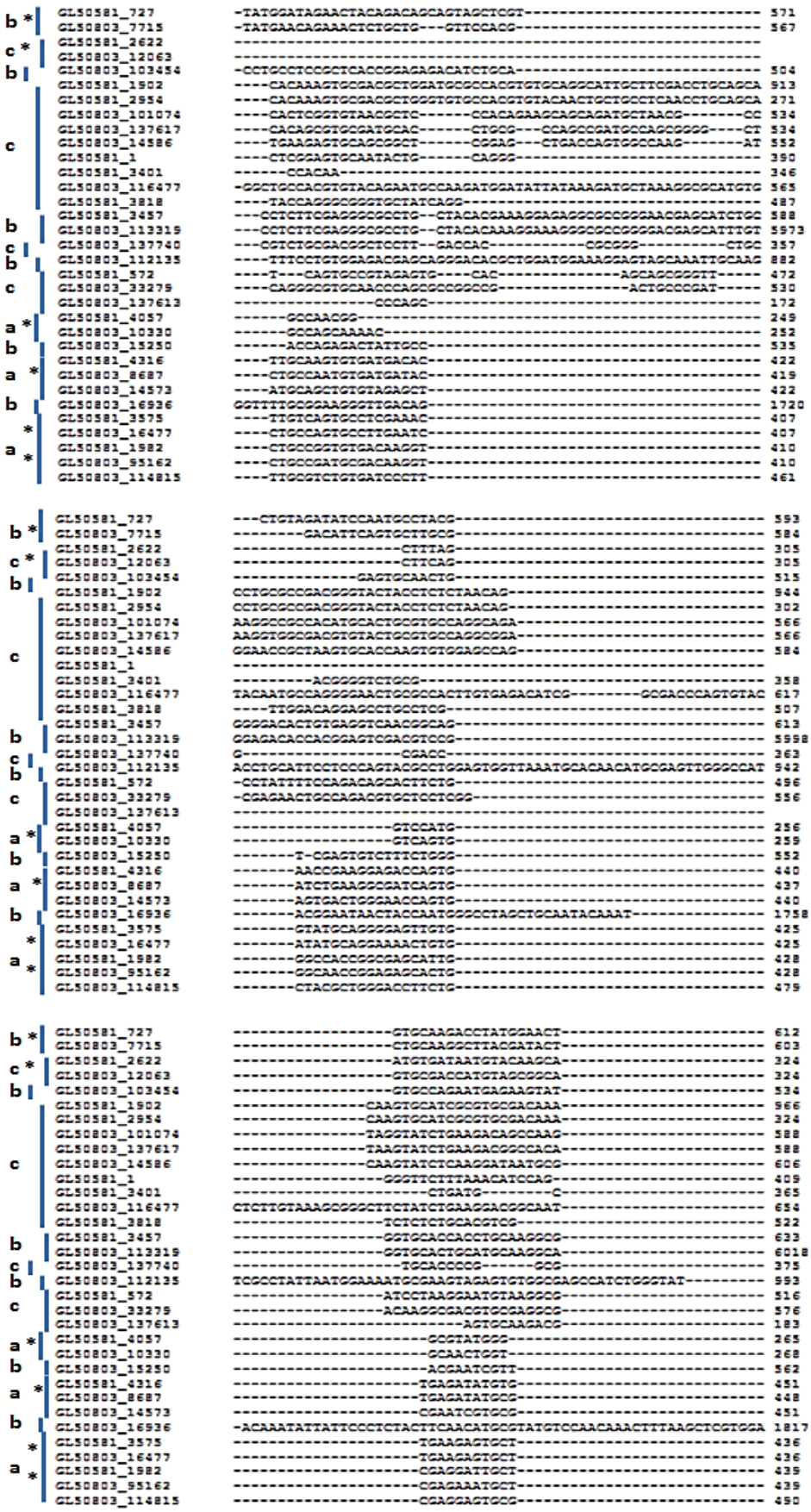

b * GL50581_727 TCATCTAATGGTTACTTTCTG 336
b * GL50803_7715 TCCTCCAATAATTATTTTCTC 336
GL50581_2622 CCAAAATC 186
c * GL50803_12063 CGAAAATC 186
b GL50803_103454 CCTTCCTC 270
GL50581_1902 AGACGGTTTCTTCAAG 642
GL50581_2954
GL50803_101074 CGAGCAGACTTTCATG 288
GL50803_137617 CGAGCAGACTTTCATG 285
c GL50803_14586 TGGGAACTCCTTTATG 306
GL50581_1 CGCGGGGTATTTCCTG 267
GL50581_3401
GL50803_116477 CGACGGGTACTTCCTC 276
GL50581_3818 CGCTGG-------CTG 290
GL50581_3457 GGGAAACTACTTCCTC 330
b GL50803_113319 AGGGAGCTACTTCCTC 5715
c GL50803_137740 GTGGCAACG 201
b GL50803_112135 TCGACACACAAGCCTAATCCG 543
GL50581_572 CTACTTC 279
c GL50803_33279 GCCGACTGCCCGATCGAG 306
GL50803_137613
GL50581_4057 TGTAATG 160
a * GL50803_10330 TGTGATG 160
b GL50803_15250 TGTAATGG 359
GL50581_4316 GGAAGTG 248
a * GL50803_8687 GAGAGTG 245
GL50803_14573 GGCAGTG 245
b GL50803_16936 TGAAGTGCACTTGTGAAAGTGGCTATGAATATATTGCGCCGGGTCGCTGCATTTCAAGCA 1279
GL50581_3575 GGGTATG 245
* GL50803_16477 GAGAATG 245
a * GL50581_1982 GAACCTG 245
GL50803_95162 GATCCTG 245
GL50803_114815 GTGTCTG 299

b * GL50581_727 TTCTATGGAGGATGCTA TCT 356
GL50803_7715 TTCTACGGAGGATGCTA TCT 356
c * GL50581_2622 CACGACG 193
GL50803_12063 CACGACG 193
b GL50803_103454 TTCTACGGCAGCTGTTA CGC 290
GL50581_1902 ATGAGCGGGGGCTGCTA CTC 662
GL50581_2954 ATGAACGGGGGCTGCTA CTC 20
GL50803_101074 TACAAGGGCGGCTGTTA CGA 308
GL50803_137617 TACAAGGGCAGCTGCTA CCA 305
c GL50803_14586 TTTGCAGGCGGCTGCTATT CCTC 329
GL50581_1 TATAAGGGCGGCTGCTA CAA 287
GL50581_3401 TGCAA CCA 238
GL50803_116477 TTCATGGGAGGGTGCTA CAA 296
GL50803_3818 TGAAGCGTGCGGGGCTA CCA 310
GL50581_3457 TACCAGGCCGGCTGCTA CTC 350
b GL50803_113319 CACCAGGGGGGCTGCTA CTC 5735
c GL50803_137740 CACCAG C 208
b GL50803_112135 AACGGAGACATTTGCTACAGATGCGAGATCCCAGACTGCCCTAGGTGT GCGG 595
GL50581_572 ACGGTTTCAAGTGCTACT GCTA 301
c GL50803_33279 AACTGCCAGACGTGCTCCT CGG 328
GL50803_137613 TGCACCA AGG 112
a * GL50581_4057 AA 162
GL50803_10330 AT 162
b GL50803_15250 GGTTGCACTAACGAATTG TAA 380
a * GL50581_4316 TAGCGAGCAGCCAAATGG TCG CTA 272
GL50803_8687 TAGAGAGCAACCAAACGG TCG TTA 269
GL50803_14573 CGTGCAGCAGCCCGGTGG AAG CTA 269
b GL50803_16936 CACTGTTGGGAACGGACAAAAGGGTTTGCAATGGTAACGGATACGTGGCTATTTCTTCTA 1339
GL50581_3575 CTCGACAG---ATTCTCC ATT 263
* GL50803_16477 CTCAGCAA---CTGCTCC ATT 263
a * GL50581_1982 CACCCAGAAGGGGGAGAG CTA 266
GL50803_95162 CACCCTGAAGGGAGAGCT CTA 266
GL50803_114815 CAAGAGGGGCCCGGATGG CTT 320

b * GL50581_727 TGCTAGTTTGTG GCCAGGTA 376
GL50803_7715 TGCTACCCGGTG GCCTGGTA 376
GL50581_2622 CATG-C ACCG---A 203
c * GL50803_12063 CGTG-C ACTG---A 203
b GL50803_103454 TATAGACAAGGT TCCGGGAA 310
GL50581_1902 CACCGGCCAGCT CCCGGGGG 682
GL50581_2954 CACCAGCCAGCT CCCGGGGT 40
GL50803_101074 AGCAGCCCAGCA GCCCGGAC 328
GL50803_137617 AACGAGTCAAGC ACCTGGCA 325
c GL50803_14586 TTCGGAGGGTCT TCCCGGTT 349
GL50581_1 ATTTGGCGAGCT GCCCGGTA 307
GL50581_3401 G 239
GL50803_116477 GGTAGGACAAGA TCCTGGCA 316
GL50581_3818 TCGGCGGAGTG 321
GL50581_3457 CGCGAGCGAGTA CCCCGGGA 370
b GL50803_113319 CGCAGGCGAGTA CCCCGGCA 5755
c GL50803_137740 CTGATGAAAGAG CAGTGGCA 228
b GL50803_112135 CTGATGATTCGT GTGCAGTT 615
GL50581_572 CTCAGGACTCCT GTGTTGCC 321
c GL50803_33279 ACAAGGCGACGT GCGAGGCG 348
GL50803_137613 AC 114
GL50581_4057 GGGTT 167
a * GL50803_10330 GGCTT 167
b GL50803_15250 GACTTGTGATAC TGCAACAG 400
GL50581_4316 CAAATGTGACTG TG 286
a * GL50803_8687 TAAGTGCGAGTG TG 283
GL50803_14573 TCAATGCAAGTG TG 283
b GL50803_16936 CTAATACTATTGGTAGCGGCGCAAATCTTGTCTGCATCTGCTCTCCTATTTACACAGGAG 1399
GL50581_3575 TACATGTACTTG TG 277
* GL50803_16477 TACATGTACTTG TG 277
a * GL50581_1982 CTCCTGCGAATG CA 280
GL50803_95162 CTCCTGTGACTG CA 280
GL50803_114815 TTCCTGCGCTTG CG 334

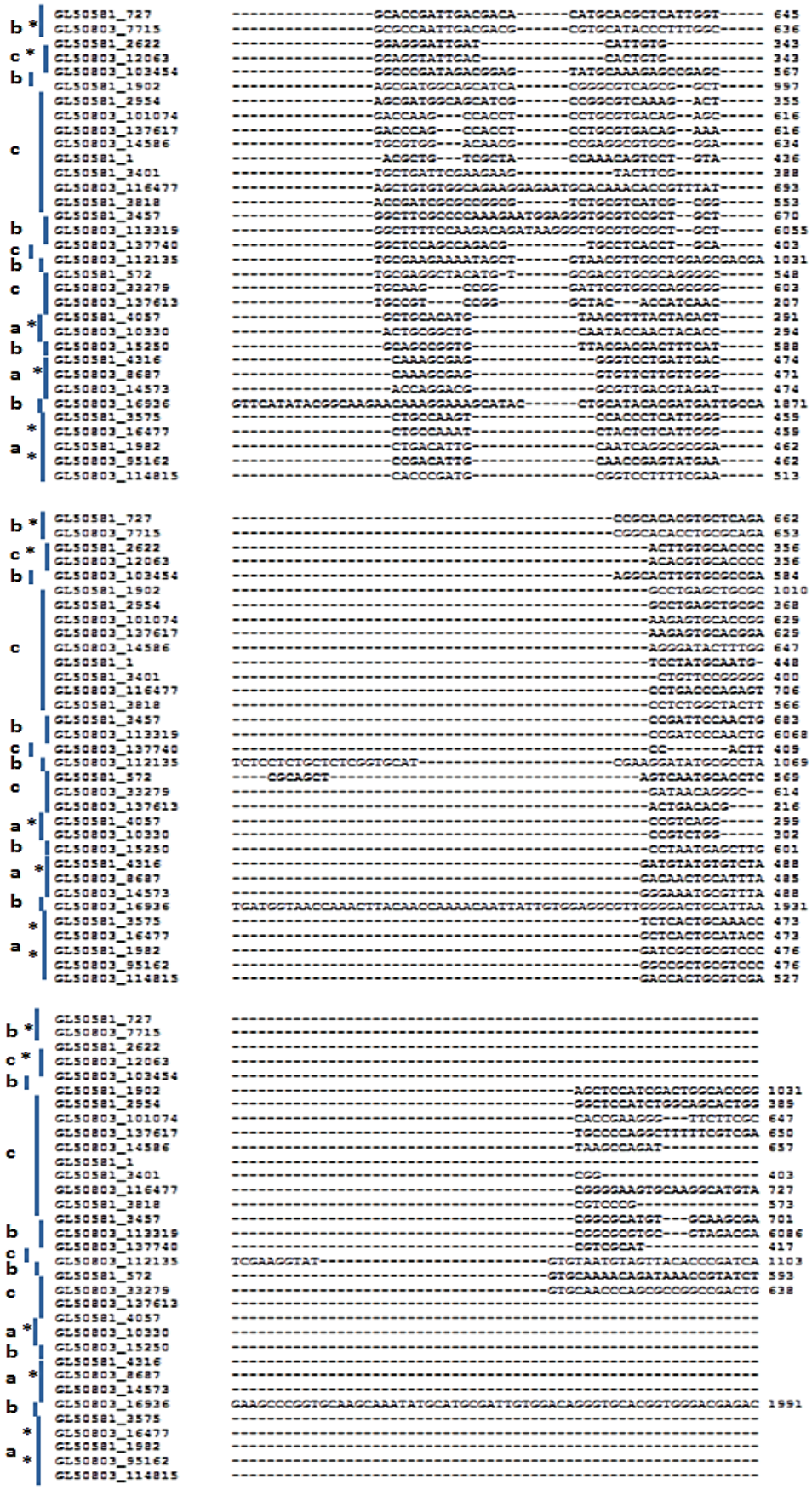

b *
GLS0581_727 ATGGCTATTGTTTGGAAGCGTCCAACGGGGTGTGTACTAGATGCAACACCAACA------ 430
GLS0803_7715 ATAGCTATTGTCTGTCGGCATCTAATGGAGTTTGCACCAGTTGCAACACCAACG------ 430
c *
GLS0581_2622 TACTGGATTGTTCCAGG----ATTTTGGCCTTTGCATG---------- 237
GLS0803_12063 CACGGGTCTCTACCAAG----ACTTTGGATTTTGCATG---------- 237
b
GLS0803_103454 ACCTTGTTTGTAGCAAAGCAGAGAGTGGCAAGTGTACGGAGTGCAGACAAGGAT------ 364
GLS0581_1902 CGACGGTGTGTGCTCAAACACAAGCTGGTGGTACTTGTAAGACATCTG---------- 730
GLS0581_2954 CGACGGTGTGCGTCACTGCGCCGACCGGAGGAACATGCACCAAATCAG---------- 88
GLS0803_101074 AGACCATGTGTCAGGCGGCAGATG---CTGGAGTATGCACACAAGCCG---------- 373
GLS0803_137617 GTAGCATGTGCAAAACAGCGGCAG---ACGGAAAGTGTACAGAAGCTCTCGAAT------ 376
c
GLS0803_14586 ATAACTTGTGCTTACGATCTGCTGGCAATGGCGTATGCGCAGAGGCTG---------- 397
GLS0581_1 AACTCATCTGCAATGACGTTGCTG----------CTGCAC---------- 337
GLS0581_3401 ---CGTTAAGTGCGAG---CAGTGCG----------ATGCGG---------- 265
GLS0803_116477 GTGAGATATGTACAAAGGCCAGTGATGGAGTGTGTACTACTTGTAAGA---------- 364
GLS0581_3818 --AACTACTGCTCAAAATGCACTGACG---------CGAATTATGCGC---------- 358
b
GLS0581_3457 CGGACATCTGCGACGAGCGCAACACGGCGGTCGTAGACGGCCAGGTCG---------- 418
GLS0803_113319 CGGGTATCTGCGATGAGCGCGATGCAGTGGTCGTCGACGGCCAGGTCG---------- 5803
c
GLS0803_137740 CA---------- 230
b
GLS0803_112135 TGTGGTCGCGGGTATCTTAAGGGAACAACCAAGCGTGTTTGCGTCTCTCAGTGTGGAGTT 675
GLS0581_572 TGCGGGGACG----------AGAACTGCGCAACGTGTACG---------- 351
c
GLS0803_33279 TGCAAGCCCGGGATTCGTGGCCAGCGGGGATAACAGGGCGTG---------- 389
GLS0803_137613 ------CCAG---------- 118
a *
GLS0581_4057 ---CGTCCT----------CGAGGGAACAG---------- 184
GLS0803_10330 ---CGTCCT----------CCAAGGGACAT---------- 184
b
GLS0803_15250 AGGGGACCTGTGCAGTTTGCAAGGATTCTT---------- 430
a *
GLS0581_4316 ATACCGGCTATGAGCTCTACAATTCTG----------CATACT---------- 319
GLS0803_8687 ACCCTGGCTATGAGCTCTATAATTCGG----------TCTACT---------- 316
GLS0803_14573 AGAACGGCTTTTCCCCCTTCGGCTCAT----------CATGGC---------- 316
b
GLS0803_16936 AGAACTGCAGTACGTGCAACACTACAGGGGAAAACGCTCCTGCTGAACAAGTGGACA--- 1456
GLS0581_3575 AGGATGGATACACAA--GCCTGGGGCACA---------- 304
*
GLS0803_16477 AAGATGGGTATATAG--CTATGGGGCATA---------- 304
a *
GLS0581_1982 TGGATGGATACATTA--ATCTCGGGGATG---------- 307
GLS0803_95162 TGGACGGCTACATTA--ATCTTGGAAGCG---------- 307
GLS0803_114815 ACGAGGGCTACACAAGCGTCTCCCCGACG---------- 363
GLS0581_727 --------ATAGGATGATCTTCATTAATCGTAACACCAGTG---------- 463
b *
GLS0803_7715 --------CCAAGATGGTCTTCACCAATCCTAACACTACCG---------- 463
GLS0581_2622 ----------TACTGCG---------- 244
c *
GLS0803_12063 ----------CACTGTG---------- 244
b
GLS0803_103454 --------CAAAGTTTCTGTTTACAAACCCCGACGACAGTG---------- 397
GLS0581_1902 --------CTGATGGCTACAAGCTGGATGCTGACACTCTTACAACGTGTCCTTCCAA--- 779
GLS0581_2954 --------AGCCGGGCTACAATGTGAATAGCGGCACCCTTGTAACCTGTGGTGTTGG--- 137
GLS0803_101074 --------CGCAAGGATACTTCGTGCCGCCGGGGCGCAGACGCCTCT---------- 411
GLS0803_137617 --------CGAAGGCTTACTTTGTTCCGCCGGGGGCAGATAAGACC---------- 414
GLS0803_14586 --------CCCCAGGGTACTTCATTCCAGCCGAGACGGCTGCCACC---------- 435
c
GLS0581_1 ----------CCGGAGGGCCGC---------- 349
GLS0581_3401 ----------CAACTTACTTCCTCTACAAGGGCGGG---------- 291
GLS0803_116477 ----------CAGATAACGGCCTGTTCAAGAACCCGGCTACAGCACCGAAGCCTGGTAG 413
GLS0581_3818 ----------CTGTCGACGGCGAGTGGGTGGACGTTT---------- 385
GLS0581_3457 --------TTTGCGGGGCGCTGCCTSGTCGATGGCGAGGTC---------- 450
b
GLS0803_113319 --------TCTGCGGGCGCTGCCTGATCGACGGAGAGGTT---------- 5835
c
GLS0803_137740 ----------CGAACTCCGGCTAATGTTGGTAGGATTT---------- 258
b
GLS0803_112135 GGCTTCTAGTAGCCCCTACTCAGCTGGGACAATGCAAAA---------- 715
GLS0581_572 ----------GAAGGGACAACTCCGCAG---AAGTGCAGCA---------- 379
c
GLS0803_33279 ----------CAACCCAGCGCCGGCCGACTGCCCGATCG---------- 418
GLS0803_137613 ----------AAGCGCCG------TGCAACGTCG---------- 136
GLS0581_4057 ----------CTAAGTGCATACCAATAA---------- 202
a *
GLS0803_10330 ----------TTAAATGTATACCAATAA---------- 202
b
GLS0803_15250 ----------ACAAGCTGATTCCAGAGA---------- 448
GLS0581_4316 ----------ACTGCGTGCCCGAAA---------- 334
a *
GLS0803_8687 ----------ACTGCATGCCTGAGA---------- 331
GLS0803_14573 ----------ATTGTGTTCCAGATA---------- 331
b
GLS0803_16936 ----------GTAACTGCGTACCAAAAAGTGTCAT---------- 1481
GLS0581_3575 ----------GCTGTGTGCCTCCAG---------- 319
*
GLS0803_16477 ----------ACTGTGTGCCTCCAG---------- 319
a *
GLS0581_1982 ----------ACTGCATACACCCGT---------- 322
GLS0803_95162 ----------ATTGCATACACCCCT---------- 322
GLS0803_114815 ----------TGTGCTCCCGACG---------- 376
GLS0581_727 ----------CCCCCGAAAAGTGCATCCTCTGCTCTGATCACATCGG---------- 500
b *
GLS0803_7715 ----------CACAAGAGAAGTGCATCCTCTGCTCAGACAACATCGG---------- 500
GLS0581_2622 ----------CCCCTGGGTGGTTCCTCCACGGATATGTCAACCTC---------- 279
c *
GLS0803_12063 ----------CCCCGGGATGGTTCCTCCACGGCTACAACAACGTC---------- 279
b
GLS0803_103454 ----------CGGCGGAACGGTGCATATTGTGCTACGACAACGTCGG---------- 434
GLS0581_1902 ----------CTGCAAGGTGTGCAGTAGT-GATACTGGGTGCACCAC---------- 815
GLS0581_2954 ----------CTGCGGCTGAGTGTACCAAC-TCTGATTCATGTACTAC---------- 173
GLS0803_101074 ----------CACCAATCGGTTATACCAT-GTGGAGACGAAGAGGAA---------- 447
GLS0803_137617 ----------CACGACTCGGTTGTTTCGT-GCGGAGACGCGACGGGC---------- 450
c
GLS0803_14586 ----------AAGCCGTCGGTTATATCAT-GCAGCGACAAAACTGGA---------- 471
GLS0581_1 ----------CTGCCGTAG---------- 358
GLS0581_3401 ----------TGCTAT----TCAACG---GCCGATG--CGCCCGG---------- 317
GLS0803_116477 TGAGTGCATCCTGTGCTCTGATATCAAGGGA-GCAGATGGATATACGGG---------- 461
GLS0581_3818 ----------CGCAATCAACTCCGGACAAGCAGTTCTGCCCCCAGA---------- 421
GLS0581_3457 ----------CCCCTGGACGGCACGTGCACGAAGCCGAAGGACGGGC---------- 487
b
GLS0803_113319 ----------CCTCTGGATGGCACGTGCGCAGAGCCGAAGGAGGAC---------- 5872
c
GLS0803_137740 ----------CTCCTAGTTTGTGCAACCGTGCTGGCAAAGGACAG---------- 293
b
GLS0803_112135 ----------AATGCTCTGGTGCGTGTCAAACCTGCGAAGGACCTAATGAAACAGATTG 764
GLS0581_572 ----------AATGTAAGGCAGCGG---------GCGAAAAACTTTA---------- 407
c
GLS0803_33279 ----------AGAACTGCCAGACGTGCTCCTCGGACAAGCCGACGTG---------- 455
GLS0803_137613 ----------AGGGCTG----------TGAGACCTG---------- 152
GLS0581_4057 ----------CATGTATTG----------TTGACGCCAGATGTG---------- 227
a *
GLS0803_10330 ----------CCTGTATTG----------TCGATGCCAAGATGTG---------- 227
b
GLS0803_15250 ----------CCAATAAGTGCGAAGCGTGCCCCCGATGCCGAATCGG---------- 485
GLS0581_4316 ----------TCTGTATTACGGGATCGTGGTTGTGTAATAGCCATGG---------- 371
a *
GLS0803_8687 ----------TCTGCATCACTGGAACTTGGCTCTGTAACGGCAGAGG---------- 368
GLS0803_14573 ----------TATGCATAACCGGCGACAAATTATGTAGTGGCAGAGG---------- 368
b
GLS0803_16936 ----TGTTAAGTCTACTCAGCTGCATGCAAAGGAAGTAACTACAAAAGTATGCGATGAAC 1537
GLS0581_3575 ----------AGTGCATATCAAACGGGACTGTATGTAATGGTAGGGG---------- 356
*
GLS0803_16477 ----------ACTGTATATCGGATGGGAAGGTGTGCAGGGGTAATGG---------- 356
a *
GLS0581_1982 ----------TCTGCATGGTCGACTACAGCGTTTGCAGTTACTATGG---------- 359
GLS0803_95162 ----------TCTGCATGGTTGACTACAGCGTCTGCAGCTACTACGG---------- 359
GLS0803_114815 ----------AGTGCGTGCACGACGGAAAGATATGCGGTGGAGCTGG---------- 413

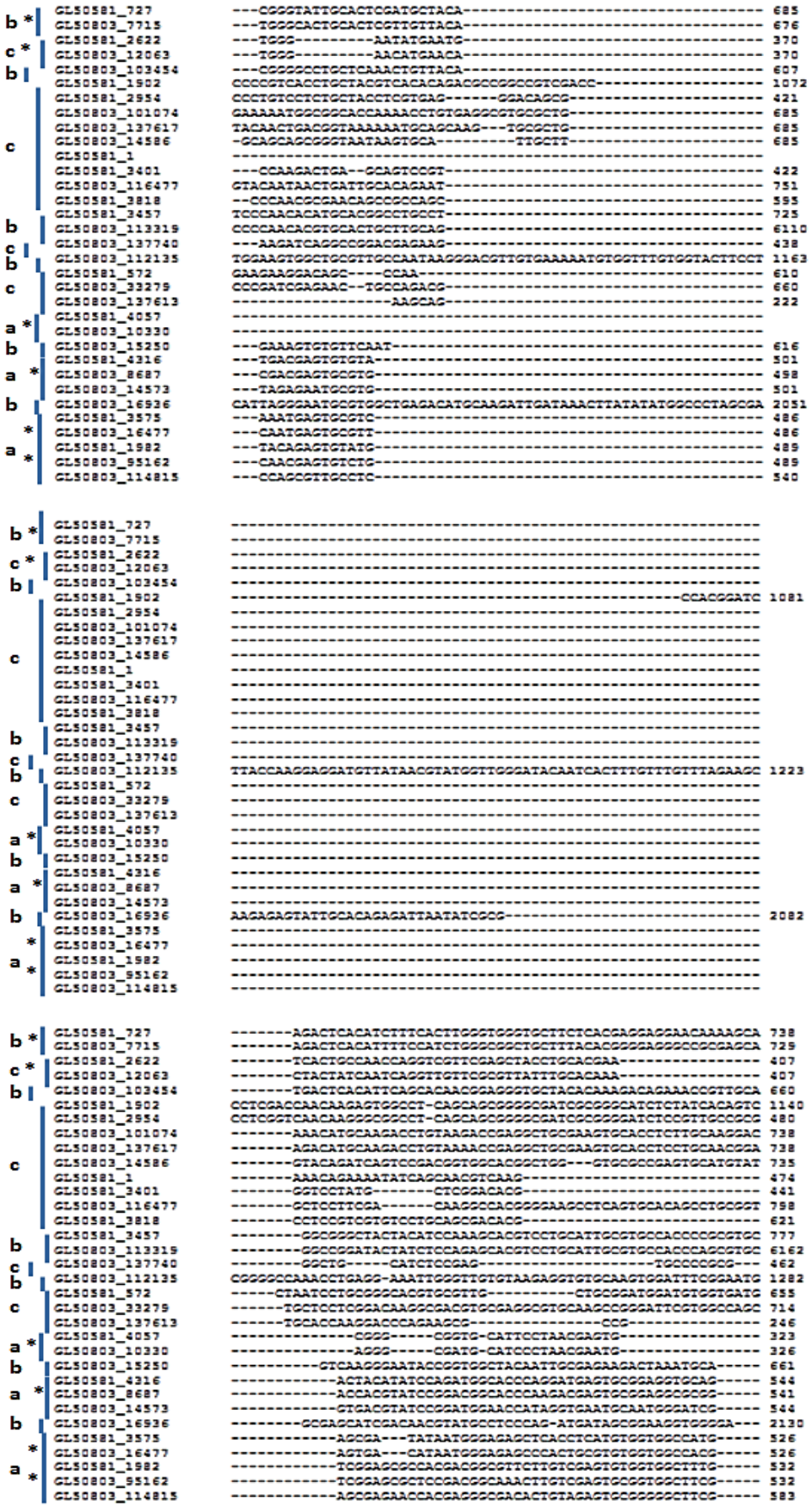

GLS0581_727
GLS0803_7715
GLS0581_2622
GLS0803_12063
GLS0803_103454
GLS0581_1902
GLS0581_2954
GLS0803_101074
GLS0803_137617
GLS0803_14586
GLS0581_1
GLS0581_3401
GLS0803_116477
GLS0581_3818
GLS0581_3457
GLS0803_113319
GLS0803_137740
GLS0803_112135
GLS0581_572
GLS0803_33279
GLS0803_137613
GLS0581_4057
GLS0803_10330
GLS0803_15250
GLS0581_4316
GLS0803_8687
GLS0803_14573
GLS0803_16936
GLS0581_3575
GLS0803_16477
GLS0581_1982
GLS0803_95162
GLS0803_114815

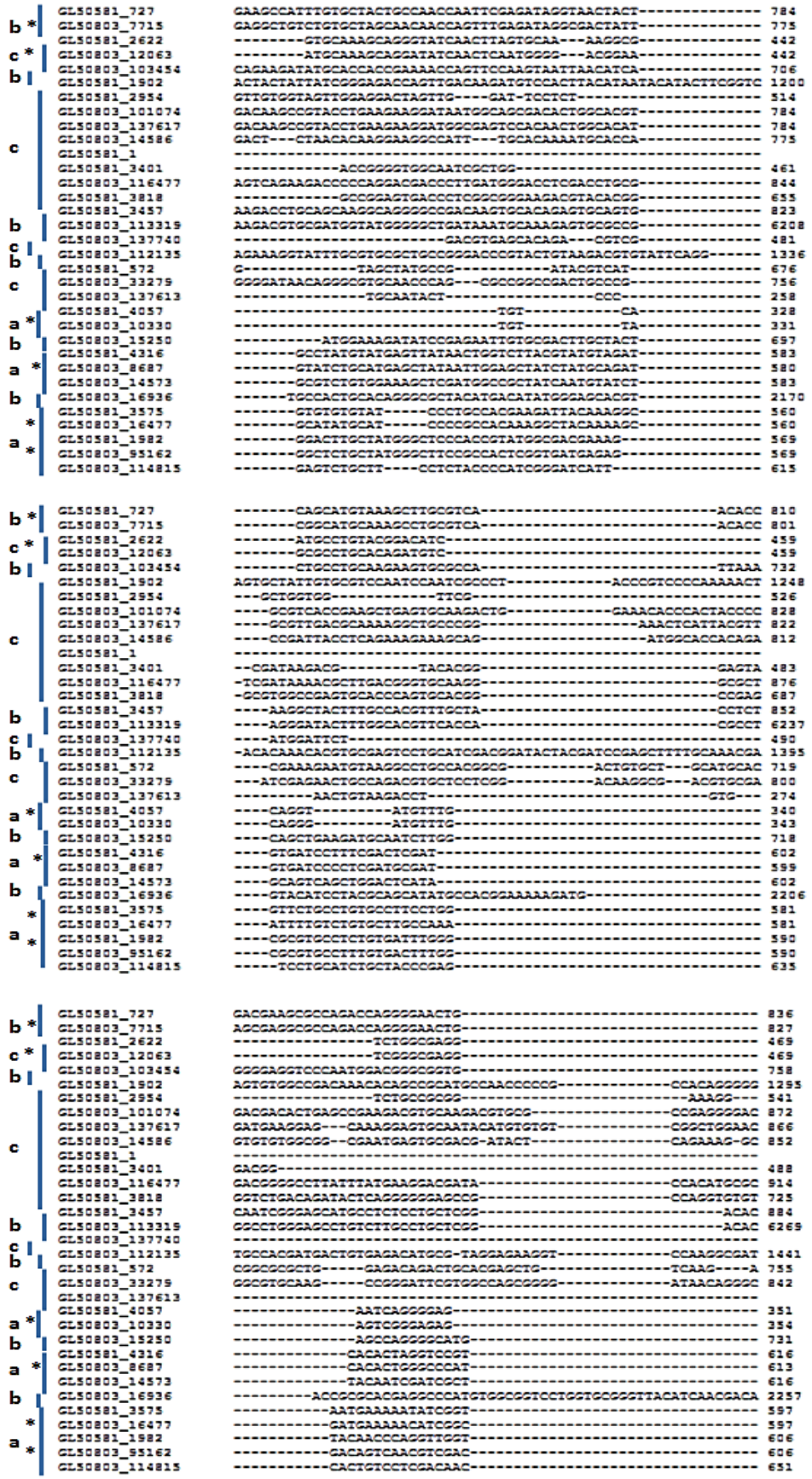

b * GL50581_727 -TATGGATAGAACTACAGACAGCAGTAGCTCGT---------- 571
GL50803_7715 -TATGAACAGAAACTCTGCTG---GTTCCACG---------- 567
c * GL50581_2622 ----------
GL50803_12063 ----------
b GL50803_103454 -CCTGCCTCCGCTCACCGGAGAGACATCTGCA---------- 504
GL50581_1902 ----CACAAAGTGCGACGCTGGATGCGCCACGTGTGCAGGCATTGCTTCGACCTGCAGCA 913
GL50581_2954 ----CACAAAGTGCGACGCTGGGTGTGCCACGTGTACAACTGCTGCCTCAACCTGCAGCA 271
GL50803_101074 ----CACTCGGTGTAACGCTC------CCACAGAAGCAGCAGATGCTAACG-------CC 534
GL50803_137617 ----CACAGCGTGCGATGCAC------CTGCG---CCAGCCGATGCCAGCGGGG----CT 534
c GL50803_14586 ----TGAAGAGTGCAGCGGCT------CGGAG---CTGACCAGTGGCCAAG-------AT 552
GL50581_1 ----CTCGGAGTGCAATACTG------CAGGG---------- 390
GL50581_3401 ------CCACAA---------- 346
GL50803_116477 -GGCTGCCACGTGTACAGAATGCCAAGATGGATATTATAAAGATGCTAAAGGCGCATGTG 565
GL50581_3818 ----TACCAGGGCGGGTGCTATCAGG---------- 487
b GL50581_3457 ---CCTCTTCGAGGGCGCCTG--CTACACGAAAGGAGAGGCGCCGGGAACGAGCATCTGC 588
b GL50803_113319 ---CCTCTTCGAGGGCGCCTG--CTACACAAAGGAAAGGGCGCCGGGAACGAGCATTTGT 5973
c GL50803_137740 ---CGTCTGCGACGGCTCCTT--GACCAC----------CGCGGG----------CTGC 357
b GL50803_112135 ----TTTCCTGTGGAGACGAGCAGGGACACGCTGGATGGAAAGGAGTAGCAAATTGCAAG 882
GL50581_572 ----T---CAGTGCCGTAGAGTG----CAC----------AGCAGCGGGTT----- 472
c GL50803_33279 ----CAGGGCGTGCAACCCAGCGCCGGCCG----------ACTGCCCGAT----- 530
GL50803_137613 ----------CCCAGC---------- 172
GL50581_4057 ------GCCAACGG---------- 249
a * GL50803_10330 ------GCCAGCAAAAC---------- 252
b GL50803_15250 ------ACCAGAGACTATTGCC---------- 535
GL50581_4316 ----TTGCAAGTGTGATGACAC---------- 422
a * GL50803_8687 ----CTGCCAATGTGATGATAC---------- 419
GL50803_14573 ----ATGCAGCTGTGTAGAGCT---------- 422
b GL50803_16936 GGTTTTGGGGAAGGGTTGACAG---------- 1720
GL50581_3575 ----TTGTCAGTGCCTCGAAAC---------- 407
* GL50803_16477 ----CTGCCAGTGCCTTGAATC---------- 407
a * GL50581_1982 ----CTGCCGGTGTGACAAGGT---------- 410
GL50803_95162 ----CTGCCGATGCGACAAGGT---------- 410
GL50803_114815 ----TTGCGTCTGTGATCCCTT---------- 461

b * GL50581_727 ---CTGTAGATATCCAATGCCTACG---------- 593
GL50803_7715 --------GACATTCAGTGCTTGCG---------- 584
c * GL50581_2622 ----------CTTTAG---------- 305
GL50803_12063 ----------CTTCAG---------- 305
b GL50803_103454 ----------GAGTGCAACTG---------- 515
GL50581_1902 CCTGCGCCGACGGGTACTACCTCTCTAACAG---------- 944
GL50581_2954 CCTGCGCCGACGGGTACTACCTCTCTAACAG---------- 302
GL50803_101074 AAGGCCGCCACATGCACTGCGTGCCAGGCAGA---------- 566
GL50803_137617 AAGGTGGCGACGTGTACTGCGTGCCAGGCGGA---------- 566
c GL50803_14586 GGAACCGCTAAGTGCACCAAGTGTGGAGCCAG---------- 584
GL50581_1 ----------
GL50581_3401 ----------ACGGGGTCTGCG---------- 358
GL50803_116477 TACAATGCCAGGGGAACTGCGCCACTTGTGAGACATCG--------GCGACCCAGTGTAC 617
GL50581_3818 ----TTGGACAGGAGCCTGCCTCG---------- 507
GL50581_3457 GGGGACACTGTGAGGTCAACGGCAG---------- 613
b GL50803_113319 GGAGACACCACGGAGTCGACGTCCG---------- 5998
c GL50803_137740 G----------CGACC---------- 363
b GL50803_112135 ACCTGCATTCCTCCCAGTACGCCTGGAGTGGTTAAATGCACAACATGCGAGTTGGGCCAT 942
GL50581_572 -CCTATTTTCCAGACAGCACTTCTG---------- 496
c GL50803_33279 -CGAGAACTGCCAGACGTGCTCCTCGG---------- 556
GL50803_137613 ----------
GL50581_4057 ----------GTCCATG---------- 256
a * GL50803_10330 ----------GTCAGTG---------- 259
b GL50803_15250 -------T-CGAGTGTCTTTCTGGG---------- 552
GL50581_4316 -------AACCGAAGGAGACCAGTG---------- 440
a * GL50803_8687 -------ATCTGAAGCCGATCAGTG---------- 437
GL50803_14573 -------AGTGACTGGGAACCAGTG---------- 440
b GL50803_16936 -------ACGGAATAACTACCAATGGGCCTAGCTGCAATACAAAT---------- 1758
GL50581_3575 -------GTATGCAGGGGAGTTGTG---------- 425
* GL50803_16477 -------ATATGCAGGAAAACTGTG---------- 425
a * GL50581_1982 -------GGCCACCGGCGAGCATTG---------- 428
GL50803_95162 -------GGCAACCGGAGAGCACTG---------- 428
GL50803_114815 -------CTACGCTGGGACCTTCTG---------- 479

b * GL50581_727 ----------GTGCAAGACCTATGGAACT---------- 612
GL50803_7715 ----------CTGCAAGGCTTACGATACT---------- 603
c * GL50581_2622 ----------ATGTGATAATGTACAAGCA---------- 324
GL50803_12063 ----------GTGCGACCATGTAGCGGCA---------- 324
b GL50803_103454 ----------GTGCCAGAATGAGAAGTAT---------- 534
GL50581_1902 ----------CAAGTGCATCGCGTGCGACAAA---------- 566
GL50581_2954 ----------CAAGTGCATCGCGTGCGACAAA---------- 324
GL50803_101074 ----------TAGGTATCTGAAGACAGCCAAG---------- 588
GL50803_137617 ----------TAAGTATCTGAAGACGGCCACA---------- 588
c GL50803_14586 ----------CAAGTATCTCAAGGATAATGCG---------- 606
GL50581_1 ----------GGGTTCTTTAAACATCCAG---------- 409
GL50581_3401 ----------CTGATG--------C---------- 365
GL50803_116477 CTCTTGTAAAGCGGGCTTCTATCTGAAGGACGGCAAT---------- 654
GL50581_3818 ----------TCTCTCTGCACGTCG---------- 522
b GL50581_3457 ----------GGTGCACCACCTGCAAGGCG---------- 633
b GL50803_113319 ----------GGTGCACTGCATGCAAGGCA---------- 6018
c GL50803_137740 ----------TGCACCCCG------GCG---------- 375
b GL50803_112135 TCGCCTATTAATGGAAAATGCGAAGTAGAGTGTGGCGAGCCATCTGGGTAT---------- 993
c GL50581_572 ----------ATCCTAAGGAATGTAAGGCG---------- 516
GL50803_33279 ----------ACAAGGCGACGTGCGAGGCG---------- 576
GL50803_137613 ----------AGTGCAAGACG---------- 183
a * GL50581_4057 ----------GCGTATGGG---------- 265
GL50803_10330 ----------GCAACTGGT---------- 268
b GL50803_15250 ----------ACGAATCGTT---------- 562
GL50581_4316 ----------TGAGATATGTG---------- 451
a * GL50803_8687 ----------TGAGATATGCG---------- 448
GL50803_14573 ----------CGAATCGTGCG---------- 451
b GL50803_16936 -ACAAATATTATTCCCTCTACTTCAACATGCGTATGTCCAACAAACTTTAAGCTCGTGGA 1817
* GL50581_3575 ----------TGAAGAGTGCT---------- 436
GL50803_16477 ----------TGAAGAGTGCT---------- 436
a * GL50581_1982 ----------CGAGGATTGCT---------- 439
GL50803_95162 ----------CGAGAAATGCT---------- 439
GL50803_114815 ----------CGAGGAGTGCG---------- 490

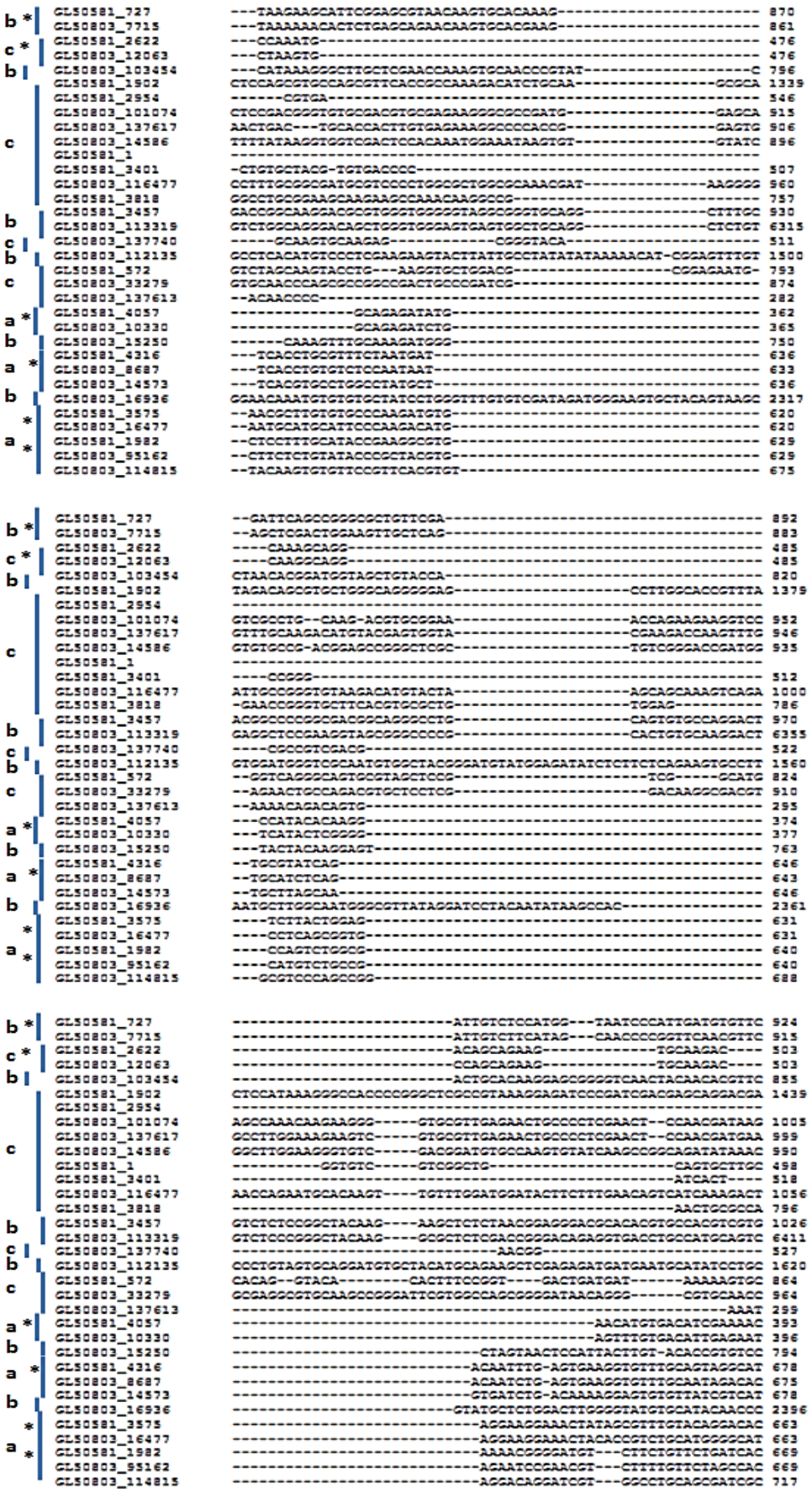



b *
c *
b
c
b
b
c
b
c
a *
b
a *
*
a *
GL50581_727
GL50803_7715
GL50581_2622
GL50803_12063
GL50803_103454
GL50581_1902
GL50581_2954
GL50803_101074
GL50803_137617
GL50803_14586
GL50581_1
GL50581_3401
GL50803_116477
GL50581_3818
GL50581_3457
GL50803_113319
GL50803_137740
GL50803_112135
GL50581_572
GL50803_33279
GL50803_137613
GL50581_4057
GL50803_10330
GL50803_15250
GL50581_4316
GL50803_8687
GL50803_14573
GL50803_16936
GL50581_3575
GL50803_16477
GL50581_1982
GL50803_95162
GL50803_114815


GLS0581_727 GAAGCCATTTGTGCTACTGCCAACCAATTCGAGATAGGTAACTACT---------- 784
b * GLS0803_7715 GAGGCTGTCTGTGCTAGCAACAACCAGTTTGAGATAGGCGACTATT---------- 775
GLS0581_2622 ---------GTGCAAAGCAGGGTATCAACTTAGTGCAA---AAGGCG---------- 442
c * GLS0803_12063 ---------ATGCAAAGCAGGATATCAACTCAATGGGG---ACGGAA---------- 442
b GLS0803_103454 CAGAAGATATGCACCACCGAAAACCAGTTCCAAGTAATTAACATCA---------- 706
GLS0581_1902 ACTACTATTATCGGGAGACCAGTTGACAAGATGTCCACTTACATAATACATACTTCGGTC 1200
GLS0581_2954 GTTGTGGTAGTTGGAGGACTAGTTG----GAT-TCCTCT---------- 514
GLS0803_101074 GACAAGCCGTACCTGAAGAAGGATAATGGCAGCGACACTGGCACGT---------- 784
GLS0803_137617 GACAAGCCGTACCTGAAGAAGGATGGCGAGTCCACAACTGGCACAT---------- 784
c GLS0803_14586 GACT---CTAACACAAGGAAGGCCATT---TGCACAAAATGCACCA---------- 775
GLS0581_1 ----------
GLS0581_3401 ------------ACCGGGGTGGCAATCGCTGG---------- 461
GLS0803_116477 AGTCAGAAGACCCCCAGGACGACCCTTGATGGGACCTCGACCTGCG---------- 844
GLS0803_3818 ------------GCCGGAGTGACCCTCGGCGGGAAGACGTACACGG---------- 655
GLS0581_3457 AAGACCTGCAGCAAGGCAGGGGCCGACAAGTGCACAGAGTGCAGTG---------- 823
b GLS0803_113319 AAGACGTGCGATGGTATGGGGGCTGATAAATGCAAAGAGTGCGCCG---------- 6208
c GLS0803_137740 ----------GACGTGAGCACAGA---CGTCG---------- 481
GLS0803_112135 AGAAAGGTATTTGCGTGCGCTGCGGGGACCCGTACTGTAAGACGTGTATTCAGG------ 1336
b GLS0581_572 G----------TAGCTATGCCG----------ATACGTCAT---------- 676
c GLS0803_33279 GGGGATAACAGGGCGTGCAACCCAG---CGCCGGCCGACTGCCCG---------- 756
GLS0803_137613 ----------TGCAATACT----------CCC---------- 258
GLS0581_4057 ----------TGT----------CA---------- 328
a * GLS0803_10330 ----------TGT----------TA---------- 331
b GLS0803_15250 ----------ATGGAAAGATATCCGAGAATTGTGCGACTTGCTACT---------- 697
GLS0581_4316 -------GCCTATGTATGAGTTATAACTGGTCTTACGTATGTAGAT---------- 583
a * GLS0803_8687 -------GTATCTGCATGAGCTATAATTGGAGCTATCTATGCAGAT---------- 580
GLS0803_14573 -------GCGTCTGTGGAAAGCTCGATGGCCGCTATCAATGTATCT---------- 583
b GLS0803_16936 ------TGCCACTGCACAGGCGCTACATGACATATGGGAGCACGT---------- 2170
GLS0581_3575 -------GTGTGTGTAT-----CCCTGCCACGAAGATTACAAAGGC---------- 560
* GLS0803_16477 -------GCATATGCAT-----CCCCGCCACAAAGGCTACAAAGC---------- 560
a GLS0581_1982 -------GGACTTGCTATGGGCTCCCACCGTATGGCGACGAAAG---------- 569
* GLS0803_95162 -------GGCTCTGCTATGGGCTTCCGCCACTCGGTGATGAGAG---------- 569
GLS0803_114815 -------GAGTCTGCTT----CCTCTACCCCATCGGGATCATT---------- 615

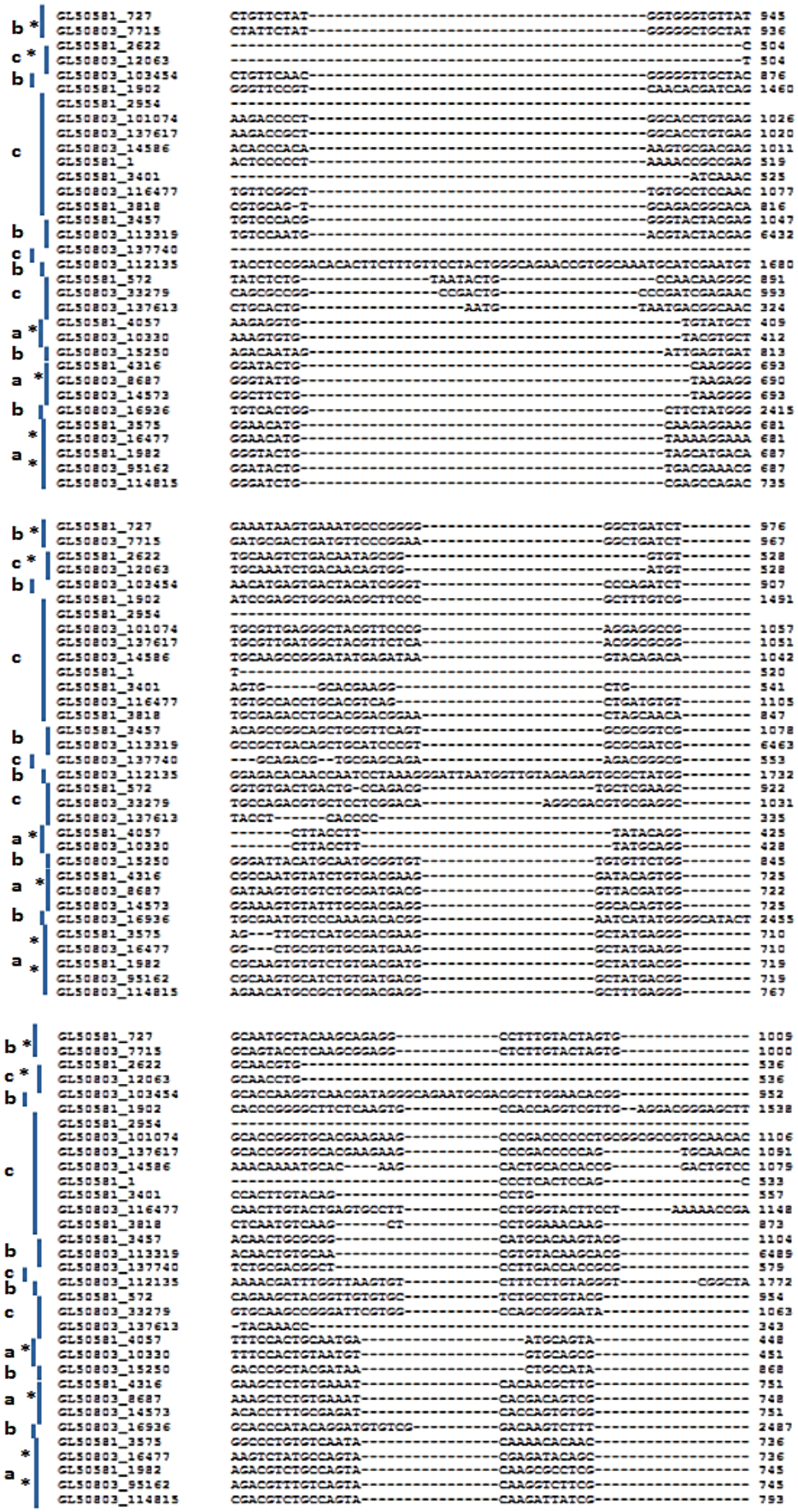
b * GLS0581_727 -------CAGCATGTAAAGCTTGCGTCA----------ACACC 810
GLS0803_7715 -------CGGCATGCAAAGCCTGCGTCA----------ACACC 801
c * GLS0581_2622 -------ATGCCTGTACGGACATC---------- 459
GLS0803_12063 -------GCGCCTGCACAGATGTC---------- 459
b GLS0803_103454 -------CTGCCTGCAAGAAGTGCGCCA----------TTAAA 732
GLS0581_1902 AGTGCTATTGTGCGTCCAATCCAATCGCCCT----------ACCCGTCCCCAAAAACT 1248
GLS0581_2954 ---GCTGGTGG----------TTCG---------- 526
GLS0803_101074 ----GCGTCACCGAAGCTGAGTGCAAGACTG----------GAAACACCCACTACCCC 828
GLS0803_137617 ----GCGTTGACGCAAAAGGCTGCCCGG----------AAACTCATTACGTT 822
c GLS0803_14586 -----CCGATTACCTCAGAAAGAAAGCAG----------ATGGCACCACAGA 812
GLS0581_1 ----------
GLS0581_3401 --CGATAAGACG----------TACACGG----------GAGTA 483
GLS0803_116477 -TCGATAAAACGCTTGACGGGTGCAAGG----------GCGCT 876
GLS0581_3818 -GCGTGGCCGAGTGCACCCAGTGCACGG----------CCGAG 687
b GLS0581_3457 ----AAGGCTACTTTGCCACGTTTGCTA----------CCTCT 852
GLS0803_113319 ----ACGGATACTTTGGCACGTTCACCA----------CGCCT 6237
c GLS0803_137740 ----ATGGATTCT---------- 490
b GLS0803_112135 -ACACAAACACGTGCGAGTCCTGCATCGACGGATACTACGATCCGAGCTTTTGCAAACGA 1395
GLS0581_572 ----CGAAAGAATGTAAGGCCTGCCACGCCG----------ACTGTGCT---GCATGCAC 719
c GLS0803_33279 ---ATCGAGAACTGCCAGACGTGCTCCTCGG----------ACAAGCCG---ACGTGCGA 800
GLS0803_137613 ----------AACTGTAAGACCT----------GTG--- 274
GLS0581_4057 ----CAGGT----------ATGTTTG---------- 340
a * GLS0803_10330 ----CAGGG----------ATGTTTG---------- 343
b GLS0803_15250 ----CAGCTGAAGATGCAATCTTGG---------- 718
GLS0581_4316 ----GTGATCCTTTCGACTCGAT---------- 602
a * GLS0803_8687 ----GTGATCCCCTCGATGCGAT---------- 599
GLS0803_14573 ----GCAGTCAGCTGGACTCATA---------- 602
b GLS0803_16936 ----GTACATCCTACGCAGCATATGCCACGGAAAAAGATG---------- 2206
GLS0581_3575 ----GTTCTGCCTGTGCCTTCCTGG---------- 581
* GLS0803_16477 ----ATTTTGTCTGTGCTTGCCAAA---------- 581
a GLS0581_1982 ----CGCGTGCCTCTGTGATTTGGG---------- 590
* GLS0803_95162 ----CGCGTGCCTTTGTGACTTTGG---------- 590
GLS0803_114815 -----TCCTGCATCTGCTACCCGAG---------- 635

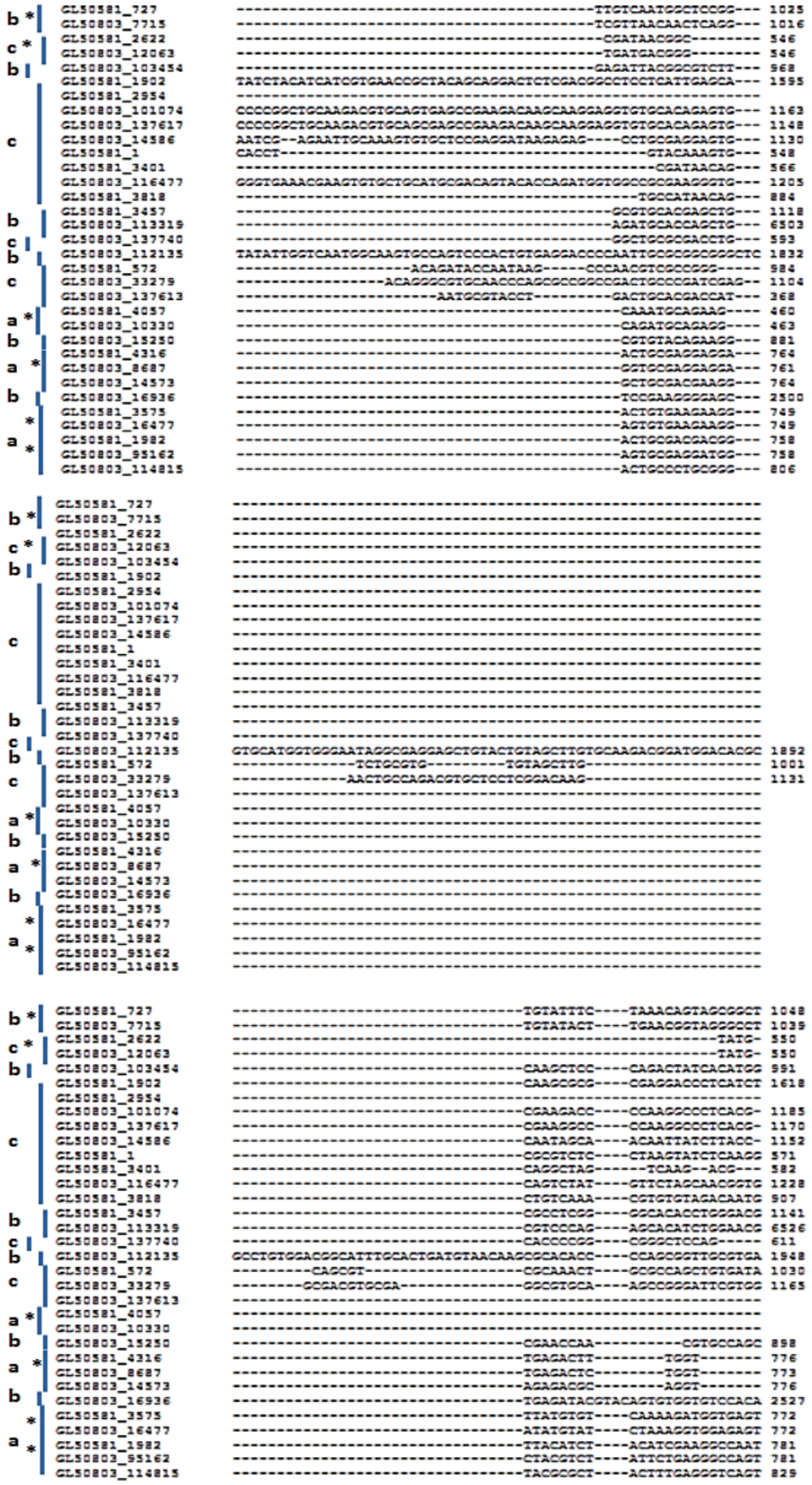
b * GLS0581_727 GACGAAGCGCCAGACCAGGGGAACTG---------- 836
GLS0803_7715 AGCGAGGCGCCAGACCAGGGGAACTG---------- 827
GLS0581_2622 ----------TCTGGCGAGG---------- 469
c * GLS0803_12063 ----------TCGGGCGAGG---------- 469
b GLS0803_103454 GGGGAGGTCCCAATGGACGGGCGGTG---------- 758
GLS0581_1902 AGTGTGGCCGACAAACACAGCCGCATGCCAACCCCCG----------CCACAGGGGG 1295
GLS0581_2954 ----------TCTGCCGCGG----------AAAGG--- 541
GLS0803_101074 GACGACACTGAGCCGAAGACGTGCAAGACGTGCG----------CCGAGGGGAC 872
GLS0803_137617 GATGAAGGAG---CAAAGGAGTGCAATACATGTGTGT----------CGGCTGGAAC 866
c GLS0803_14586 GTGTGTGGCGG--CGAATGAGTGCGACG-ATACT----------CAGAAAG-GC 852
GLS0581_1 ----------
GLS0581_3401 GACGG---------- 488
GLS0803_116477 GACGGGGCCTTATTTATGAAGGACGATA----------CCACATGCGC 914
GLS0581_3818 GGTCTGACAGATACTCAGGGGGGAGCCG----------CCAGGTGTGT 725
GLS0581_3457 CAATCGGGAGCATGCCTCTCCTGCTCGG----------ACAC 884
b GLS0803_113319 GGCCTGCGAGCCCTGTCTTGCCTGCTCGG----------ACAC 6269
GLS0803_137740 ----------
c GLS0803_112135 TGCCACGATGACTGTGAGACATGCG-TAGGAGAAGGT----------CCAAGGCGAT 1441
b GLS0581_572 CGGCGCGCTG-----GAGACAGACTGCACGAGCTG----------TCAAG----A 755
c GLS0803_33279 GGCGTGCAAG-----CCGGGATTCGTGGCCAGCGGGG----------ATAACAGGGC 842
GLS0803_137613 ----------
GLS0581_4057 ----------AATCAGGGGAG---------- 351
a * GLS0803_10330 ----------AGTCGGGAGAG---------- 354
b GLS0803_15250 ----------AGCCAGGGGCATG---------- 731
GLS0581_4316 ----------CACACTAGGTCCGT---------- 616
a * GLS0803_8687 ----------CACACTGGGCCCAT---------- 613
GLS0803_14573 ----------TACAATCGATCGCT---------- 616
b GLS0803_16936 ---------ACCGCGCACGAGGCCCATGTGGCGGTCCTGGTGCGGGTTACATCAACGACA 2257
GLS0581_3575 ----------AATGAAAAATATCGGT---------- 597
* GLS0803_16477 ----------GATGAAAAACATCGGC---------- 597
a GLS0581_1982 ----------TACAACCCAGGTTGGT---------- 606
* GLS0803_95162 ----------GACAGTCAACGTCGAC---------- 606
GLS0803_114815 ----------CACTGTCCTCGACAAC---------- 651


b *
GLS0581_727 ---TAAGAAGCATTCGGAGCGTAACAAGTGCACAAAG---------- 870
GLS0803_7715 ---TAAAAAACACTCTGAGCAGAACAAGTGCACGAAG---------- 861
c *
GLS0581_2622 ---CCAAATG---------- 476
GLS0803_12063 ---CTAAGTG---------- 476
b
GLS0803_103454 ---CATAAAGGGCTTGCTCGAACCAAAGTGCAACCCGTAT----------C 796
GLS0581_1902 CTCCAGCGTGCCAGCGTTCACCGCCAAAGACATCTGCAA----------GCGCA 1339
GLS0581_2954 ------CGTGA---------- 546
GLS0803_101074 CTCCGACGGGTGTGCGACGTGCGAGAAGGGCGCCGATG----------GAGCA 915
GLS0803_137617 AACTGAC---TGCACCACTTGTGAGAAAGGCCCCACCG----------GAGTG 906
c
GLS0803_14586 TTTTATAAGGTGGTCGACTCCACAAATGGAAATAAGTGT----------GTATC 896
GLS0581_1 ----------
GLS0581_3401 -CTGTGCTACG-TGTGACCCC---------- 507
GLS0803_116477 CCTTTGCGGCGATGCGTCCCCTGGCGCTGGCGCAAACGAT----------AAGGGG 960
GLS0581_3818 GGCCTGCGGAAGCAAGAAGCCAAACAAGGCCG---------- 757
GLS0581_3457 GACCGGCAAGGACGCGTGGGTGGGGGTAGGCGGGTGCAGG----------CTTTGC 930
b
GLS0803_113319 GTCTGGCAGGGACAGCTGGGTGGGAGTGAGTGGCTGCAGG----------CTCTGT 6315
c
GLS0803_137740 -----GCAAGTGCAAGAG----------CGGGTACA---------- 511
b
GLS0803_112135 GCCTCACATGTCCCTCGAAGAAGTACTTATTGCCTATATATAAAAACAT-CGGAGTTTGT 1500
GLS0581_572 GTCTAGCAAGTACCTG---AAGGTGCTGGACG----------CGGAGAATG- 793
c
GLS0803_33279 GTGCAACCCAGCGGCCGGCGACTGCCCGATCG---------- 874
GLS0803_137613 --ACAACCCC---------- 282
GLS0581_4057 ----------GCAGAGATATG---------- 362
a *
GLS0803_10330 ----------GCAGAGATCTG---------- 365
b
GLS0803_15250 ------CAAAGTTTGCAAAGATGGG---------- 750
GLS0581_4316 ---TCACCTGCGTTTCTAATGAT---------- 636
a *
GLS0803_8687 ---TCACCTGTGTCTCCAATAAT---------- 633
GLS0803_14573 ---TCACGTGCCTGGCCTATGCT---------- 636
b
GLS0803_16936 GGAACAAATGTGTGTGCTATCCTGGGTTTGTGTCGATAGATGGGAAGTGCTACAGTAAGC 2317
GLS0581_3575 --AACGCTTGTGTGCCCAAGATGTG---------- 620
*
GLS0803_16477 --AATGCATGCATTCCCAAGACATG---------- 620
a *
GLS0581_1982 --CTCCTTTGCATACCGAAGGCGTG---------- 629
GLS0803_95162 --CTTCTCTGTATACCCGCTACGTG---------- 629
GLS0803_114815 --TACAAGTGTGTTCCGTTCACGTGT---------- 675

GLS0581_727 --GATTCAGCCGGGCGCTGTTCGA---------- 892
b *
GLS0803_7715 --AGCTCGACTGGAAGTTGCTCAG---------- 883
GLS0581_2622 ----CAAAGCAGG---------- 485
c *
GLS0803_12063 ----CAAGGCAGG---------- 485
b
GLS0803_103454 CTAACACGGATGGTAGCTGTACCA---------- 820
GLS0581_1902 TAGACAGCGTGCTGGGCAGGGGGAG----------CCTTGGCACCGTTTA 1379
GLS0581_2954 ----------
GLS0803_101074 GTCGCCTG--CAAG-ACGTGCGGAA----------ACCAGAAGAAGGTCC 952
GLS0803_137617 GTTTGCAAGACATGTACGAGTGGTA----------CGAAGACCAAGTTTG 946
c
GLS0803_14586 GTGTGCCG-ACGGAGCCGGGCTCGC----------TGTCGGGACCGATGG 935
GLS0581_1 ----------
GLS0581_3401 ----CCGGG---------- 512
GLS0803_116477 ATTGCCGGGTGTAAGACATGTACTA----------AGCAGCAAAGTCAGA 1000
GLS0581_3818 -GAACGGGGTGCTTCACGTGCGCTG----------TGGAG---------- 786
GLS0581_3457 ACGGCCCCGGCGACGGCAGGGCCTG----------CAGTGTGCCAGGACT 970
b
GLS0803_113319 GAGGCTCCGAAGGTAGCGGGCCCCG----------CACTGTGCAAGGACT 6355
c
GLS0803_137740 ----CGCCGTCGACG---------- 522
b
GLS0803_112135 GTGGATGGGTCGCAATGTGGCTACGGGATGTATGGAGATATCTCTCTCAGAAGTGCCTT 1560
GLS0581_572 --GGTCAGGGCAGTGCGTAGCTCCG----------TCG-----GCATG 824
c
GLS0803_33279 --AGAACTGCCAGACGTGCTCCTCG----------GACAAGGCGACGT 910
GLS0803_137613 --AAAACAGACAGTG---------- 295
GLS0581_4057 ---CCATACACAAGG---------- 374
a *
GLS0803_10330 ---TCATACTCGGGG---------- 377
b
GLS0803_15250 ---TACTACAAGGAGT---------- 763
GLS0581_4316 --TGCGTATCAG---------- 646
a *
GLS0803_8687 --TGCATCTCAG---------- 643
GLS0803_14573 --TGCTTAGCAA---------- 646
b
GLS0803_16936 AATGCTTGGCAATGGGCGTTATAGGATCCTACAATATAAGCCAC---------- 2361
GLS0581_3575 ----TCTTACTGGAG---------- 631
*
GLS0803_16477 ----CCTCAGCGGTG---------- 631
a *
GLS0581_1982 ----CCAGTCTGGGG---------- 640
GLS0803_95162 ----CATGTCTGCCG---------- 640
GLS0803_114815 ---GCGTCCCAGCCGG---------- 688

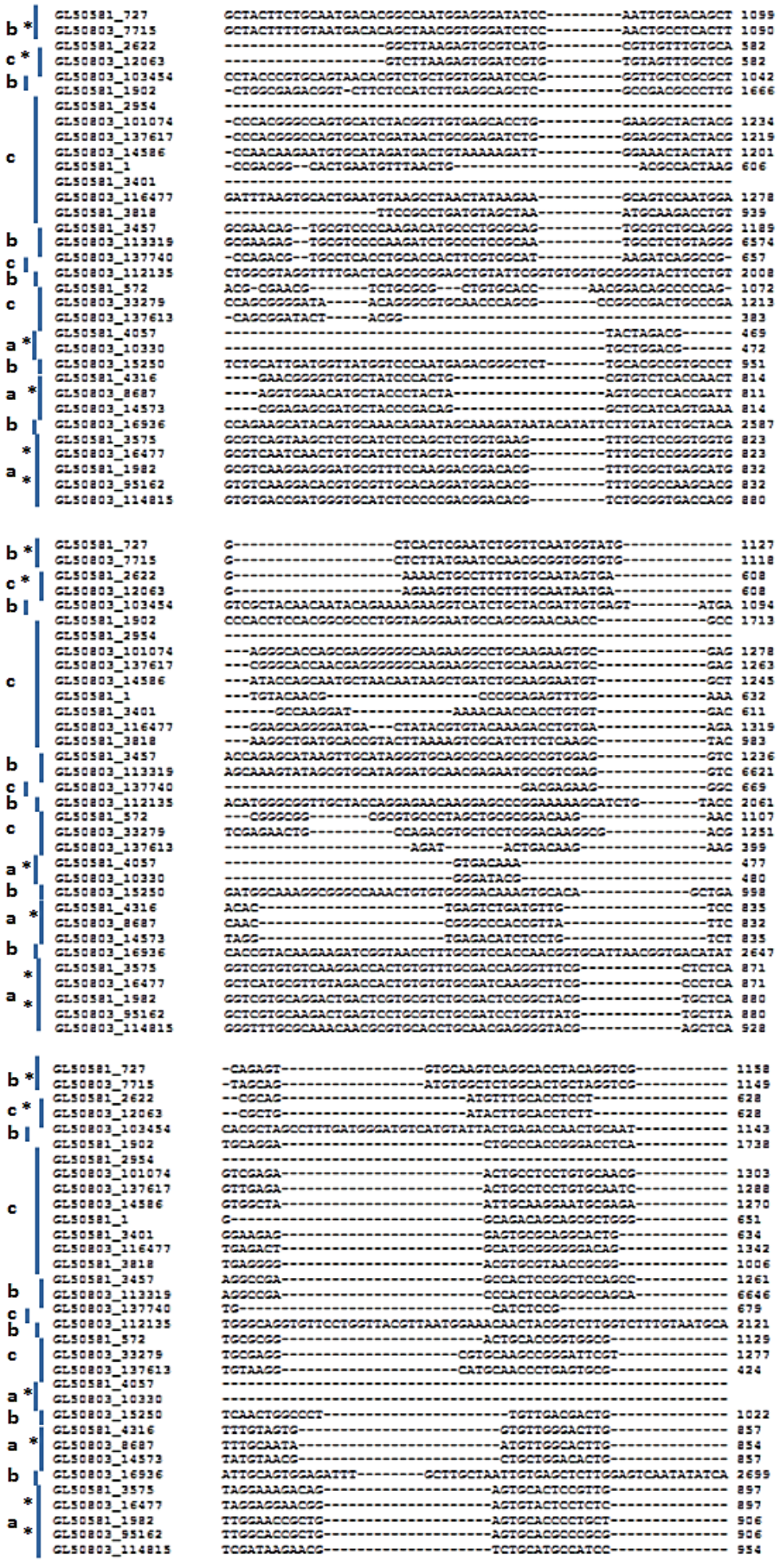
b *
GLS0581_727 ----------ATTGTCTCCATGG---TAATCCCATTGATGTGTTC 924
GLS0803_7715 ----------ATTGTCTTCATAG---CAACCCCGGTTCAACGTTC 915
c *
GLS0581_2622 ----------ACAGCAGAAG----------TGCAAGAC---- 503
GLS0803_12063 ----------CCAGCAGAAG----------TGCAAGAC---- 503
b
GLS0803_103454 ----------ACTGCACAAGGAGCGGGGTCAACTACAACACGTTC 855
GLS0581_1902 CTCCATAAAGGGCCACCCCGGGCTCGCCGTAAAGGAGATCCCGATCGACGAGCAGGACGA 1439
GLS0581_2954 ----------
GLS0803_101074 AGCCAAACAAGAAGGG-----GTGCGTTGAGAACTGCCCCTCGAACT--CCAACGATAAG 1005
GLS0803_137617 GCCTTGGAAAGAAGTC-----GTGCGTTGAGAACTGCCCCTCGAACT--CCAACGATGAA 999
c
GLS0803_14586 GGCTTGGAAGGGTGTC-----GACGGATGTGCCAAGTGTATCAAGCCGGCAGATATAAAC 990
GLS0581_1 ----------GGTGTC-----GTCGGCTG----------CAGTGCTTGC 498
GLS0581_3401 ----------ATCACT---- 518
GLS0803_116477 AACCAGAATGCACAAGT----TGTTTGGATGGATACTTCTTTGAACAGTCATCAAAGACT 1056
GLS0581_3818 ----------AACTGCGCCA 796
b
GLS0581_3457 GTCTCTCCGGCTACAAG----AAGCTCTCTAACGGAGGGACGCACACGTGCCACGTCGTG 1026
GLS0803_113319 GTCTCCCGGGCTACAAG----GCGCTCTCGACCGGGACAGAGGTGACCTGCCATGCAGTC 6411
c
GLS0803_137740 ----------AACGG---------- 527
b
GLS0803_112135 CCCTGTAGTGCAGGATGTGCTACATGCAGAAGCTCGAGAGATGATGAATGCATATCCTGC 1620
GLS0581_572 CACAG--GTACA---------CACTTTCCGGT----GACTGATGAT------AAAAAGTGC 864
c
GLS0803_33279 GCGAGGCGTGCAAGCCGGGATTCGTGGCCAGCGGGGATAACAGGG------CGTGCAACC 964
GLS0803_137613 ----------AAAT 299
GLS0581_4057 ----------AACATGTGACATCGAAAAC 393
a *
GLS0803_10330 ----------AGTTTGTGACATTGAGAAT 396
b
GLS0803_15250 ----------CTAGTAACTCCATTACTTGT-ACACCGTGTCC 794
GLS0581_4316 ----------ACAATTTG-AGTGAAGGTGTTTGCAGTAGGCAT 678
a *
GLS0803_8687 ----------ACAATCTG-AGTGAAGGTGTTTGCAATAGACAC 675
GLS0803_14573 ----------GTGATCTG-ACAAAAGGAGTGTGTTATCGTCAT 678
b
GLS0803_16936 ----------GTATGCTCTGGACTTGGGGTATGTGCATACAACCC 2396
GLS0581_3575 ----------AGGAAGGAAACTATAGCGTTTGTACAGGACAC 663
*
GLS0803_16477 ----------AGGAAGGAAACTACACCGTCTGCATGGGGCAT 663
a *
GLS0581_1982 ----------AAAACGGGGATGT---CTTCTGTTCTGATCAC 669
GLS0803_95162 ----------AGAATCCGAACGT---CTTTTGTTCTAGCCAC 669
GLS0803_114815 ----------AGGACAGGATCGT---GGCCTGCAGCGATCGC 717

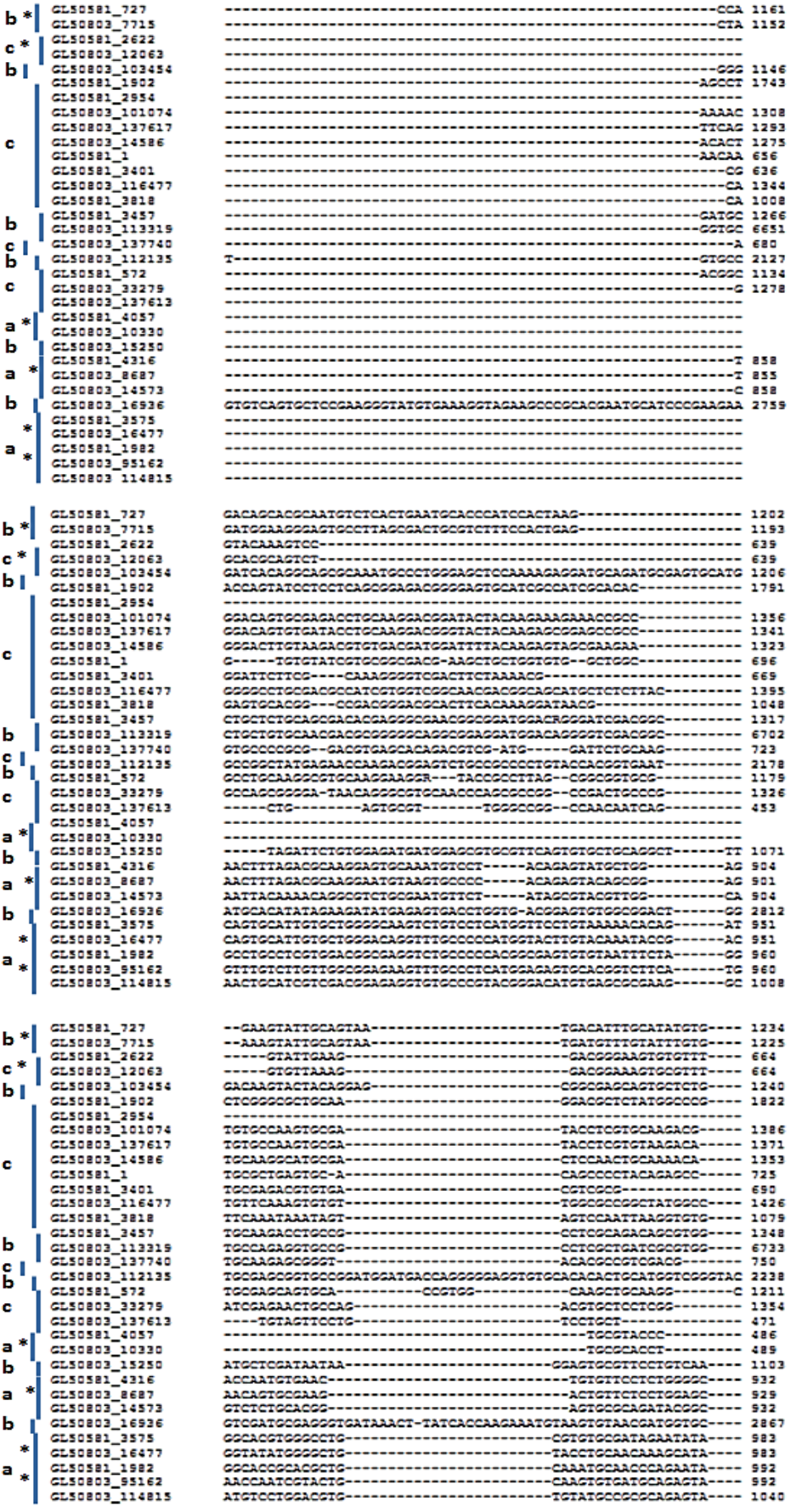

b * GL50581_727 CTGTTCTAT---------- GGTGGGTGTTAT 945
GL50803_7715 CTATTCTAT---------- GGGGGCTGCTAT 936
c * GL50581_2622 ---------- C 504
GL50803_12063 ---------- T 504
b GL50803_103454 CTGTTCAAC---------- GGGGGTTGCTAC 876
GL50581_1902 GGGTTCCGT---------- CAACACGATCAG 1460
GL50581_2954 ----------
GL50803_101074 AAGACCCCT---------- GGCACCTGTGAG 1026
GL50803_137617 AAGACCGCT---------- GGCACCTGTGAG 1020
c GL50803_14586 ACACCCACA---------- AAGTGCGACGAG 1011
GL50581_1 ACTCCCCCT---------- AAAACCGCCGAG 519
GL50581_3401 ---------- ATCAAAC 525
GL50803_116477 TGTTCGGCT---------- TGTGCCTCCAAC 1077
GL50581_3818 CGTGCAG-T---------- GCAGACGGCACA 816
GL50581_3457 TGTCCCACG---------- GGGTACTACGAG 1047
b GL50803_113319 TGTCCAATG---------- ACGTACTACGAG 6432
c GL50803_137740 ----------
b GL50803_112135 TACCTCCGGACACACTTCTTTGTTCCTACTGGGCAGAACCGTGGCAAATGCATCGAATGT 1680
GL50581_572 TATCTCTG---------- TAATACTG---------- CCAACAAGGGC 891
c GL50803_33279 CAGCGCCGG---------- CCGACTG---------- CCCGATCGAGAAC 993
GL50803_137613 CTGCACTG---------- AATG---------- TAATGACGGCAAC 924
GL50581_4057 AAGAGGTG---------- TGTATGCT 409
a * GL50803_10330 AAAGTGTG---------- TACGTGCT 412
b GL50803_15250 AGACAATAG---------- ATTGAGTGAT 813
GL50581_4316 GGATACTG---------- CAAGGGG 693
a * GL50803_8687 GGGTATTG---------- TAAGAGG 690
GL50803_14573 GGCTTCTG---------- TAAGGGG 693
b GL50803_16936 TGTCACTGG---------- CTTCTATGGG 2415
GL50581_3575 GGAACATG---------- CAAGAGGAAG 681
* GL50803_16477 GGAACATG---------- TAAAAGGAAA 681
a * GL50581_1982 GGGTACTG---------- TAGCATGACA 687
GL50803_95162 GGATACTG---------- TGACGAAACG 687
GL50803_114815 GGGATCTG---------- CGAGCCAGAC 735
b * GL50581_727 GAAATAAGTGAAATGCCCGGGG---------- GGCTGATCT-------- 976
GL50803_7715 GATGCGACTGATGTTCCCGGAA---------- GGCTGATCT-------- 967
c * GL50581_2622 TGCAAGTCTGACAATAGCGG---------- GTGT-------- 528
GL50803_12063 TGCAAATCTGACAACAGTGG---------- ATGT-------- 528
b GL50803_103454 AACATGAGTGACTACATCGGGT---------- CCCAGATCT-------- 907
GL50581_1902 ATCCGAGCTGGCGACGCTTCCC---------- GCTTTGTCG-------- 1491
GL50581_2954 ----------
GL50803_101074 TGCGTTGAGGGCTACGTTCCCG---------- AGGAGGCCG-------- 1057
GL50803_137617 TGCGTTGATGGCTACGTTCTCA---------- ACGCCGCGG-------- 1051
c GL50803_14586 TGCAAGCCGGCATATGAGATAA---------- GTACAGACA-------- 1042
GL50581_1 T---------- 520
GL50581_3401 AGTG------GCACGAAGG---------- CTG---------- 541
GL50803_116477 TGTGCCACCTGCACGTCAG---------- CTGATGTGT-------- 1105
GL50581_3818 TCCGAGACCTGCACGGACGGAA---------- CTAGCAACA-------- 847
GL50581_3457 ACAGCCGGCAGCTGCGTTCAGT---------- GCGCGGTCG-------- 1078
b GL50803_113319 GCCGCTGACAGCTGCATCCCGT---------- GCGCGATCG-------- 6463
c GL50803_137740 ---GCAGACG--TGCGAGCAGA---------- AGACGGCCG-------- 553
b GL50803_112135 GGAGACACAACCAATCCTAAAGGGATTAATGGTTGTAGAGAGTGCGCTATGG-------- 1732
GL50581_572 GGTGTGACTGACTG-CCAGACG---------- TGCTCGAAGC-------- 922
c GL50803_33279 TGCCAGACGTGCTCCTCGGACA---------- AGGCGACGTGCGAGGC-------- 1031
GL50803_137613 TACCT------CACCCC---------- 335
GL50581_4057 -------CTTACCTT---------- TATACAGG-------- 425
a * GL50803_10330 -------CTTACCTT---------- TATGCAGG-------- 428
b GL50803_15250 GGGATTACATGCAATGCGGTGT---------- TGTGTTCTGG-------- 845
GL50581_4316 CGCCAATGTATCTGTGACGAAG---------- GATACAGTGG-------- 725
a * GL50803_8687 GATAAGTGTGTCTGCGATGACG---------- GTTACGATGG-------- 722
GL50803_14573 GGAAAGTGTATTTGCGACGAGG---------- GGCACAGTGG-------- 725
b GL50803_16936 TCCGAATGTCCCAAAGACACGG---------- AATCATATGGGGCATACT 2455
GL50581_3575 AG---TTGCTCATGCGACGAAG---------- GCTATGAGGG-------- 710
* GL50803_16477 GG---CTGCGTGTGCGATGAAG---------- GCTATGAAGG-------- 710
a * GL50581_1982 CGCAAGTGTGTCTGTGACGATG---------- GCTATGACGG-------- 719
GL50803_95162 CGCAAGTGCATCTGTGATGACG---------- GCTATGACGG-------- 719
GL50803_114815 AGAACATGCCGCTGCGACGAGG---------- GCTTTGAGGG-------- 767
b * GL50581_727 GCAATGCTACAAGCAGAGG---------- CCTTTGTACTAGTG---------- 1009
GL50803_7715 GCAGTACCTCAAGCGGAGG---------- CTCTTGTACTAGTG---------- 1000
c * GL50581_2622 GCAACGTG---------- 536
GL50803_12063 GCAACCTG---------- 536
b GL50803_103454 GCACCAAGGTCAACGATAGGGCAGAATGCGACGCTTGGAACACGG---------- 952
GL50581_1902 CACCCGGGGCTTCTCAAGTG---------- CCACCAGGTCGTTG--AGGACGGGAGCTT 1538
GL50581_2954 ----------
GL50803_101074 GCACCGGGTGCACGAAGAAG---------- CCCGACCCCCCTGCGGGCGCGTGCAACAC 1106
GL50803_137617 GCACCGGGTGCACGAAGAAG---------- CCCGACCCCCAG---------TGCAACAC 1091
c GL50803_14586 AAACAAAATGCAC----AAG---------- CACTGCACCACCG---------GACTGTCC 1079
GL50581_1 ---------- CCCTCACTCCAG---------- C 533
GL50581_3401 CCACTTGTACAG---------- CCTG---------- 557
GL50803_116477 CAACTTGTACTGAGTGCCTT---------- CCTGGGTACTTCCT------AAAAACCGA 1148
GL50581_3818 CTCAATGTCAAG------CT---------- CCTGGAAACAAG---------- 873
GL50581_3457 ACAACTGCGCGG---------- CATGCACAAGTACG---------- 1104
b GL50803_113319 ACAACTGTGCAA---------- CGTGTACAAGCACG---------- 6489
c GL50803_137740 TCTGCGACGGCT---------- CCTTGACCACCGCG---------- 579
b GL50803_112135 AAAACGATTTGGTTAAGTGT---------- CTTTCTTGTAGGGT---------CGGCTA 1772
GL50581_572 CAGAAGCTACGGTTGTGTGC---------- TCTGCCTGTACG---------- 954
c GL50803_33279 GTGCAAGCCGGGGATTCGTGG---------- CCAGCGGGGATA---------- 1063
GL50803_137613 -TACAAACC---------- 343
GL50581_4057 TTTCCACTGCAATGA---------- ATGCAGTA---------- 448
a * GL50803_10330 TTTCCACTGTAATGT---------- GTGCAGCG---------- 451
b GL50803_15250 GACCCGCTACGATAA---------- CTGCCATA---------- 868
GL50581_4316 GAAGCTCTGTGAAAT---------- CACAACGCTTG---------- 751
a * GL50803_8687 AAAGCTCTGTGAAAT---------- CACGACAGTCG---------- 748
GL50803_14573 ACACCTTTGCGAGAT---------- CACCAGTGTGG---------- 751
b GL50803_16936 GCACCCATACAGGATGTGTCG---------- GACAAGTCTTT---------- 2487
GL50581_3575 GGCCCTGTGTCAATA---------- CAAAACACAAC---------- 736
* GL50803_16477 AAGTCTATGCCAGTA---------- CGAGATACAGC---------- 736
a * GL50581_1982 AGACGTCTGCCAGTA---------- CAAGCGCCTCG---------- 745
GL50803_95162 AGACGTTTGTCAGTA---------- CAAGGTCTTCG---------- 745
GL50803_114815 CGACGTCTGCCAGTA---------- CAAGATTATCG---------- 793

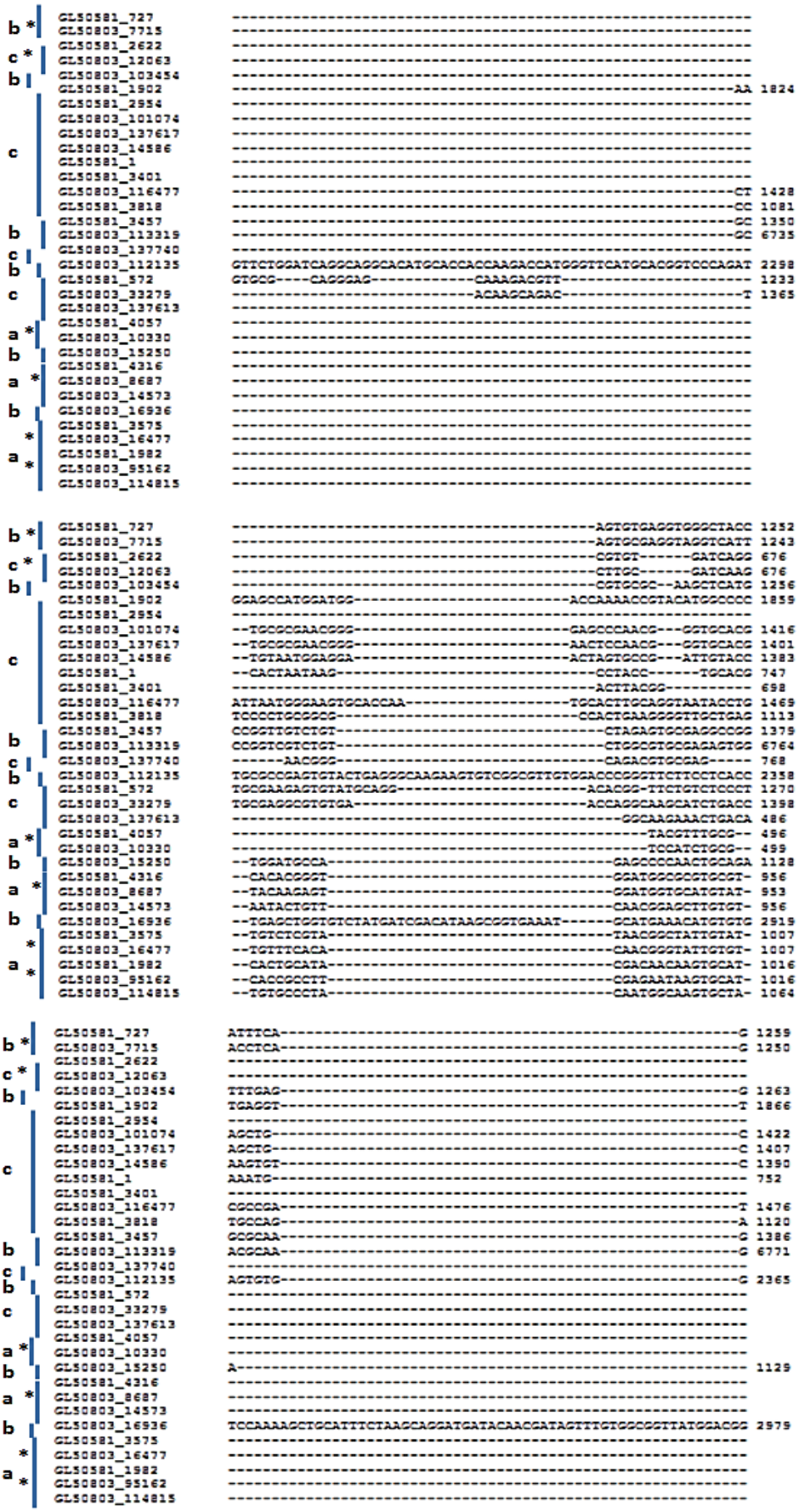

b * GL50581_727 ----------TTGTCAATGGCTCCGG--- 1025
GL50803_7715 ----------TCGTTAACAACTCAGG--- 1016
c * GL50581_2622 ----------CGATAACGGC-------- 546
GL50803_12063 ----------TGATGACGGG-------- 546
b GL50803_103454 ----------GAGATTACGGCGTCTT--- 968
GL50581_1902 TATCTACATCATCGTGAACCGCTACAGCAGGACTCTCGACGGCCTCCTCATTGAGCA--- 1595
GL50581_2954 ----------
GL50803_101074 CCCCGGCTGCAAGACGTGCAGTGAGCCGAAGACAAGCAAGGAGGTGTGCACAGAGTG--- 1163
GL50803_137617 CCCCGGCTGCAAGACGTGCAGCGAGCCGAAGACAAGCAAGGAGGTGTGCACAGAGTG--- 1148
c GL50803_14586 AATCG--AGAATTGCAAAGTGTGCTCCGAGGATAAGAGAG----CCTGCGAGGAGTG--- 1130
GL50581_1 CACCT----------GTACAAAGTG--- 548
GL50581_3401 ----------CGATAACAG--- 566
GL50803_116477 GGGTGAAACGAAGTGTGCTGCATGCGACAGTACACCAGATGGTGGCCGCGAAGGGTG--- 1205
GL50581_3818 ----------TGCCATAACAG--- 884
b GL50581_3457 ----------GCGTGCACGAGCTG--- 1118
GL50803_113319 ----------AGATGCACCAGCTG--- 6503
c GL50803_137740 ----------GGCTGCGCGACCTG--- 593
b GL50803_112135 TATATTGGTCAATGGCAAGTGCCAGTCCCACTGTGAGGACCCCAATTGCGCGCGGGGCTC 1832
GL50581_572 ----------ACAGATACCAATAAG-----CCCAACGTCGCCGGG----- 984
c GL50803_33279 ----------ACAGGGCGTGCAACCCAGCGCCGGCCGACTGCCCGATCGAG-- 1104
GL50803_137613 ----------AATGCGTACCT----------GACTGCACGACCAT--- 368
GL50581_4057 ----------CAAATGCAGAAG---- 460
a * GL50803_10330 ----------CAGATGCAGAGG---- 463
b GL50803_15250 ----------CGTGTACAGAAGG--- 881
GL50581_4316 ----------ACTGCGAGGAGGA--- 764
a * GL50803_8687 ----------GGTGCGAGGAGGA--- 761
GL50803_14573 ----------GCTGCGACGAAGG--- 764
b GL50803_16936 ----------TCCGAAGGCGAGC--- 2500
GL50581_3575 ----------ACTGTGAAGAAGG--- 749
* GL50803_16477 ----------AGTGTGAAGAAGG--- 749
a * GL50581_1982 ----------ACTGCGACGACGG--- 758
GL50803_95162 ----------AGTGCGAGGATGG--- 758
GL50803_114815 ----------ACTGCCCTGCGGG--- 806

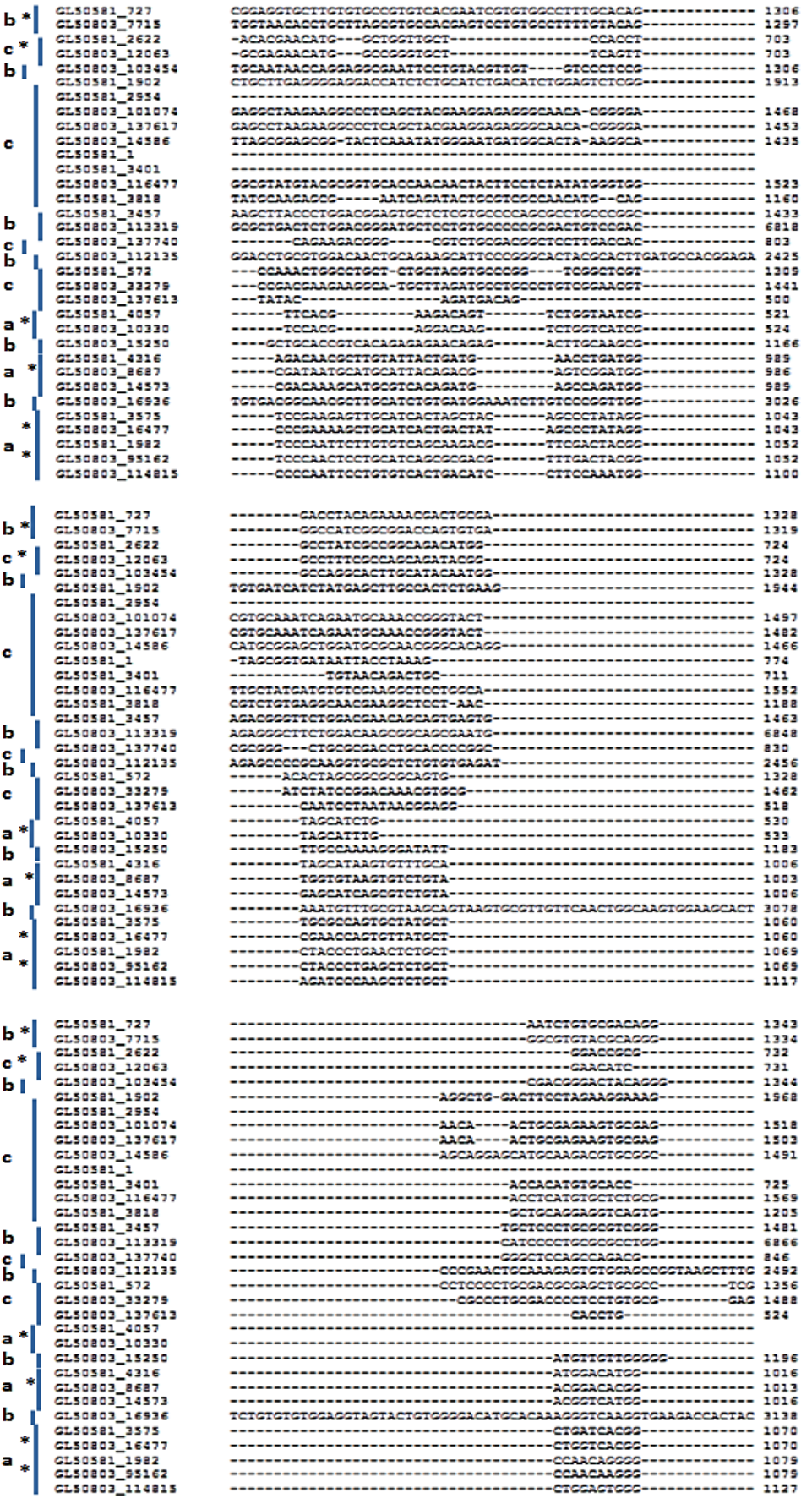
b * GL50581_727 ----------
GL50803_7715 ----------
c * GL50581_2622 ----------
GL50803_12063 ----------
b GL50803_103454 ----------
GL50581_1902 ----------
GL50581_2954 ----------
GL50803_101074 ----------
GL50803_137617 ----------
c GL50803_14586 ----------
GL50581_1 ----------
GL50581_3401 ----------
GL50803_116477 ----------
GL50581_3818 ----------
b GL50581_3457 ----------
GL50803_113319 ----------
c GL50803_137740 ----------
b GL50803_112135 GTGCATGGTGGGAATAGGCGAGGAGCTGTACTGTAGCTTGTGCAAGACGGATGGACACGC 1892
GL50581_572 ----------TCTGCGTG----------TGTAGCTTG---------- 1001
c GL50803_33279 ----------AACTGCCAGACGTGCTCCTCGGACAAG---------- 1131
GL50803_137613 ----------
GL50581_4057 ----------
a * GL50803_10330 ----------
b GL50803_15250 ----------
GL50581_4316 ----------
a * GL50803_8687 ----------
GL50803_14573 ----------
b GL50803_16936 ----------
GL50581_3575 ----------
* GL50803_16477 ----------
a * GL50581_1982 ----------
GL50803_95162 ----------
GL50803_114815 ----------

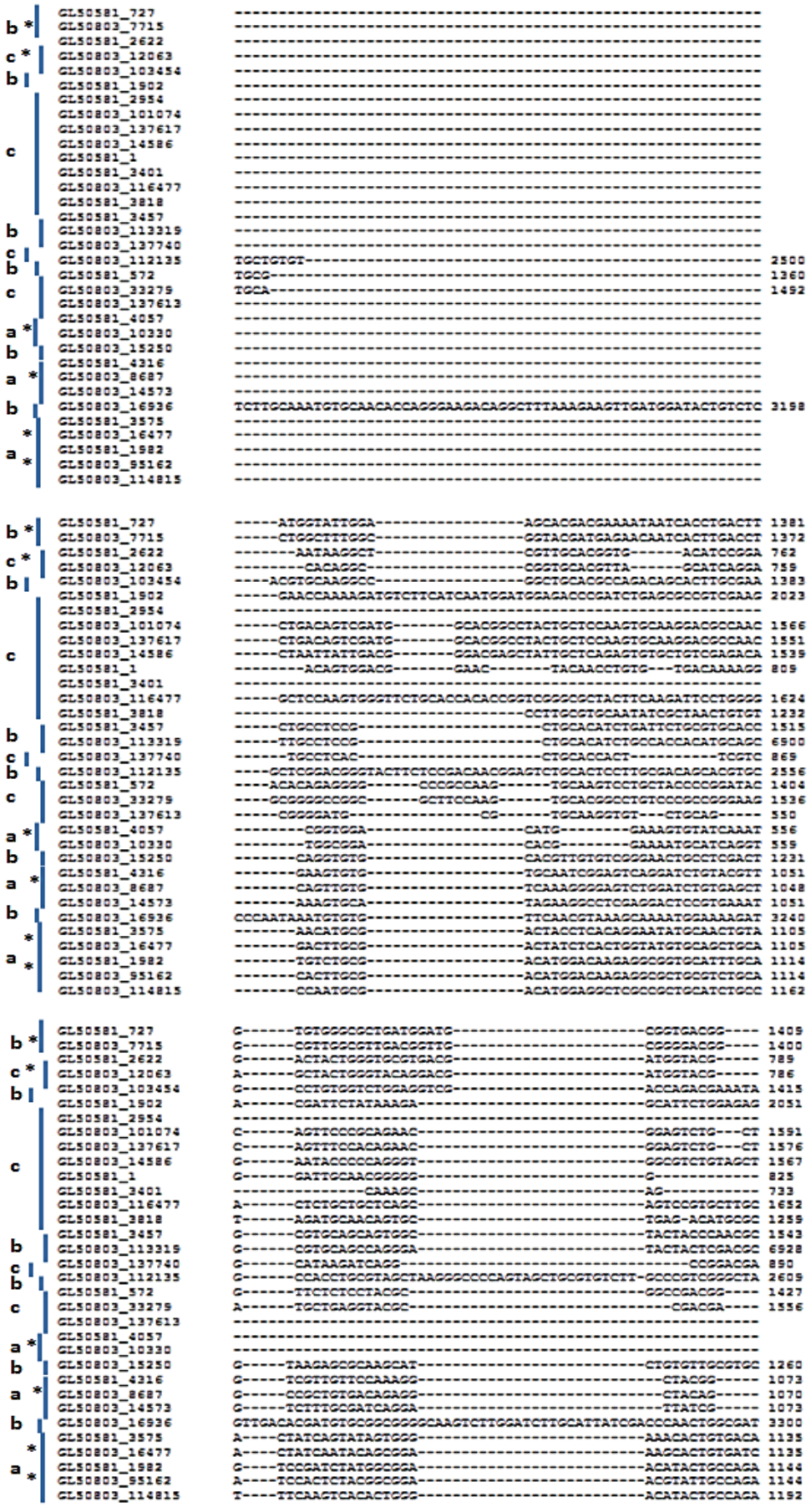
b * GL50581_727 ----------TGTATTTC----TAAACAGTAGCGGCT 1048
GL50803_7715 ----------TGTATACT----TGAACGGTAGCGGCT 1039
c * GL50581_2622 ----------TATG- 550
GL50803_12063 ----------TATG- 550
b GL50803_103454 ----------CAAGCTCC----CAGACTATCACATGG 991
GL50581_1902 ----------CAAGCGCG----CGAGGACCCTCATCT 1618
GL50581_2954 ----------
GL50803_101074 ----------CGAAGACC----CCAAGGCCCTCACG- 1185
GL50803_137617 ----------CGAAGCCC----CCAAGCCCCTCACG- 1170
c GL50803_14586 ----------CAATAGCA----ACAATTATCTTACC- 1152
GL50581_1 ----------CGCGTCTC----CTAAGTATCTCAAGG 571
GL50581_3401 ----------CAGGCTAG------TCAAG--ACG--- 582
GL50803_116477 ----------CAGTCTAT----GTTCTAGCAACGGTG 1228
GL50581_3818 ----------CTGTCAAA----CGTGTGTAGACAATG 907
b GL50581_3457 ----------CGCCTCGG----GGCACACCTGGGACG 1141
GL50803_113319 ----------CGTCCCAG----AGCACATCTGGAACG 6526
c GL50803_137740 ----------CACCCCGG----CGGGCTCCAG----- 611
b GL50803_112135 GCCTGTGGACGGCATTTGCACTGATGTAACAAGCGCACACC----CCAGCGGTTGCGTGA 1948
GL50581_572 ----------CAGCGT----------CGCAAACT----GCGCCAGCTGTGATA 1030
c GL50803_33279 ----------GCGACGTCCGA----------GGCGTGCA----AGCCGGGATTCGTGG 1165
GL50803_137613 ----------
GL50581_4057 ----------
a * GL50803_10330 ----------
b GL50803_15250 ----------CGAACCAA----------CGTGCCAGC 898
GL50581_4316 ----------TGAGACTT---------TGGT-------- 776
a * GL50803_8687 ----------TGAGACTC---------TGGT-------- 773
GL50803_14573 ----------AGAGACGC---------AGGT-------- 776
b GL50803_16936 ----------TGAGATACGTACAGTGTGGTGTCCACA 2527
GL50581_3575 ----------TTATGTGT----CAAAAGATGGTGAGT 772
* GL50803_16477 ----------ATATGTAT----CTAAAGGTGGAGAGT 772
a * GL50581_1982 ----------TTACATCT----ACATCGAAGGCCAAT 781
GL50803_95162 ----------CTACGTCT----ATTCTGAGGGCCAGT 781
GL50803_114815 ----------TACGCGCT----ACTTTGAGGGTCAGT 829

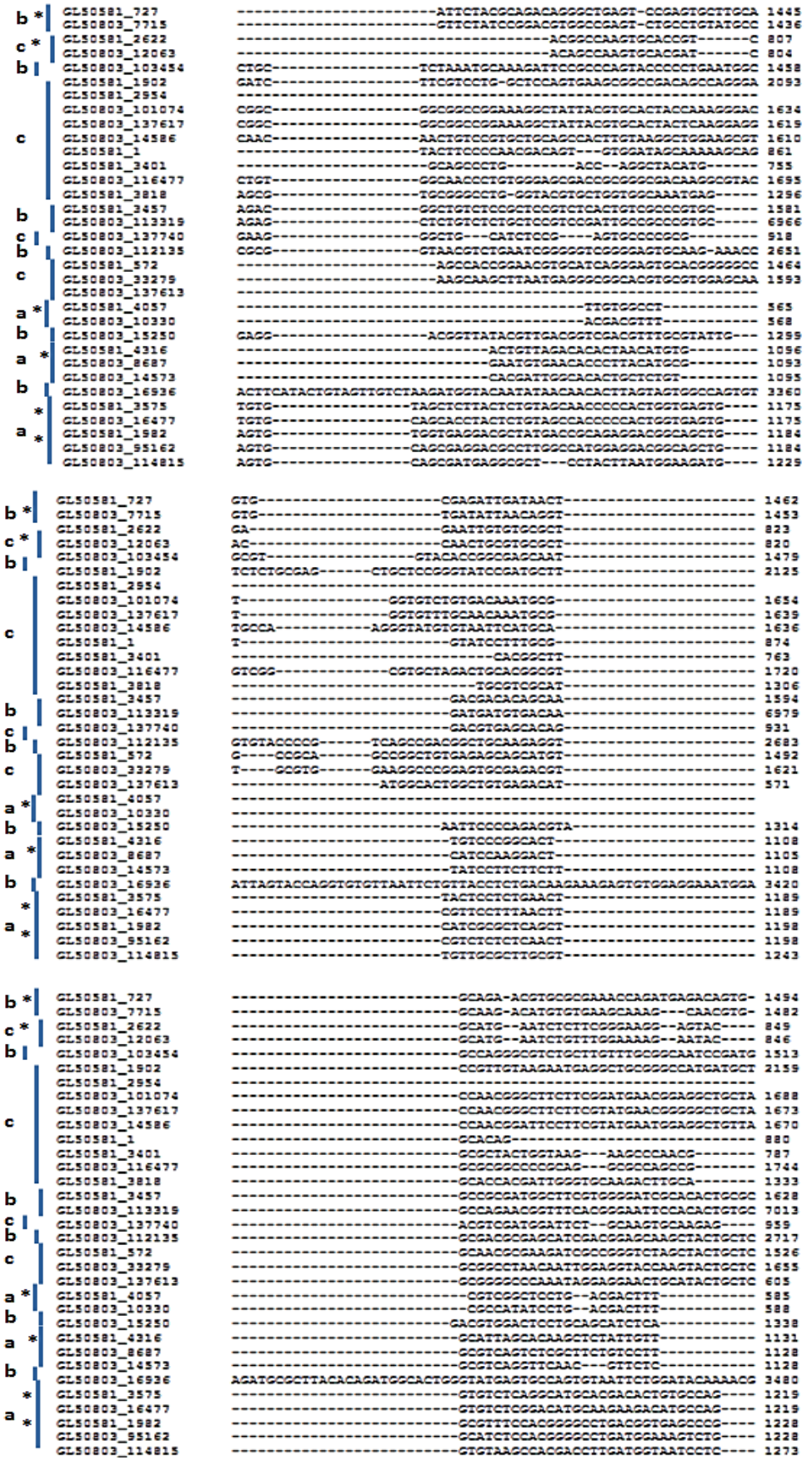

b *
c *
b
c
b
c
b
a *
b
a *
b *
c *
b
c
b
a *
b
a *
b *
c *
b
c
b
a *
b
a *


GL50581_727 ----------CCA 1161
GL50803_7715 ----------CTA 1152
GL50581_2622 ----------
GL50803_12063 ----------
GL50803_103454 ----------GGG 1146
GL50581_1902 ----------AGCCT 1743
GL50581_2954 ----------
GL50803_101074 ----------AAAAC 1308
GL50803_137617 ----------TTCAG 1293
GL50803_14586 ----------ACACT 1275
GL50581_1 ----------AACAA 656
GL50581_3401 ----------CG 636
GL50803_116477 ----------CA 1344
GL50581_3818 ----------CA 1008
GL50581_3457 ----------GATGC 1266
GL50803_113319 ----------GGTGC 6651
GL50803_137740 ----------A 680
GL50803_112135 T----------GTGCC 2127
GL50581_572 ----------ACGGC 1134
GL50803_33279 ----------G 1278
GL50803_137613 ----------
GL50581_4057 ----------
GL50803_10330 ----------
GL50803_15250 ----------
GL50581_4316 ----------T 858
GL50803_8687 ----------T 855
GL50803_14573 ----------C 858
GL50803_16936 GTGTCAGTGCTCCGAAGGGTATGTGAAAGGTAGAAGCCCGCACGAATGCATCCCGAAGAA 2759
GL50581_3575 ----------
GL50803_16477 ----------
GL50581_1982 ----------
GL50803_95162 ----------
GL50803_114815 ----------
GL50581_727 GACAGCACGCAATGTCTCACTGAATGCACCCATCCACTAAG---------- 1202
GL50803_7715 GATGGAAGGGAGTGGCTTAGCGACTGCGTCTTTCCACTGAG---------- 1193
GL50581_2622 GTACAAAGTCC---------- 639
GL50803_12063 GCACGCAGTCT---------- 639
GL50803_103454 GATCACAGGCAGCGCAAATGCCCTGGGAGCTCCAAAAGAGGATGCAGATGCGAGTGCATG 1206
GL50581_1902 ACCAGTATCCTCCTCAGCGGAGACGGGGAGTGCATCGCCATCGCACAC---------- 1791
GL50581_2954 ----------
GL50803_101074 GGACAGTGCGAGACCTGCAAGGACGGATACTACAAGAAAGAAACCGCC---------- 1356
GL50803_137617 GGACAGTGTGATACCTGCAAGGACGGGTACTACAAGAGCGGAGCCGCC---------- 1341
GL50803_14586 GGGACTTGTAAGACGTGTGACGATGGATTTTACAAGAGTAGCGAAGAA---------- 1323
GL50581_1 G-----TGTGTATCGTGGCGCGACG-AAGCTGCTGGTGTG--GCTGGC---------- 696
GL50581_3401 GGATTCTTCG----CAAAGGGGTCGACTTCTAAAACG---------- 669
GL50803_116477 GGGGCCTGCGACGCCATCGTGGTCGGCAACGACGGCAGCATGCTCTCTTAC--------- 1395
GL50581_3818 GAGTGCACGG---CCGACGGGACGCACTTCACAAAGGATAACG---------- 1048
GL50581_3457 CTGCTCTGCAGCGACACGAGGGCGAACGGCGGATGGACRGGGATCGACGGC--------- 1317
GL50803_113319 CTGCTGTGCAACGACGCGGGGGCAGGCGGAGGATGGACAGGGGTCGACGGC--------- 6702
GL50803_137740 GTGCCCCGCG--GACGTGAGCACAGACGTCG-ATG-----GATTCTGCAAG---------- 723
GL50803_112135 GCCGGCTATGAGAACCAAGACGGAGTCTGCCGCCCCTGTACCACGGTGAAT--------- 2178
GL50581_572 GCCTGCAAGGCGTGCAAGGAAGGR---TACCGCCTTAG--CGGCGGTGCG---------- 1179
GL50803_33279 GCCAGCGGGGA-TAACAGGCGTGCAACCCAGCCCCG--CCGACTGCCCG--------- 1326
GL50803_137613 -----CTG--------AGTGCGT-------TGGCCCG--CCAACAATCAG--------- 453
GL50581_4057 ----------
GL50803_10330 ----------
GL50803_15250 -----TAGATTCTGTGGAGATGATGGAGGGTGCGTTCAGTGTGCTGCAGGCT------TT 1071
GL50581_4316 AACTTTAGACGCAAGGAGTGCAAATGTCCT-----ACAGAGTATGCTGG---------AG 904
GL50803_8687 AACTTTAGACGCAAGGAATGTAAGTGCCCC-----ACAGAGTACAGCGG---------AG 901
GL50803_14573 AATTACAAAACAGGCGTCTGCGAATGTTCT-----ATAGCGTACGTTGG---------CA 904
GL50803_16936 ATGCACATATAGAAGATATGAGAGTGACCTGGTG-ACGGAGTGTGGCGGACT------GG 2812
GL50581_3575 CAGTGCATTGTGCTGGGGCAAGTCTGTCCTCATGGTTCCTGTAAAAACACAG------AT 951
GL50803_16477 CAGTGCATTGTGCTGGGACAGGTTTGCCCCCATGGTACTTGTACAAATACCG------AC 951
GL50581_1982 GCCTGCCTCGTGGACGGCGAGGTCTGCCCCCACGGCGAGTGTGTAATTTCTA------GG 960
GL50803_95162 GTTTGTCTTGTTGGCGGAGAAGTTTGCCCTCATGGAGAGTGCACGGTCTTCA------TG 960
GL50803_114815 AACTGCATCGTCGACGGAGAGGTGTGCCCGTACGGGACATGTGAGCGCGAAG------GC 1008
GL50581_727 --GAAGTATTGCAGTAA----------TGACATTTGCATATGTG---- 1234
GL50803_7715 --AAAGTATTGCAGTAA----------TGATGTTTGTATTTGTG---- 1225
GL50581_2622 -----GTATTGAAG----------GACGGGAAGTGTGTTT---- 664
GL50803_12063 -----GTGTTAAAG----------GACGGAAAGTGCGTTT---- 664
GL50803_103454 GACAAGTACTACAGGAG----------CGGCGAGCAGTGCTCTG---- 1240
GL50581_1902 CTCGGGCGCTGCAA----------GGACGCTCTATGGCCCG---- 1822
GL50581_2954 ----------
GL50803_101074 TGTGCCAAGTGCGA----------TACCTCGTGCAAGACG----- 1386
GL50803_137617 TGTGCCAAGTGCGA----------TACCTCGTGTAAGACA----- 1371
GL50803_14586 TGCAAGGCATGCGA----------CTCCAACTGCAAAACA----- 1353
GL50581_1 TGCGCTGAGTGC-A----------CAGCCCCTACAGAGCC----- 725
GL50581_3401 TGCGAGACGTGTGA----------CGTCGCG---------- 690
GL50803_116477 TGTTCAAAGTGTGT----------TGGCGCCGGCTATGGCC---- 1426
GL50581_3818 TTCAAATAAATAGT----------AGTCCAATTAAGGTGTG---- 1079
GL50581_3457 TGCAAGACCTGCCG----------CCTCGCAGACAGCGTGG---- 1348
GL50803_113319 TGCCAGAGGTGCCG----------CCTCGCTGATCGGCTGG---- 6733
GL50803_137740 TGCAAGAGCGGGT----------ACACGCCGTCGACG-------- 750
GL50803_112135 TGCGAGCGGTGCCGGATGGATGACCAGGGGGAGGTGTGCACACACTGCATGGTCGGGTAC 2238
GL50581_572 TGCGAGCAGTGCA----------CCGTGG----------CAAGCTGCAAGG--------C 1211
GL50803_33279 ATCGAGAACTGCCAG----------ACGTGCTCCTCGG--------- 1354
GL50803_137613 ----TGTAGTTCCTG----------TCCTGCT---------- 471
GL50581_4057 ----------TGCGTACCC---------- 486
GL50803_10330 ----------TGCGCACCT---------- 489
GL50803_15250 ATGCTCGATAATAA----------GGAGTGCGTTCCTGTCAA---- 1103
GL50581_4316 ACCAATGTGAAC----------TGTGTTCCTCTGGGGC---- 932
GL50803_8687 AACAGTGCGAAG----------ACTGTTCTCCTGGAGC---- 929
GL50803_14573 GTCTCTGCACGG----------AGTGCGCAGATACGGC---- 932
GL50803_16936 GTCGATGCGAGGGTGATAAACT-TATCACCAAGAAATGTAAGTGTAACGATGGTGC---- 2867
GL50581_3575 GGCACGTGGGCCTG----------CGTGTGCGATAGAATATA---- 983
GL50803_16477 GGTATATGGGGCTG----------TACCTGCAACAAAGCATA---- 983
GL50581_1982 GGCACCGCACGCTG----------CAAATGCAACCCAGAATA---- 992
GL50803_95162 AACCAATCGTACTG----------CAAGTGTGATGCAGAGTA---- 992
GL50803_114815 ATGTCCTGGACGTG----------TGTATGCCGCGCAGAGTA---- 1040


b *
c *
b
c
b
c
b
c
a *
b
a *
b
*
a *
GL50581_727
GL50803_7715
GL50581_2622
GL50803_12063
GL50803_103454
GL50581_1902
GL50581_2954
GL50803_101074
GL50803_137617
GL50803_14586
GL50581_1
GL50581_3401
GL50803_116477
GL50581_3818
GL50581_3457
GL50803_113319
GL50803_137740
GL50803_112135
GL50581_572
GL50803_33279
GL50803_137613
GL50581_4057
GL50803_10330
GL50803_15250
GL50581_4316
GL50803_8687
GL50803_14573
GL50803_16936
GL50581_3575
GL50803_16477
GL50581_1982
GL50803_95162
GL50803_114815

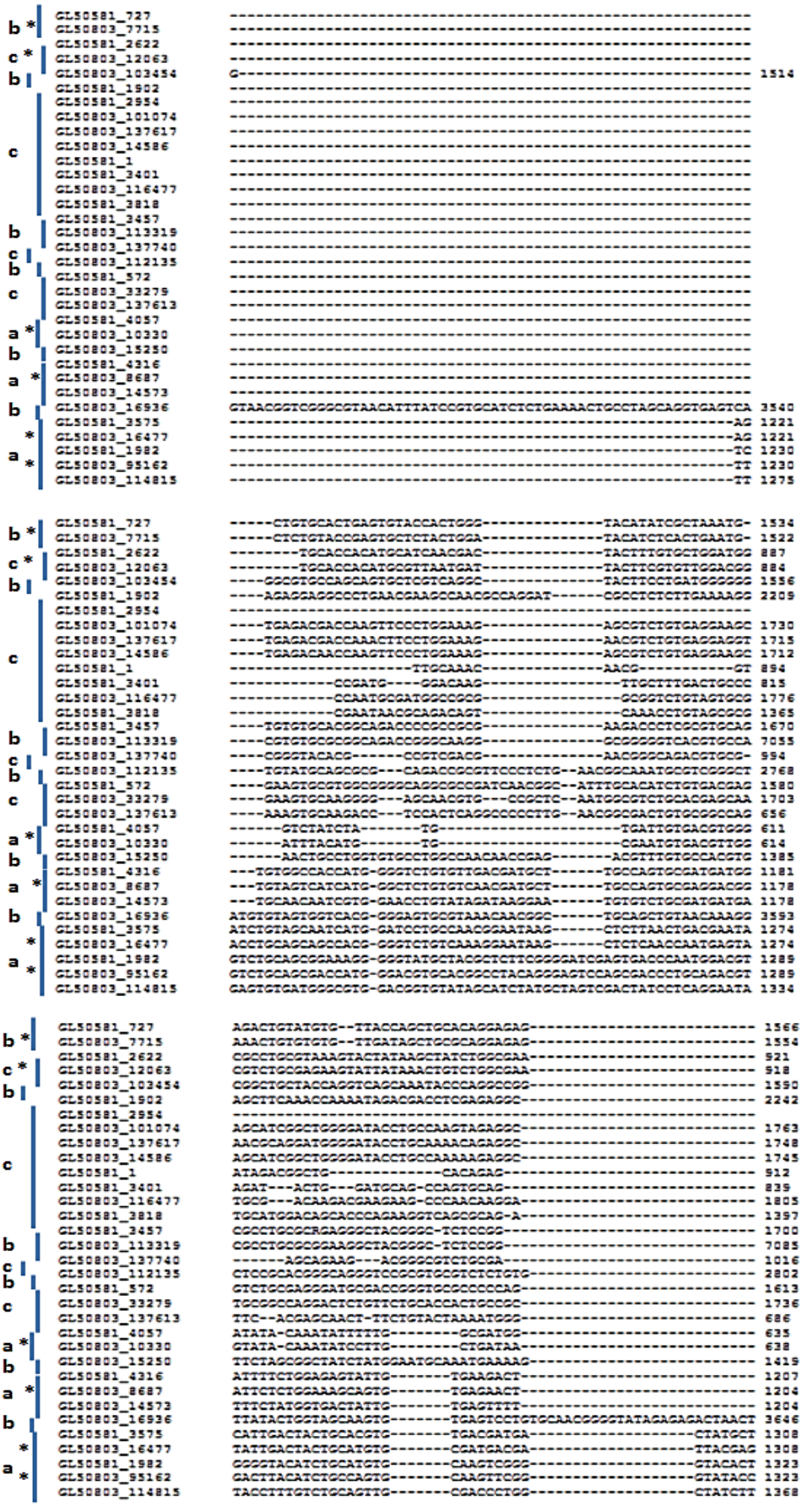

b *
c *
b
c
b
c
b
a *
b
a *
b
*
a *
GL50581_727 CGGAGGTGCTTGTGTGCCGTGTCACGAATCGTGTGGCCTTTGCACAG---------- 1306
GL50803_7715 TGGTAACACCTGCTTAGCGTGCCACGAGTCCTGTGCCTTTTGTACAG---------- 1297
GL50581_2622 -ACACGAACATG---GCTGGTTGCT----------CCACCT---------- 703
GL50803_12063 -GCGAGAACATG---GCCGGGTGCT----------TCAGTT---------- 703
GL50803_103454 TGCAATAACCAGGAGCGCAATTCCTGTACGTTGT----GTCCCTCCG---------- 1306
GL50581_1902 CTGCTTGAGGGGAGGACCATCTCTGCATCTGACATCTGGAGTCTCGG---------- 1913
GL50581_2954 ----------
GL50803_101074 GAGGCTAAGAAGGCCCTCAGCTACGAAGGAGAGGGCAACA-CGGGGA---------- 1468
GL50803_137617 GAGCCTAAGAAGGCCCTCAGCTACGAAGGAGAGGGCAACA-CGGGGA---------- 1453
GL50803_14586 TTAGCGGAGCGG-TACTCAAATATGGGAATGATGGCACTA-AAGGCA---------- 1435
GL50581_1 ----------
GL50581_3401 ----------
GL50803_116477 GGCGTATGTACGCGGTGCACCAACAACTACTTCCTCTATATGGGTGG---------- 1523
GL50581_3818 TATGCAAGAGCG-----AATCAGATACTGCGTCGCCAACATG--CAG---------- 1160
GL50581_3457 AAGCTTACCCTGGACGGAGTGCTCTCGTGCCCCACGCGCCTGCCCGGC---------- 1433
GL50803_113319 GCGCTGACTCTGGACGGGATGCTCCTGTGCCCCCGCGACTGTCCGAC---------- 6818
GL50803_137740 -------CAGAAGACGGG-----CGTCTGCGACGGCTCCTTGACCAC---------- 803
GL50803_112135 GGACCTGCGTGGACAACTGCAGAAGCATTCCCGGGCACTACGCACTTGATGCCACGGAGA 2425
GL50581_572 ---CCAAACTGGCCTGCT-CTGCTACGTGCCCGG----TCGGCTCGT---------- 1309
GL50803_33279 ---CCGACGAAGAAGGCA-TGCTTAGATGCCTGCCCTGTCGGAACGT---------- 1441
GL50803_137613 ---TATAC----------AGATGACAG---------- 500
GL50581_4057 ------TTCACG---------AAGACAGT-------TCTGGTAATCG---------- 521
GL50803_10330 ------TCCACG---------AGGACAAG-------TCTGGTCATCG---------- 524
GL50803_15250 ----GCTGCACCGTCACAGAGAGAACAGAG------ACTTGCAAGCG---------- 1166
GL50581_4316 -----AGACAACGCTTGTATTACTGATG---------AACCTGATGG---------- 989
GL50803_8687 -----CGATAATGCATGCATTACAGACG---------AGTCGGATGG---------- 986
GL50803_14573 -----CGACAAAGCATGGGTCACAGATG---------AGCCAGATGG---------- 989
GL50803_16936 TGTGACGGCAACGCTTGCATCTGTGATGGAAATCTTGTCCCGGTTGG---------- 3026
GL50581_3575 -----TCCGAAGAGTTGCATCACTAGCTAC------AGCCCTATAGG---------- 1043
GL50803_16477 -----CCCGAAAAGCTGCATCACTGACTAT------AGCCCTATAGG---------- 1043
GL50581_1982 -----TCCCAATTCTTGTGTCAGCAAGACG------TTCGACTACGG---------- 1052
GL50803_95162 -----TCCCAACTCCTGCATCAGCGCGACG------TTTGACTACGG---------- 1052
GL50803_114815 -----CCCCAATTCCTGTGTCACTGACATC------CTTCCAAATGG---------- 1100
b *
c *
b
c
b
c
b
a *
b
a *
b
*
a *
GL50581_727 --------GACCTACAGAAAACGACTGCGA---------- 1328
GL50803_7715 --------GGCCATCGGCGGACCAGTGTGA---------- 1319
GL50581_2622 --------GCCTATCGCCGGCAGACATGG---------- 724
GL50803_12063 --------GCCTTTCGCCAGCAGATACGG---------- 724
GL50803_103454 --------GCCAGGCACTTGCATACAATGG---------- 1328
GL50581_1902 TGTGATCATCTATGAGCTTGCCACTCTGAAG---------- 1944
GL50581_2954 ----------
GL50803_101074 CGTGCAAATCAGAATGCAAACCGGGTACT---------- 1497
GL50803_137617 CGTGCAAATCAGAATGCAAACCGGGTACT---------- 1482
GL50803_14586 CATGCGGAGCTGGATGCGCAACGGGCACAGG---------- 1466
GL50581_1 -TACCGGTGATAATTACCTAAAG---------- 774
GL50581_3401 ----------TGTAACAGACTGC---------- 711
GL50803_116477 TTGCTATGATGTGTCGAAGGCTCCTGGCA---------- 1552
GL50581_3818 CGTCTGTGAGGCAACGAAGGCTCCT-AAC---------- 1188
GL50581_3457 AGACGGGTTCTGGACGAACAGCAGTGAGTG---------- 1463
GL50803_113319 AGAGGGCTTCTGGACAAGCGGCAGCGAATG---------- 6848
GL50803_137740 CGCGGG---CTGCGCGACCTGCACCCCGGC---------- 830
GL50803_112135 AGAGCCCCGCAAGGTGCGCTCTGTGTGAGAT---------- 2456
GL50581_572 ------ACACTAGCGGGCGGCAGTG---------- 1328
GL50803_33279 ------ATCTATCCGGACAAACGTGCG---------- 1462
GL50803_137613 --------CAATCCTAATAACGGAGG---------- 518
GL50581_4057 --------TAGCATCTG---------- 530
GL50803_10330 --------TAGCATTTG---------- 533
GL50803_15250 --------TTGCCAAAAGGGATATT---------- 1183
GL50581_4316 --------TAGCATAAGTGTTTGCA---------- 1006
GL50803_8687 --------TGGTGTAAGTGTCTGTA---------- 1003
GL50803_14573 --------GAGCATCAGCGTCTGTA---------- 1006
GL50803_16936 --------AAATGTTTGCGTAAGCAGTAAGTGCGTTGTTCAACTGGCAAGTGGAAGCACT 3078
GL50581_3575 --------TGCGCCAGTGCTATGCT---------- 1060
GL50803_16477 --------CGAACCAGTGTTATGCT---------- 1060
GL50581_1982 --------CTACCCTGAACTCTGCT---------- 1069
GL50803_95162 --------CTACCCTGAGCTCTGCT---------- 1069
GL50803_114815 --------AGATCCCAAGCTCTGCT---------- 1117
b *
c *
b
c
b
c
b
a *
b
a *
b
*
a *
GL50581_727 ----------AATCTGTGCGACAGG---------- 1343
GL50803_7715 ----------GGCGTGTACGCAGGG---------- 1334
GL50581_2622 ----------GGACCGCG---------- 732
GL50803_12063 ----------GAACATC---------- 731
GL50803_103454 ----------CGACGGGACTACAGGG---------- 1344
GL50581_1902 ----------AGGCTG-GACTTCCTAGAAGGAAAG---------- 1968
GL50581_2954 ----------
GL50803_101074 ----------AACA----ACTGCGAGAAGTGCGAG---------- 1518
GL50803_137617 ----------AACA----ACTGCGAGAAGTGCGAG---------- 1503
GL50803_14586 ----------AGCAGGAGCATGCAAGACGTGCGGC---------- 1491
GL50581_1 ----------
GL50581_3401 ----------ACCACATGTGCACC---------- 725
GL50803_116477 ----------ACCTCATGTGCTCTGCG---------- 1569
GL50581_3818 ----------GCTGCAGGAGGTCAGTG---------- 1205
GL50581_3457 ----------TGCTCCCTGCGCGTCGGG---------- 1481
GL50803_113319 ----------CATCCCCTGCGCGGCTGG---------- 6866
GL50803_137740 ----------GGGCTCCAGCCAGACG---------- 846
GL50803_112135 ----------CCCGAACTGCAAAGAGTGTGGAGCCGGTAAGCTTTG 2492
GL50581_572 ----------CCTCCCCTGCGACGCGAGCTGCGCC---------TCG 1356
GL50803_33279 ----------CGCCCTGCGACCCCTCCTGTGCG---------GAG 1488
GL50803_137613 ----------CACCTG---------- 524
GL50581_4057 ----------
GL50803_10330 ----------
GL50803_15250 ----------ATGTTGTTGGGGG---------- 1196
GL50581_4316 ----------ATGGACATGG---------- 1016
GL50803_8687 ----------ACGGACACGG---------- 1013
GL50803_14573 ----------ACGGTCATGG---------- 1016
GL50803_16936 TCTGTGTGTGGAGGTAGTACTGTGGGGACATGCACAAAGGGTCAAGGTGAAGACCACTAC 3138
GL50581_3575 ----------CTGATCACGG---------- 1070
GL50803_16477 ----------CTGGTCACGG---------- 1070
GL50581_1982 ----------CCAACAGGGG---------- 1079
GL50803_95162 ----------CCAACAAGGG---------- 1079
GL50803_114815 ----------CTGGAGTGGG---------- 1127

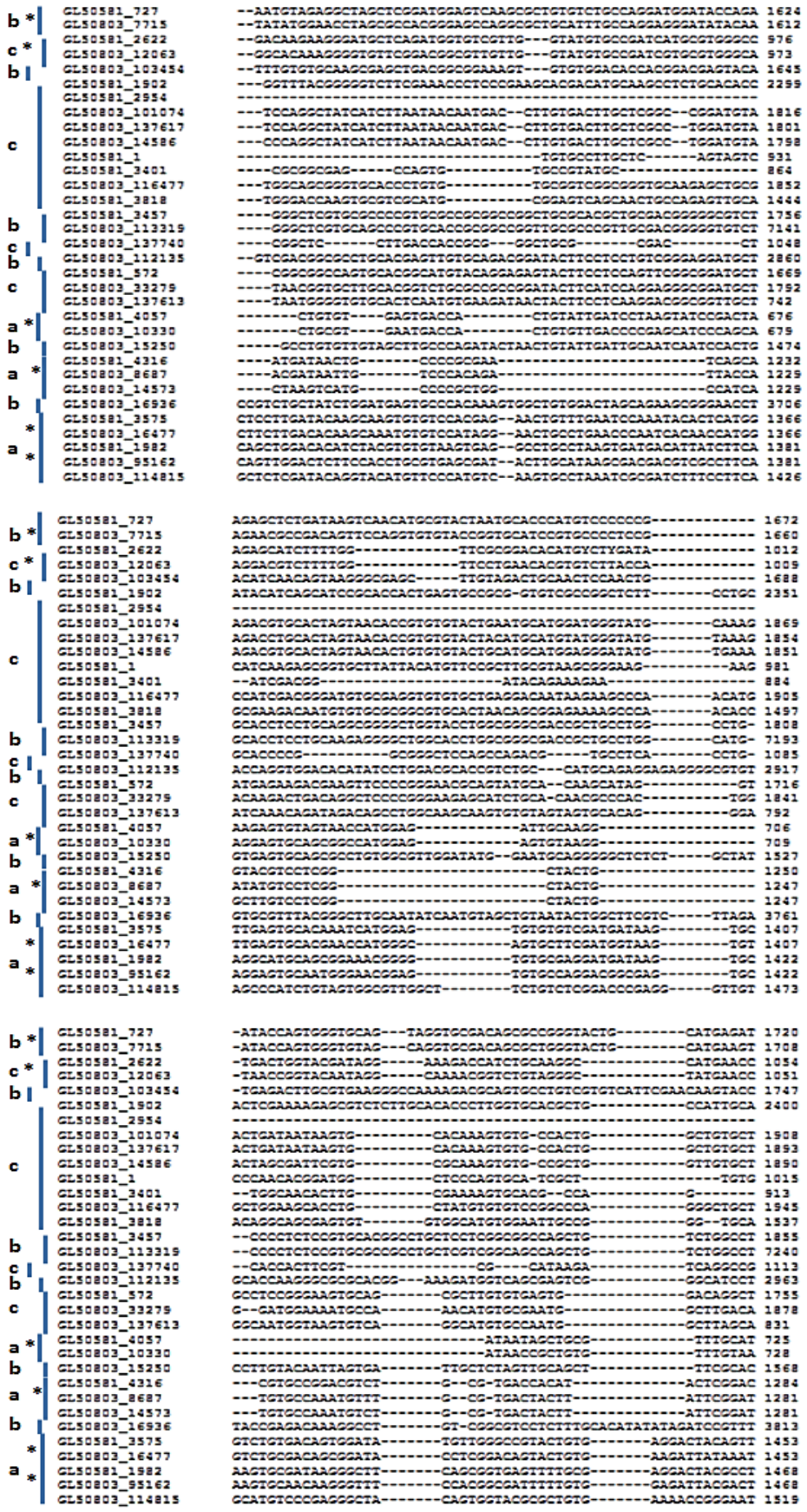

b *
c *
b
c
b
c
b
a *
b
a *
b
*
a *
GL50581_727
GL50803_7715
GL50581_2622
GL50803_12063
GL50803_103454
GL50581_1902
GL50581_2954
GL50803_101074
GL50803_137617
GL50803_14586
GL50581_1
GL50581_3401
GL50803_116477
GL50581_3818
GL50581_3457
GL50803_113319
GL50803_137740
GL50803_112135
GL50581_572
GL50803_33279
GL50803_137613
GL50581_4057
GL50803_10330
GL50803_15250
GL50581_4316
GL50803_8687
GL50803_14573
GL50803_16936
GL50581_3575
GL50803_16477
GL50581_1982
GL50803_95162
GL50803_114815

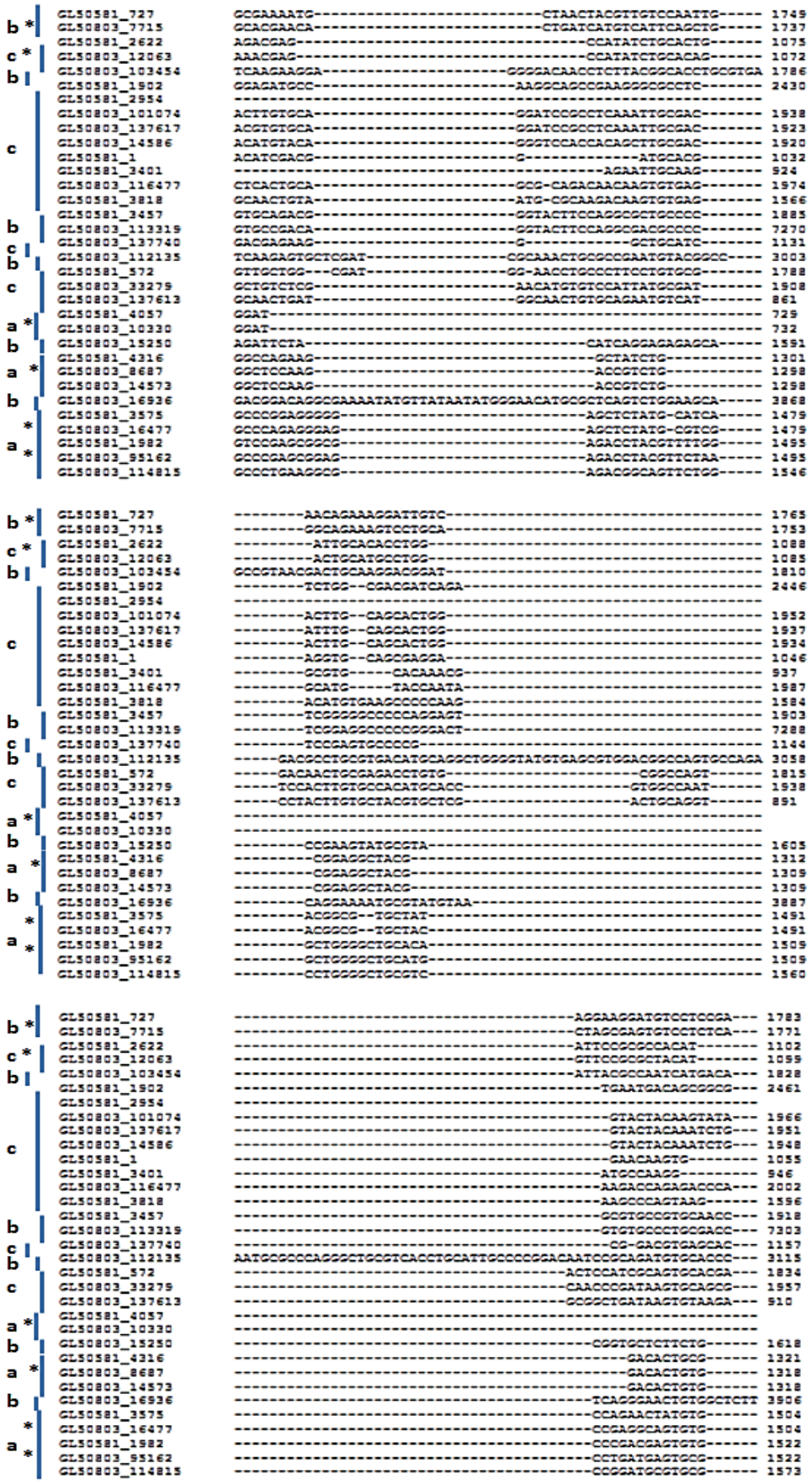

GL50581_727 ----------ATTCTACGCAGACAGGGCTGAGT-CGGAGTGCTTGCA 1445
GL50803_7715 ----------GTTCTATCGGGACGTCGGCGAGT-CTGCCTGTATGCC 1436
GL50581_2622 ----------ACGGCCAAGTGCACCGT------C 807
GL50803_12063 ----------ACAGCCAAGTGCACGAT------C 804
GL50803_103454 CTGC----------TCTAAATGCAAAGATTCCGCCCAGTACCCCCTGAATGCC 1458
GL50581_1902 GATC----------TTCGTCCTG-GCTCCAGTGAAGCGGCCGCACAGCCAGGCA 2093
GL50581_2954 ----------
GL50803_101074 CGGC----------GGCGGCCGGAAAGGCTATTACGTGCACTACCAAAGGGAC 1634
GL50803_137617 CGGC----------GGCGGCCGGAAAGGCTATTACGTGCACTACTCAAGAGG 1619
GL50803_14586 CAAC----------AACTGTCCGTGCTGCAGCCACTTGTAAGGCTGGAAGCGT 1610
GL50581_1 ----------TACTTCCCCAACGACAGT---GTGGATAGCAAAAAGCAG 861
GL50581_3401 ----------GCAGCCCTG--------ACC--AGGCTACATG------ 755
GL50803_116477 CTGT----------GGCAACCCTGTGGGAGCGACCGCGGGCGACAAGCGTAC 1695
GL50581_3818 AGCG----------TGCGGGCCTG-GGTACGTGCTGGTGGCAAATGAG----- 1296
GL50581_3457 AGAC----------GGCTGTCTCCGCTCCGTCTCACTGTCGCCCGTGC----- 1581
GL50803_113319 AGAG----------CTCTGTCTCTGCTCCGTCCGATTGCCGCCCCGTGC----- 6966
GL50803_137740 GAAG----------GGCTG---CATCTCCG----AGTGCCCCGCG-------- 918
GL50803_112135 CGCG----------GTAACGTCTGAATCGGGGGTCGGGGAGTGCAAG-AAACC 2651
GL50581_572 ----------AGCCACCGGAACGTGCATCAGGGAGTGCACGGGGGCC 1464
GL50803_33279 ----------AAGCAAGCTTAATGAGGGCGGCACGTGCGTGGAGCAA 1593
GL50803_137613 ----------
GL50581_4057 ----------TTGTGGCCT----------- 565
GL50803_10330 ----------ACGACGTTT----------- 568
GL50803_15250 GAGG----------ACGGTTATACGTTGACGGTCGACGTTTGCGTATTG---- 1299
GL50581_4316 ----------ACTGTTAGACACACTAACATGTG-------- 1096
GL50803_8687 ----------GAATGTGAACACCCTTACATGCG-------- 1093
GL50803_14573 ----------CACGATTGGCACACTGCTCTGT--------- 1095
GL50803_16936 ACTTCATACTGTAGTTGTCTAAGATGGTACAATATAACAACACTTAGTAGTGGCCAGTGT 3360
GL50581_3575 TGTG----------TAGCTCTTACTCTGTAGCAACCCCCACTGGTGAGTG----- 1175
GL50803_16477 TGTG----------CAGCACCTACTCTGTAGCCACCCCCACTGGTGAGTG----- 1175
GL50581_1982 AGTG----------TGGTGAGGACGCTATGACCGCAGAGGACGGCAGCTG----- 1184
GL50803_95162 AGTG----------CAGCGAGGACGCCTTGGCCATGGAGGACGGCAGCTG----- 1184
GL50803_114815 AGTG----------CAGCGATGAGGCGCT---CCTACTTAATGGAAGATG----- 1229

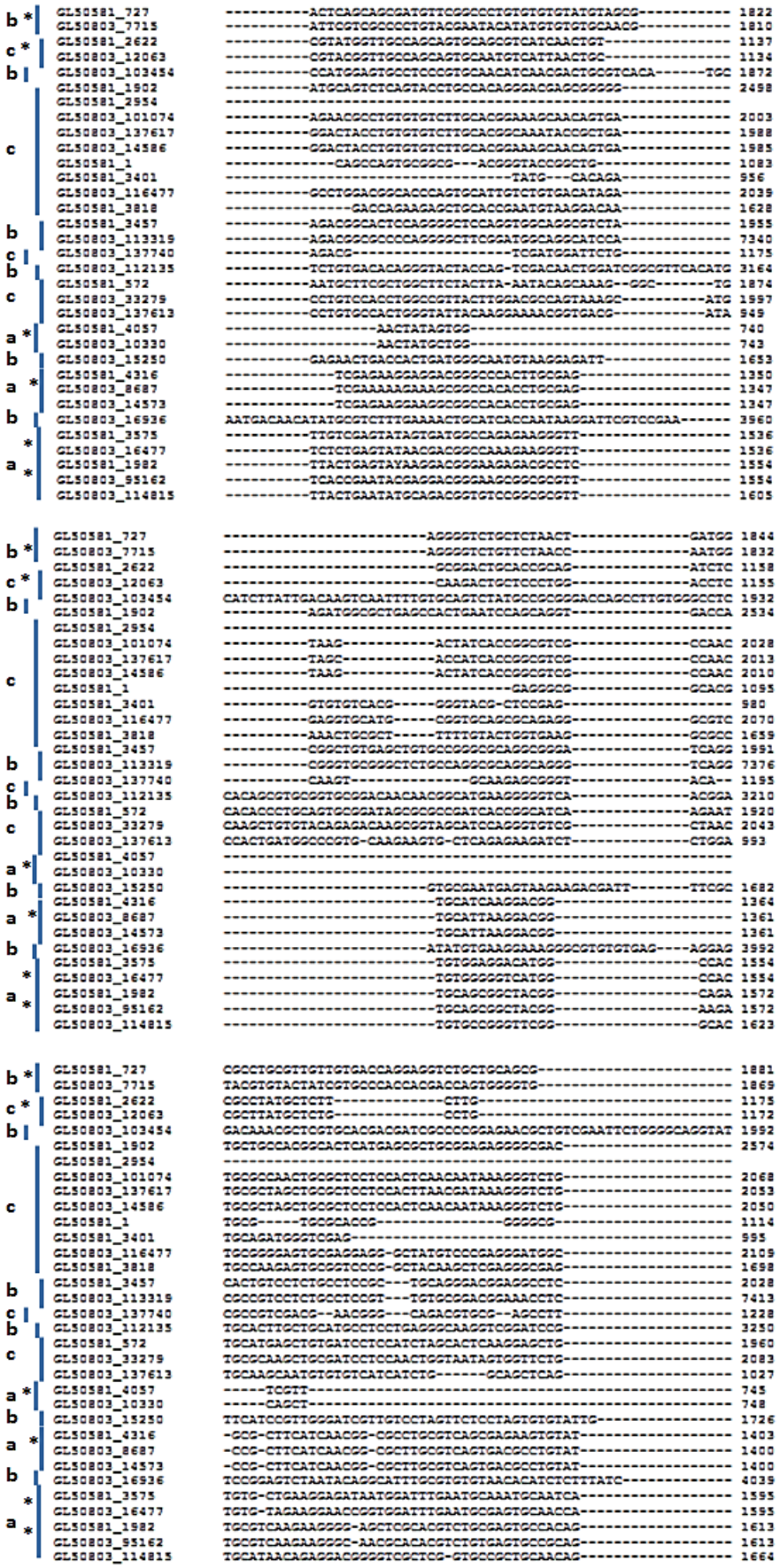
GL50581_727 GTG----------CGAGATTGATAACT---------- 1462
GL50803_7715 GTG----------TGATATTAACAGGT---------- 1453
GL50581_2622 GA----------GAATTGTGTGCGCT---------- 823
GL50803_12063 AC----------CAACTGCGTGCGCT---------- 820
GL50803_103454 GCGT----------GTACACCGGCGAGCAAT---------- 1479
GL50581_1902 TCTCTGCGAG------CTGCTCCGGGTATCCGATGCTT---------- 2125
GL50581_2954 ----------
GL50803_101074 T----------GGTGTCTGTGACAAATGCG---------- 1654
GL50803_137617 T----------GGTGTTTGCAACAAATGCG---------- 1639
GL50803_14586 TGCCA----------AGGGTATGTGTAATTCATGCA---------- 1636
GL50581_1 T----------GTATCCTTTGCG---------- 874
GL50581_3401 ----------CACGGCTT---------- 763
GL50803_116477 GTCGG----------CGTGCTAGACTGCACGGCGT---------- 1720
GL50581_3818 ----------TGCGTCGCAT---------- 1306
GL50581_3457 ----------GACGACACAGCAA---------- 1594
GL50803_113319 ----------GATGATGTGACAA---------- 6979
GL50803_137740 ----------GACGTGAGCACAG---------- 931
GL50803_112135 GTGTACCCCG------TCAGCCGACGGCTGCAAGAGGT---------- 2683
GL50581_572 G----CCGCA------GCCGGCTGTGAGAGCAGCATGT---------- 1492
GL50803_33279 T----GCGTG------GAAGGCCCGGAGTGCGAGACGT---------- 1621
GL50803_137613 ----------ATGGCACTGGCTGTGAGACAT---------- 571
GL50581_4057 ----------
GL50803_10330 ----------
GL50803_15250 ----------AATTCCCCAGACGTA---------- 1314
GL50581_4316 ----------TGTCCCGGCACT---------- 1108
GL50803_8687 ----------CATCCAAGGACT---------- 1105
GL50803_14573 ----------TATCCTTCTTCTT---------- 1108
GL50803_16936 ATTAGTACCAGGTGTGTTAATTCTGTTACCTCTGACAAGAAAGAGTGTGGAGGAAATGGA 3420
GL50581_3575 ----------TACTCCTCTGAACT---------- 1189
GL50803_16477 ----------CGTTCCTTTAACTT---------- 1189
GL50581_1982 ----------CATCGCGCTCAGCT---------- 1198
GL50803_95162 ----------CGTCTCTCTCAACT---------- 1198
GL50803_114815 ----------TGTTGCGCTTGCGT---------- 1243

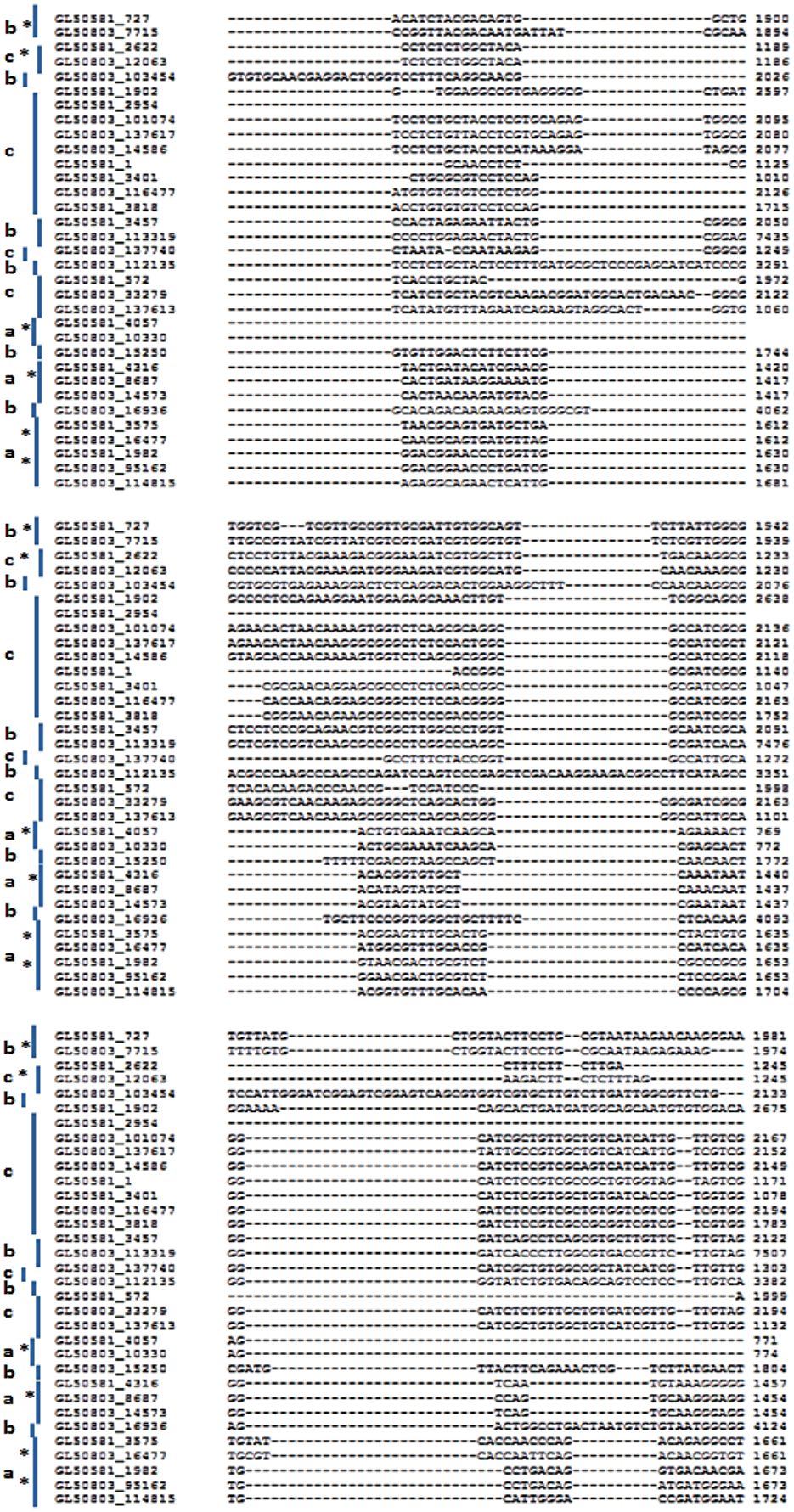
GL50581_727 ----------GCAGA-ACGTGCGCGAAACCAGATGAGACAGTG- 1494
GL50803_7715 ----------GCAAG-ACATGTGTGAAGCAAAG---CAACGTG- 1482
GL50581_2622 ----------GCATG--AATCTCTTCGGGAAGG--AGTAC---- 849
GL50803_12063 ----------GCATG--AATCTGTTTGGAAAAG--AATAC---- 846
GL50803_103454 ----------GCCAGGGCGTCTGCTTGTTTGCGGCAATCCGATG 1513
GL50581_1902 ----------CCGTTGTAAGAATGAGGCTGCGGGCCATGATGCT 2159
GL50581_2954 ----------
GL50803_101074 ----------CCAACGGGCTTCTTCGGATGAACGGAGGCTGCTA 1688
GL50803_137617 ----------CCAACGGGCTTCTTCGTATGAACGGGGGCTGCTA 1673
GL50803_14586 ----------CCAACGGATTCCTTCGTATGAATGGAGGCTGTTA 1670
GL50581_1 ----------GCACAG---------- 880
GL50581_3401 ----------GCGCTACTGGTAAG---AAGCCCAACG-------- 787
GL50803_116477 ----------GGCGGGCCCCGCAG---GCGCCAGCCG-------- 1744
GL50581_3818 ----------GCACCACGATTGGGTGCAAGACTTGCA-------- 1333
GL50581_3457 ----------GCCGCGATGGCTTCGTGGGGATCGCACACTGCGC 1628
GL50803_113319 ----------GCCAGAACGGTTTCACGGGAATTCCACACTGTGC 7013
GL50803_137740 ----------ACGTCGATGGATTCT--GCAAGTGCAAGAG---- 959
GL50803_112135 ----------GCGACGCGAGCATCGACGGAGCAAGCTACTGCTC 2717
GL50581_572 ----------GCAACGCGAAGATCGCCGGGTCTAGCTACTGCTC 1526
GL50803_33279 ----------GCGGCCTAACAATTGGAGGTACCAAGTACTGCTC 1655
GL50803_137613 ----------GCGGGGCCCAAATAGGAGGAACTGCATACTGCTC 605
GL50581_4057 ----------CGTCGGCTCCTG--ACGACTTT---------- 585
GL50803_10330 ----------CGCCATATCCTG--ACGACTTT---------- 588
GL50803_15250 ----------GACGTGGACTCCTGCAGCATCTCA---------- 1338
GL50581_4316 ----------GCATTAGCACAAGCTCTATTGTT---------- 1131
GL50803_8687 ----------GGGTCAGTCTCGCTTCTGTCCTT---------- 1128
GL50803_14573 ----------GGGTCAGGTTCAAC---GTTCTC---------- 1128
GL50803_16936 AGATGCGCTTACACAGATGGCACTGGGTATGAGTGCCAGTGTAATTCTGGATACAAAACG 3480
GL50581_3575 ----------GTGTCTCAGGCATGCACGACACTGTGCCAG---- 1219
GL50803_16477 ----------GTGTCTCGGACATGCAAGAAGACATGCCAG---- 1219
GL50581_1982 ----------GGGTTTCCACGGGGGCCTGACGGTGAGCCCG---- 1228
GL50803_95162 ----------GCATCTCCACGGGGGCCTGATGGAAAGTCTG---- 1228
GL50803_114815 ----------GTGTAAGCCACGACCTTGATGGTAATCCTC---- 1273

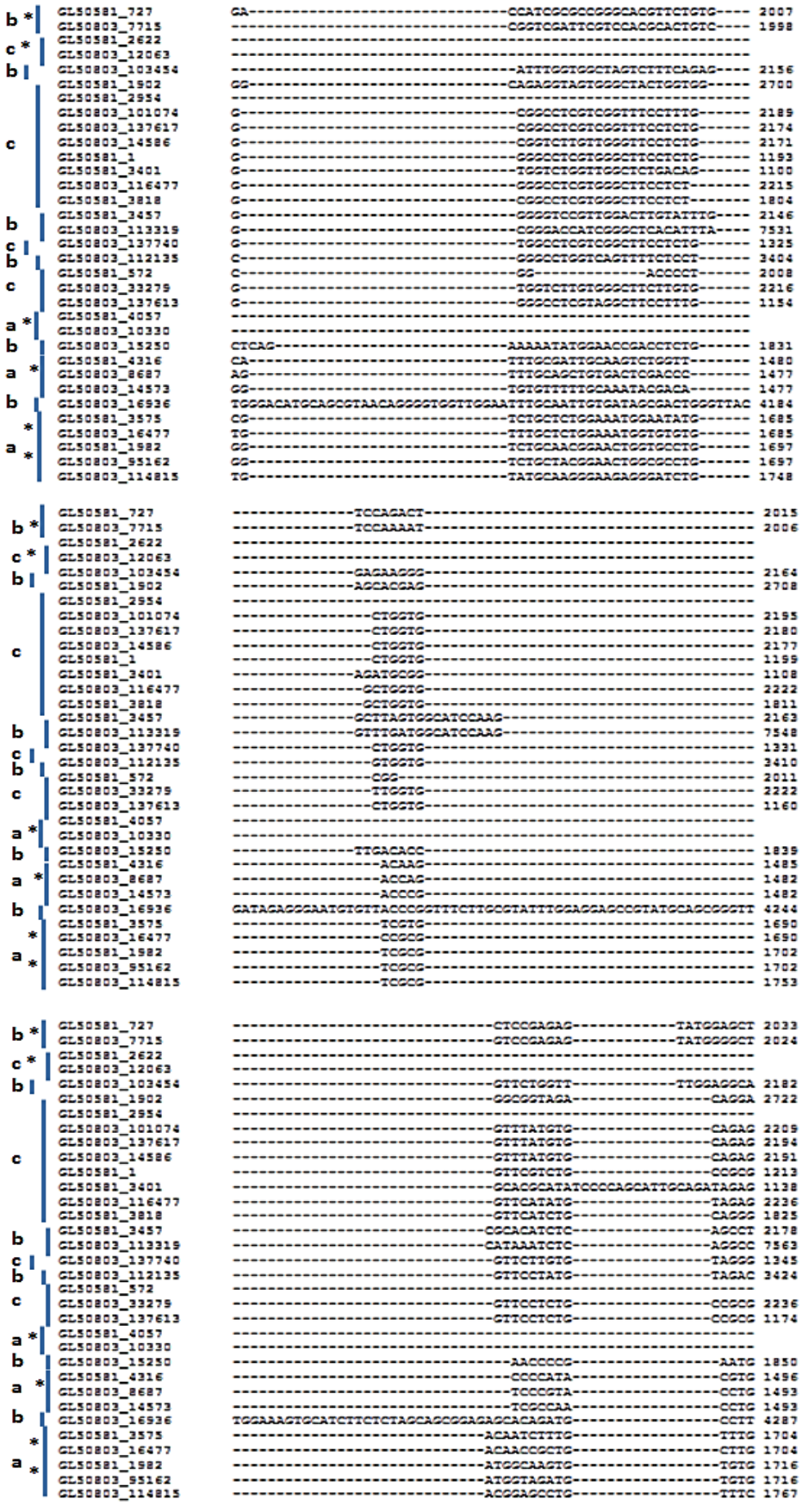

b * GL50581_727 ----------
GL50803_7715 ----------
GL50581_2622 ----------
c * GL50803_12063 ----------
b GL50803_103454 G---------- 1514
GL50581_1902 ----------
GL50581_2954 ----------
GL50803_101074 ----------
GL50803_137617 ----------
c GL50803_14586 ----------
GL50581_1 ----------
GL50581_3401 ----------
GL50803_116477 ----------
GL50581_3818 ----------
GL50581_3457 ----------
b GL50803_113319 ----------
c GL50803_137740 ----------
b GL50803_112135 ----------
GL50581_572 ----------
c GL50803_33279 ----------
GL50803_137613 ----------
GL50581_4057 ----------
a * GL50803_10330 ----------
b GL50803_15250 ----------
GL50581_4316 ----------
a * GL50803_8687 ----------
GL50803_14573 ----------
b GL50803_16936 GTAACGGTCGGGCGTAACATTTATCCGTGCATCTCTGAAAACTGCCTAGCAGGTGAGTCA 3540
GL50581_3575 ----------AG 1221
* GL50803_16477 ----------AG 1221
a * GL50581_1982 ----------TC 1230
GL50803_95162 ----------TT 1230
GL50803_114815 ----------TT 1275
b * GL50581_727 -----CTGTGCACTGAGTGTACCACTGGG----------TACATATCGCTAAATG- 1534
GL50803_7715 -----CTCTGTACCGAGTGCTCTACTGGA----------TACATCTCACTGAATG- 1522
c * GL50581_2622 --------TGCACCACATGCATCAACGAC----------TACTTTGTGCTGGATGG 887
GL50803_12063 --------TGCACCACATGCGTTAATGAT----------TACTTCGTGTTGGACGG 884
b GL50803_103454 ----GGCGTGCCAGCAGTGCTCGTCAGGC----------TACTTCCTGATGGGGGG 1556
GL50581_1902 ----AGAGGAGGCCCTGAACGAAGCCAACGCCAGGAT------CGCCTCTCTTGAAAAGG 2209
GL50581_2954 ----------
GL50803_101074 ----TGAGACGACCAAGTTCCCTGGAAAG----------AGCGTCTGTGAGGAAGC 1730
GL50803_137617 ----TGAGACGACCAAACTTCCTGGAAAG----------AACGTCTGTGAGGAGGT 1715
c GL50803_14586 ----TGAGACAACCAAGTTCCCTGGAAAG----------AGCGTCTGTGAGGAAGC 1712
GL50581_1 ----------TTGCAAAC----------AACG----------GT 894
GL50581_3401 ----------CCGATG----GGACAAG----------TTGCTTTGACTGCCC 815
GL50803_116477 ----------CCAATGCGATGGCCGCG----------GCGGTCTGTAGTGCG 1776
GL50581_3818 ----------CGAATAACGCAGACAGT----------CAAACCTGTAGCGCG 1365
GL50581_3457 ----TGTGTGCACGGCAGACCCGCCGCG----------AAGACCCTCGCGTGCAG 1670
b GL50803_113319 ----CGTGTCGCGCGGCAGACCGGGCAAGG----------GCGGGGGTCACGTGCCA 7055
c GL50803_137740 ----CGGGTACACG------CCGTCGACG----------AACGGGCAGACGTGCG- 994
b GL50803_112135 ----TGTATGCAGCGCG---CAGACCGCGTTCCCTCTG--AACGGCAAATGCGTCGGGCT 2768
GL50581_572 ----GAAGTGCGTGGCGGGGCAGGCGCCGATCAACGGC--ATTTGCACATCTGTGACGAG 1580
c GL50803_33279 ----GAAGTGCAAGGGG---AGCAACGTG---CCGCTC--AATGGCGTCTGCACGAGCAA 1703
GL50803_137613 ----AAAGTGCAAGACC---TCCACTCAGGCCCCTTG--AACGGCGACTGTGCGGCCAG 656
GL50581_4057 ------GTCTATCTA-------TG----------TGATTGTGACGTGGG 611
a * GL50803_10330 ------ATTTACATG-------TG----------CGAATGTGACGTTGG 614
b GL50803_15250 ------AACTGCCTGGTGTGCCTGGCCAACAACCGAG-------ACGTTTGTGCCACGTG 1385
GL50581_4316 ---TGTGGCCACCATG-GGGTCTGTGTTGACGATGCT------TGCCAGTGCGATGATGG 1181
a * GL50803_8687 ---TGTAGTCATCATG-GGCTCTGTGTCAACGATGCT------TGCCAGTGCGAGGACGG 1178
GL50803_14573 ---TGCAACAATCGTG-GAACCTGTATAGATAAGGAA------TGTGTCTGCGATGATGA 1178
b GL50803_16936 ATGTGTAGTGGTCACG-GGGAGTGCGTAAACAACGGC------TGCAGCTGTAACAAAGG 3593
GL50581_3575 ATCTGTAGCAATCATG-GATCCTGCCAACGGAATAAG------CTCTTAACTGACGAATA 1274
* GL50803_16477 ACCTGCAGCAGCCACG-GGGTCTGTCAAAGGAATAAG------CTCTCAACCAATGAGTA 1274
a * GL50581_1982 GTCTGCAGCGGAAAGG-GGGTATGCTACGCTCTTCGGGGATCGAGTGACCCAATGGACGT 1289
GL50803_95162 GTCTGCAGCGACCATG-GGACGTGCACGGCCTACAGGGAGTCCAGCGACCCTGCAGACGT 1289
GL50803_114815 GAGTGTGATGGGCGTG-GACGGTGTATAGCATCTATGCTAGTCGACTATCCTCAGGAATA 1334
b * GL50581_727 AGACTGTATGTG--TTACCAGCTGCACAGGAGAG---------- 1566
GL50803_7715 AAACTGTGTGTG--TTGATAGCTGCGCAGGAGAG---------- 1554
c * GL50581_2622 CGCCTGCGTAAAGTACTATAAGCTATCTGGCGAA---------- 921
GL50803_12063 CGTCTGCGAGAAGTATTATAAACTGTCTGGCGAA---------- 918
b GL50803_103454 CGGCTGCTACCAGGTCAGCAAATACCCAGGCCGG---------- 1590
GL50581_1902 AGCTTCAAACCAAAATAGACGACCTCGAGAGGC---------- 2242
GL50581_2954 ----------
GL50803_101074 AGCATCGGCTGGGGATACCTGCCAAGTAGAGGC---------- 1763
GL50803_137617 AACGCAGGATGGGGATACCTGCAAAACAGAGGC---------- 1748
c GL50803_14586 AGCATCGGCTGGGGATACCTGCCAAAAAGAGGC---------- 1745
GL50581_1 ATAGACGGCTG----------CACAGAG---------- 912
GL50581_3401 AGAT---ACTG---GATGCAG-CCAGTGCAG---------- 839
GL50803_116477 TGCG---ACAAGACGAAGAAG-CCCAACAAGGA---------- 1805
GL50581_3818 TGCATGGACAGCACCCAGAAGGTCAGCGCAG-A---------- 1397
GL50581_3457 CGCCTGCGCRGAGGGCTACGGGC-TCTCCGG---------- 1700
b GL50803_113319 CGCCTGCGCGGAAGGCTACGGGC-TCTCCGG---------- 7085
c GL50803_137740 ------AGCAGAAG---ACGGGCGTCTGCGA---------- 1016
b GL50803_112135 CTCCGCACGGGCAGGGTCCGCGTGCGTCTCTGTG---------- 2802
GL50581_572 GTCTGCGAGGGATGCGACCGGGTGCGCCCCCAG---------- 1613
c GL50803_33279 TGCGGCCAGGACTCTGTTCTGCACCACTGCCGC---------- 1736
GL50803_137613 TTC--ACGAGCAACT-TTCTGTACTAAAATGGG---------- 686
GL50581_4057 ATATA-CAAATATTTTTG--------GCGATGG---------- 635
a * GL50803_10330 GTATA-CAAATATCCTTG--------CTGATAA---------- 638
b GL50803_15250 TTCTAGCGGCTATCTATGGAATGCAAATGAAAAG---------- 1419
GL50581_4316 ATTTTCTGGAGAGTATTG-------TGAAGACT---------- 1207
a * GL50803_8687 ATTCTCTGCAAAGCAGTG-------TGAGAACT---------- 1204
GL50803_14573 TTTCTATGGTGACTATTG-------TGAGTTTT---------- 1204
b GL50803_16936 TTATACTGGTAGCAAGTG-------TGAGTCCTGTGCAACGGGGTATAGAGAGACTAACT 3646
GL50581_3575 CATTGACTACTGCACGTG-------TGACGATGA----------CTATGCT 1308
* GL50803_16477 TATTGACTACTGCATGTG-------CGATGACGA----------TTACGAG 1308
a * GL50581_1982 GGGGTACATCTGCATGTG-------CAAGTCGGG----------GTACACT 1323
GL50803_95162 GACTTACATCTGCCAGTG-------CAAGTTCGG----------GTATACC 1323
GL50803_114815 TACCTTTGTCTGCAGTTG-------CGACCCTGG----------CTATCTT 1368

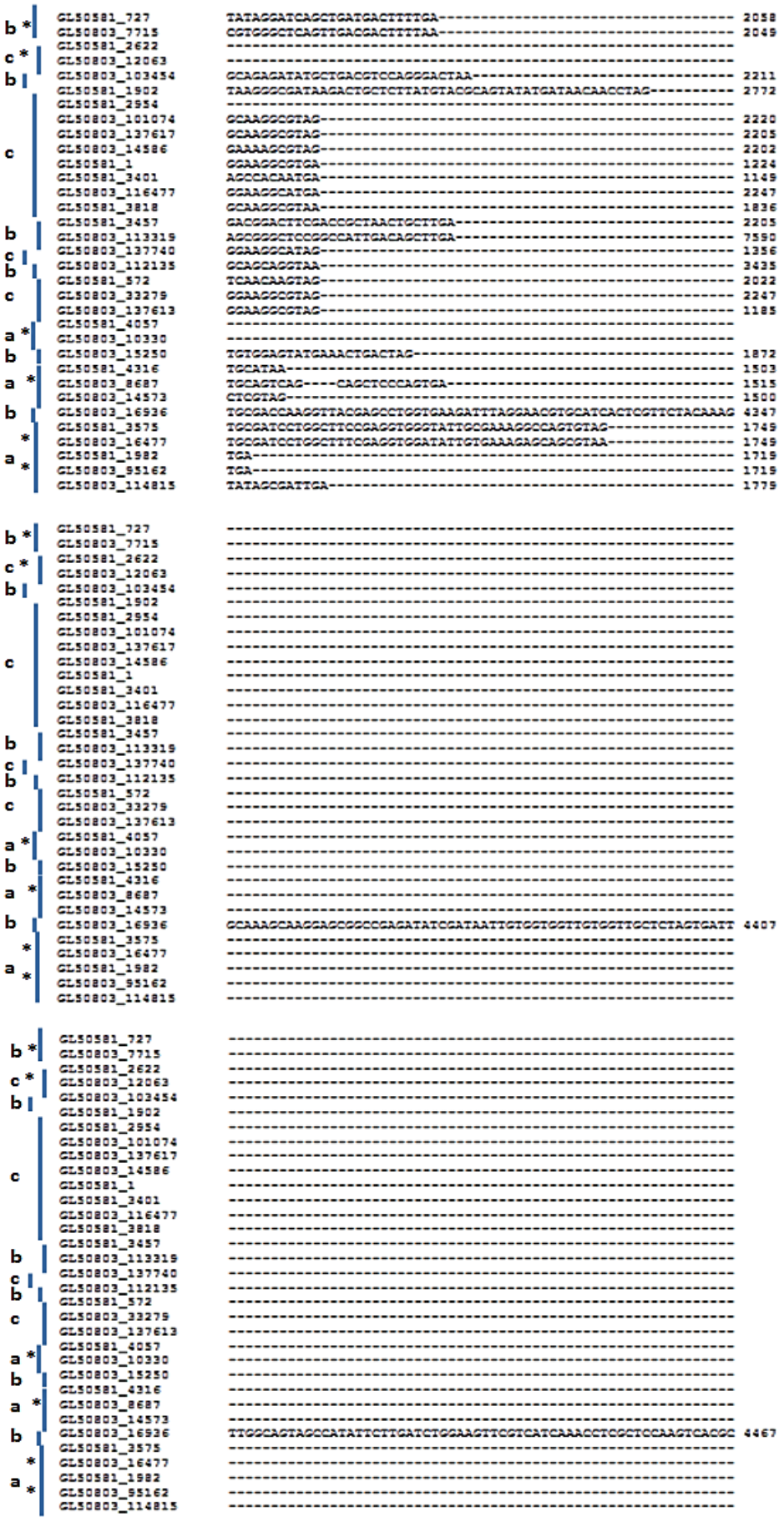

b *
c *
b
c
b
b
c
b
c
a *
b
a *
b
c
b
*
a *
b *
c *
b
c
b
c
a *
b
a *
b *
c *
b
c
b
c
a *
b
a *


b *
c *
b
c
b
b
c
b
c
a *
b
a *
a *
b *
c *
b
c
b
c
b
a *
a *
b *
c *
b
c
b
c
b
a *
a *
b *
c *
b
c
b
c
b
a *
a *



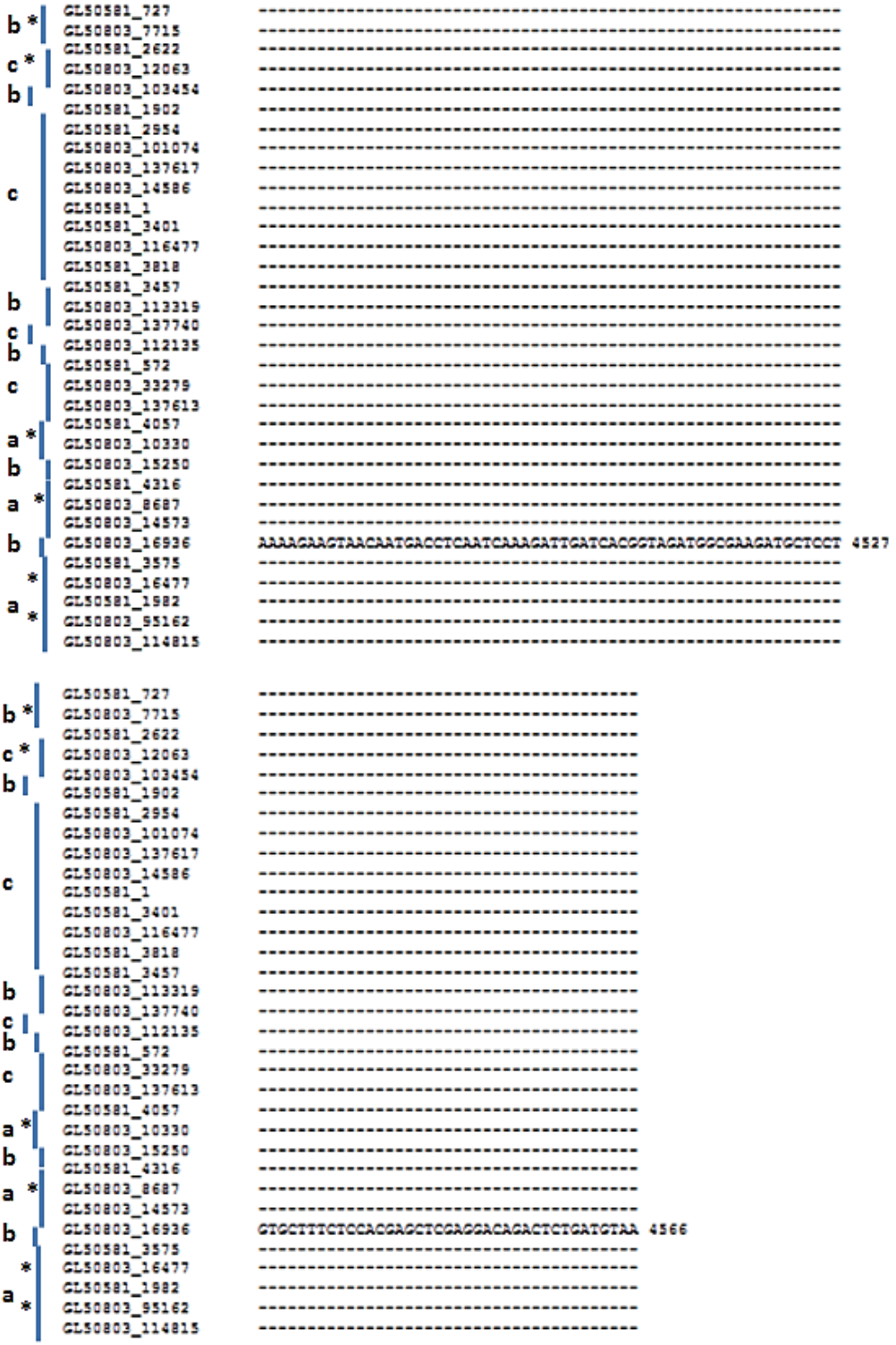

b *
c *
b
c
b
c
a *
b
a *
b
*
a *
b
c
b
c
a *
b
a *
b
*
a *


b * GL50581_727 GA---------- CCATCGCGCCGGGCACGTTCTGTG---- 2007
GL50803_7715 ---------- CGGTCGATTCGTCCACGCACTGTC---- 1998
c * GL50581_2622 ----------
GL50803_12063 ----------
b GL50803_103454 ---------- ATTTGGTGGCTAGTCTTTCAGAG---- 2156
GL50581_1902 GG---------- CAGAGGTAGTGGGCTACTGGTGG----- 2700
GL50581_2954 ----------
GL50803_101074 G---------- CGGCCTCGTCGGTTTCCTTTG------ 2189
GL50803_137617 G---------- CGGCCTCGTCGGTTTCCTCTG------ 2174
c GL50803_14586 G---------- CGGTCTTGTTGGGTTCCTCTG------ 2171
GL50581_1 G---------- CGGCCTCGTCGGGTTCCTCTG------ 1193
GL50581_3401 G---------- TGGTCTGGTTGGCTCTGACAG------ 1100
GL50803_116477 G---------- CGGCCTCGTGGGCTTCCTCT------- 2215
GL50581_3818 G---------- CGGCCTCGTGGGCTTCCTCT------- 1804
b GL50581_3457 G---------- GGGGTCCGTTGGACTTGTATTTG---- 2146
GL50803_113319 G---------- CGGGACCATCGGGCTCACATTTA---- 7531
c GL50803_137740 G---------- TGGCCTCGTCGGCTTCCTCTG------ 1325
b GL50803_112135 C---------- GGGCCTGGTCAGTTTTCTCCT------ 3404
GL50581_572 C---------- GG------------ ACCCCT------ 2008
c GL50803_33279 G---------- TGGTCTTGTGGGCTTCTTGTG------ 2216
GL50803_137613 G---------- GGGCCTCGTAGGCTTCCTTTG------ 1154
GL50581_4057 ----------
a * GL50803_10330 ----------
b GL50803_15250 CTCAG---------- AAAAATATGGAACCGACCTCTG------ 1831
GL50581_4316 CA---------- TTTGCGATTGCAAGTCTGGTT------- 1480
a * GL50803_8687 AG---------- TTTGCAGCTGTGACTCGACGC------- 1477
GL50803_14573 GG---------- TGTGTTTTTGCAAATACGACA------- 1477
b GL50803_16936 TGGGACATGCAGCGTAACAGGGGTGGTTGGAATTTGCAATTGTGATAGCGACTGGGTTAC 4184
GL50581_3575 CG---------- TCTGCTCTGGAAATGGAATATG------ 1685
* GL50803_16477 TG---------- TTTGCTCTGGAAATGGTGTGTG------ 1685
a * GL50581_1982 GG---------- TCTGCAACGGAACTGGTGCCTG------ 1697
GL50803_95162 GG---------- TCTGCTACGGAACTGGCGCCTG------ 1697
GL50803_114815 TG---------- TATGCAAGGGAAGAGGGATCTG------ 1748
GL50581_727 ---------- TCCAGACT---------- 2015
b * GL50803_7715 ---------- TCCAAAAT---------- 2006
GL50581_2622 ----------
c * GL50803_12063 ----------
b GL50803_103454 ---------- GAGAAGGG---------- 2164
GL50581_1902 ---------- AGCACGAG---------- 2708
GL50581_2954 ----------
GL50803_101074 ---------- CTGGTG---------- 2195
GL50803_137617 ---------- CTGGTG---------- 2180
c GL50803_14586 ---------- CTGGTG---------- 2177
GL50581_1 ---------- CTGGTG---------- 1199
GL50581_3401 ---------- AGATGCGG---------- 1108
GL50803_116477 ---------- GCTGGTG---------- 2222
GL50581_3818 ---------- GCTGGTG---------- 1811
GL50581_3457 ---------- GCTTAGTGGCATCCAAG---------- 2163
b GL50803_113319 ---------- GTTTGATGGCATCCAAG---------- 7548
c GL50803_137740 ---------- CTGGTG---------- 1331
b GL50803_112135 ---------- GTGGTG---------- 3410
GL50581_572 ---------- CGG---------- 2011
c GL50803_33279 ---------- TTGGTG---------- 2222
GL50803_137613 ---------- CTGGTG---------- 1160
GL50581_4057 ----------
a * GL50803_10330 ----------
b GL50803_15250 ---------- TTGACACC---------- 1839
GL50581_4316 ---------- ACAAG---------- 1485
a * GL50803_8687 ---------- ACCAG---------- 1482
GL50803_14573 ---------- ACCCG---------- 1482
b GL50803_16936 GATAGAGGGAATGTGTTACCCGGTTTCTTGCGTATTTGGAGGAGCCGTATGCAGCGGGTT 4244
GL50581_3575 ---------- TCGTG---------- 1690
* GL50803_16477 ---------- CCGCG---------- 1690
a * GL50581_1982 ---------- TCGCG---------- 1702
GL50803_95162 ---------- TCGCG---------- 1702
GL50803_114815 ---------- TCGCG---------- 1753
b * GL50581_727 ---------- CTCCGAGAG---------- TATGGAGCT 2033
GL50803_7715 ---------- GTCCGAGAG---------- TATGGGGCT 2024
c * GL50581_2622 ----------
GL50803_12063 ----------
b GL50803_103454 ---------- GTTCTGGTT---------- TTGGAGGCA 2182
GL50581_1902 ---------- GGCGGTAGA---------- CAGGA 2722
GL50581_2954 ----------
GL50803_101074 ---------- GTTTATGTG---------- CAGAG 2209
GL50803_137617 ---------- GTTTATGTG---------- CAGAG 2194
c GL50803_14586 ---------- GTTTATGTG---------- CAGAG 2191
GL50581_1 ---------- GTTCGTCTG---------- CCGCG 1213
GL50581_3401 ---------- GCACGCATATCCCCAGCATTGCAGATAGAG 1138
GL50803_116477 ---------- GTTCATATG---------- TAGAG 2236
GL50581_3818 ---------- GTTCATCTG---------- CAGGG 1825
b GL50581_3457 ---------- CGCACATCTC---------- AGCCT 2178
GL50803_113319 ---------- CATAAATCTC---------- AGGCC 7563
c GL50803_137740 ---------- GTTCTTGTG---------- TAGGG 1345
b GL50803_112135 ---------- GTTCCTATG---------- TAGAC 3424
GL50581_572 ----------
c GL50803_33279 ---------- GTTCCTCTG---------- CCGCG 2236
GL50803_137613 ---------- GTTCCTCTG---------- CCGCG 1174
GL50581_4057 ----------
a * GL50803_10330 ----------
b GL50803_15250 ---------- AACCCCG---------- AATG 1850
GL50581_4316 ---------- CCCCATA---------- CGTG 1496
a * GL50803_8687 ---------- TCCCGTA---------- CCTG 1493
GL50803_14573 ---------- TCGCCAA---------- CCTG 1493
b GL50803_16936 TGGAAAGTGCATCTTCTCTAGCAGCGGAGAGCACAGATG---------- CCTT 4287
* GL50581_3575 ---------- ACAATCTTTG---------- TTTG 1704
GL50803_16477 ---------- ACAACCGCTG---------- CTTG 1704
a * GL50581_1982 ---------- ATGGCAAGTG---------- TGTG 1716
GL50803_95162 ---------- ATGGTAGATG---------- TGTG 1716
GL50803_114815 ---------- ACGGAGCCTG---------- TTTC 1767


b * GL50581_727 TATAGGATCAGCTGATGACTTTTGA---------- 2058
b * GL50803_7715 CGTGGGCTCAGTTGACGACTTTTAA---------- 2049
c * GL50581_2622 ----------
GL50803_12063 ----------
b GL50803_103454 GCAGAGATATGCTGACGTCCAGGGACTAA---------- 2211
GL50581_1902 TAAGGGCGATAAGACTGCTCTTATGTACGCAGTATATGATAACAACCTAG---------- 2772
GL50581_2954 ----------
GL50803_101074 GCAAGGCGTAG---------- 2220
GL50803_137617 GCAAGGCGTAG---------- 2205
c GL50803_14586 GAAAAGCGTAG---------- 2202
GL50581_1 GGAAGGCGTGA---------- 1224
GL50581_3401 AGCCACAATGA---------- 1149
GL50803_116477 GGAAGGCATGA---------- 2247
GL50581_3818 GCAAGGCGTAA---------- 1836
GL50581_3457 GACGGACTTCGACCGCTAACTGCTTGA---------- 2205
b GL50803_113319 AGCGGGCTCCGGCCATTGACAGCTTGA---------- 7590
c GL50803_137740 GGAAGGCATAG---------- 1356
b GL50803_112135 GCAGCAGGTAA---------- 3435
GL50581_572 TCAACAAGTAG---------- 2022
c GL50803_33279 GGAAGGCGTAG---------- 2247
GL50803_137613 GGAAGGCGTAG---------- 1185
GL50581_4057 ----------
a * GL50803_10330 ----------
b GL50803_15250 TGTGGAGTATGAAACTGACTAG---------- 1872
GL50581_4316 TGCATAA---------- 1503
a * GL50803_8687 TGCAGTCAG----CAGCTCCCAGTGA---------- 1515
GL50803_14573 CTCGTAG---------- 1500
b GL50803_16936 TGCGACCAAGGTTACGAGCCTGGTGAAGATTTAGGAACGTGCATCACTCGTTCTACAAAG 4347
GL50581_3575 TGCGATCCTGGCTTCCGAGGTGGGTATTGCGAAAGGCCAGTGTAG---------- 1749
* GL50803_16477 TGCGATCCTGGCTTTCGAGGTGGATATTGTGAAAGAGCAGCGTAA---------- 1749
a * GL50581_1982 TGA---------- 1719
GL50803_95162 TGA---------- 1719
GL50803_114815 TATAGCGATTGA---------- 1779

b * GL50581_727 ----------
b * GL50803_7715 ----------
c * GL50581_2622 ----------
GL50803_12063 ----------
b GL50803_103454 ----------
GL50581_1902 ----------
GL50581_2954 ----------
GL50803_101074 ----------
GL50803_137617 ----------
c GL50803_14586 ----------
GL50581_1 ----------
GL50581_3401 ----------
GL50803_116477 ----------
GL50581_3818 ----------
GL50581_3457 ----------
b GL50803_113319 ----------
c GL50803_137740 ----------
b GL50803_112135 ----------
GL50581_572 ----------
c GL50803_33279 ----------
GL50803_137613 ----------
GL50581_4057 ----------
a * GL50803_10330 ----------
b GL50803_15250 ----------
GL50581_4316 ----------
a * GL50803_8687 ----------
GL50803_14573 ----------
b GL50803_16936 GCAAAGCAAGGAGCGGCCGAGATATCGATAATTGTGGTGGTTGTGGTTGCTCTAGTGATT 4407
GL50581_3575 ----------
* GL50803_16477 ----------
a * GL50581_1982 ----------
GL50803_95162 ----------
GL50803_114815 ----------

b * GL50581_727 ----------
b * GL50803_7715 ----------
c * GL50581_2622 ----------
GL50803_12063 ----------
b GL50803_103454 ----------
GL50581_1902 ----------
GL50581_2954 ----------
GL50803_101074 ----------
GL50803_137617 ----------
c GL50803_14586 ----------
GL50581_1 ----------
GL50581_3401 ----------
GL50803_116477 ----------
GL50581_3818 ----------
GL50581_3457 ----------
b GL50803_113319 ----------
c GL50803_137740 ----------
b GL50803_112135 ----------
GL50581_572 ----------
c GL50803_33279 ----------
GL50803_137613 ----------
GL50581_4057 ----------
a * GL50803_10330 ----------
b GL50803_15250 ----------
GL50581_4316 ----------
a * GL50803_8687 ----------
GL50803_14573 ----------
b GL50803_16936 TTGGCAGTAGCCATATTCTTGATCTGGAAGTTCGTCATCAAACCTCGCTCCAAGTCACGC 4467
GL50581_3575 ----------
* GL50803_16477 ----------
a * GL50581_1982 ----------
GL50803_95162 ----------
GL50803_114815 ----------


b *
c *
b
c
b
c
a *
b
a *
b
*
a *
GL50581_727
GL50803_7715
GL50581_2622
GL50803_12063
GL50803_103454
GL50581_1902
GL50581_2954
GL50803_101074
GL50803_137617
GL50803_14586
GL50581_1
GL50581_3401
GL50803_116477
GL50581_3818
GL50581_3457
GL50803_113319
GL50803_137740
GL50803_112135
GL50581_572
GL50803_33279
GL50803_137613
GL50581_4057
GL50803_10330
GL50803_15250
GL50581_4316
GL50803_8687
GL50803_14573
GL50803_16936
GL50581_3575
GL50803_16477
GL50581_1982
GL50803_95162
GL50803_114815
AAAAGAAGTAACAATGACCTCAATCAAAGATTGATCACGGTAGATGGCGAAGATGCTCCT 4527
GTGCTTTCTCCACGAGCTCGAGGACAGACTCTGATGTAA 4566

Appendix XII:

CLustalW sequence alignments of the conserved secreted hypothetical protein coding genes. All the GL50581_ accession numbers represent GS isolate proteins and all the GL50803_ represent WB isolate proteins.

```
CLUSTAL 2.1 multiple sequence alignment


GL50581_2767        ----------------------ATGAC-CGACAGGCCTAGC-------CAAGAGGAT--- 27
GL50803_5258        ----------------------ATGAC-TGATAGACCGAGT-------CAGGAGGAC--- 27
GL50581_352         ----------------------ATGTCGCGCTGGACCTTCCTGCGGTGCTTGAGAACGAG 38
GL50803_16522       ----------------------ATGTCACGCTGGACCTTCCTGCGGTGCTTGAGGACGAG 38
GL50581_4180        ATGCTTCTTGCGCTTCTTGGGCTCCTGGCGCTATTCCGCTCTTCGTGCCCTGAGGGAGAA 60
GL50803_113038      ATGCTTCCTGTACTTCTTGGGCTCCTAGCGGGGTTCCTCTCCGCGTGCCCCGAGGGCGAA 60
GL50581_4133        ----------------------ATGGTTGATCCCAAGCTTATCGAGCTCAACGAGGGTG-- 37
GL50803_5810        ----------------------ATGGTGGATCCTAAGCTTATCGAGCTCAACGAAGGCG-- 37
                                           : :          *          .  **...

GL50581_2767        CTCGAT----------------------GAGCTGCAGCG----CGAG----------AT 50
GL50803_5258        CTTGAC----------------------GAGCTGCAGCG----CGAG----------AT 50
GL50581_352         CTCGGG----------------------GACAGGCATCCGGTTCGGGGACTCAACTAAT 75
GL50803_16522       CTCGGG----------------------GATAGGCATCTCGTTCGGGGGCTCAACTAAT 75
GL50581_4180        ATAGAGGTCGACGGAGAGTGCTATCCAGGGGCGTGCGTATTCGACGGGACGGTCTGCGCG 120
GL50803_113038      ATCGAGGTCGACGGGGAGTGCTACCCGGAGGCGTGCGTCTTCGACGGAACAGTCTGCGCG 120
GL50581_4133        CAAAGC----------------------ACATAGCTACTATTGCATAC---------CT 65
GL50803_5810        CAAAGC----------------------ACATAGCTACTATCGCGTAT---------CT 65
                    .: ..                      .    **  .    *. .          .

GL50581_2767        GGATAAG---------AAGGAGGCAGAGGCCAAG------------------------- 75
GL50803_5258        GGACAAG---------AAGGAGGCAGAGGCCAAG------------------------- 75
GL50581_352         AATCACT---------TAGAAATCAAATTCAAATGG------------------------ 102
GL50803_16522       AATCACC---------TAGAAATCAAATTCAAATGG------------------------ 102
GL50581_4180        GGTCACGGAACTTGCATGGTGAACTATTGCTATTGCGATGATGGCTACAGACTGTCGAAC 180
GL50803_113038      GGCCACGGGGTCTGCATGGTGAACTATTGCTATTGTGACGACGGCTACAGACTGTCGAAC 180
GL50581_4133        GAATGAA---------AAGGGCTATCCAGACAGCG------------------------ 91
GL50803_5810        GAACGAG---------AAGGGCTATCCAGATAGCG------------------------ 91
                    ..  ..          :.* .  .:    . *

GL50581_2767        ---------------------------AAGGCCGGAA---------------------- 85
GL50803_5258        ---------------------------AAGGCCGGGG---------------------- 85
GL50581_352         --------------------TTTATAGTATTATTGGG---------------------- 119
GL50803_16522       --------------------TTTACAATATTATTGGG---------------------- 119
GL50581_4180        GCTGGATGCACACCATCGGACTGCACAGACAGCATGGGCCTTGTTTGTGGCGACCATGGC 240
GL50803_113038      GCCGGGTGCACGCCTGCGGAGTGTACGGATAGCATGGGGCTTGTCTGCGGCGATCACGGT 240
GL50581_4133        ----------------------------AGCCCACG---------------------- 99
GL50803_5810        ----------------------------AACCTACG---------------------- 99
                                               .  . .

GL50581_2767        ------------------------------------------TACCTTTTAACAGGGCC 102
GL50803_5258        ------------------------------------------TGCCCTTCAATAGAGCG 102
GL50581_352         ------------------------------------------TACGATGTCTCAGGGAC 136
GL50803_16522       ------------------------------------------CAGGCT-TCCCAGAATA 135
GL50581_4180        ACCTGCAAGCACGGCTCTGACGGCTTTGCATGCGTCTGCGAGGAGGGCTACACAAGCGTC 300
GL50803_113038      GTCTGCAAGAGGGGCCCGGATGGCTTTTCCTGCGCTTGCGACGAGGGCTACACAAGCGTC 300
GL50581_4133        ------------------------------------------ACTGTTCTGTATTC 113
GL50803_5810        ------------------------------------------ACTGTCCTGTACAC 113
                                                               *     .
```

```
GL50581_2767        GCAGCGCTTC---------------GCAGGATTGCCCAG----TC--------------- 128
GL50803_5258        GCTGCGCTGC---------------GCAGGATTGCCCAG----TC--------------- 128
GL50581_352         TAC-AGATGA---------------GCGAAAATTCGGGAGAAATCG-------------- 166
GL50803_16522       GTCGAGATGG---------------ACGGGAATCCGGGGGAGATCG-------------- 166
GL50581_4180        GCCCCGACGTGCGCCCCCAATGAGTGCGTGCATGACGGAAAGATCTGCGGCGGATTTGGG 360
GL50803_113038      TCCCCGACGTGTGCTCCCGACGAGTGCGTGCACGACGGAAAGATATGCGGTGGAGCTGGA 360
GL50581_4133        GAAGACACTTG--------------ACAGGCTCGTTACAGCTGTCA-------------- 145
GL50803_5810        GAAGACACTTG--------------ACAGACTCGTCACGGCTGTCA-------------- 145
                        . .                  .*. ..:      .    *.

GL50581_2767        ------------------------------------------------------CAAGGGAAGA 138
GL50803_5258        ------------------------------------------------------CAAGGGAAGG 138
GL50581_352         ---------------------------------------AAATTATCGACGCACGCGAGGA 188
GL50803_16522       ---------------------------------------AGGTCGTTAACGCGCGAGAGGG 188
GL50581_4180        ACGTGCGTTGTCGAGGACGGTGTTCCTGCCTGTGTCTGTGATCCCTTTTTCGCGGGAACC 420
GL50803_113038      GCGTGTGTTGCTGAGGATGGCACGCCCGCTTGCGTCTGTGATCCCTTCTACGCTGGGACC 420
GL50581_4133        ------------------------------------------------ATAAGAACAGC 156
GL50803_5810        ------------------------------------------------ACAAGAACAGC 156
                                                                       ..  . .

GL50581_2767        TTCTACG-----------------------------------ACTCTGCAGAG-------- 156
GL50803_5258        TTCTATG-----------------------------------ACTCTGCAGAG-------- 156
GL50581_352         AGCTACGG----------------------------CTGAGCTCAAGGGAAGCACGC-- 217
GL50803_16522       GGCAACGG----------------------------CAGAGCTCAGCGGGAGTGTGC-- 217
GL50581_4180        TACTGCGAGGAGTGCGAGTCTGGTGCAGTTCTCTTTGGCGACCACTGCGTCGACCCGCGC 480
GL50803_113038      TTCTGCGAGGAGTGCGCACCCGATGCGGTCCTTTTCGAAGACCACTGCGTCGACCAGCGT 480
GL50581_4133        GCCAAGG-----------------------------------TCAAGGCGGTGACCAATAA 182
GL50803_5810        GCCAAGG-----------------------------------TCAAGGCCGTGACTAACAA 182
                      *:. *                                     *:. .    .

GL50581_2767        TTC---------ATTCTAAAG-------CAGGGTAGGGGAG---------AAAC------ 185
GL50803_5258        TTC---------ATCTTGAAG-------CAAGGTAGAGGAG---------AGAC------ 185
GL50581_352         TTCGCTTGGCGTGTCCGGGAGG---CATCGGGCATGTGCAG---------TTTCGGCTCT 265
GL50803_16522       TTCGCCTAGCATGTCCGGGAGG---GATCGGGCATGTGCAG---------TTTCGGCTCC 265
GL50581_4180        TGCCTCAGCGAGAANCANGAGGGCAGCACCGTCGAGTGCGGGGGCTTTGGAGTCTGCGTT 540
GL50803_113038      TGCCTCAGCGAGAACCACGAGGGCGACACCGTCGAGTGCGGGGGCTTCGGGGTCTGCTTC 540
GL50581_4133        CCCCATGTTTACCTTCAAGATTG---GCCAGCCGTTTGGTG--------ACAACGCC--- 228
GL50803_5810        CCCCATGTTCACCTTCAAGATTG---GCCAGCCCTTCGGTG--------ACAACGCC--- 228
                      *          :    .*       * .   :  *  *          :*

GL50581_2767        ----------------------CCAGCCT-------------AGTGCTTACTTCAATAAA 210
GL50803_5258        ----------------------CCAGCCT-------------AGTGCTTACTTCAACAAG 210
GL50581_352         CTCACCCAACGCA---------TCTGACTGT-------AGTCATCGCTATTTCCAATTGC 309
GL50803_16522       GCAGCGCCACGCG---------CCTGACCGT-------GGCCATTGCAATCTCTAACTGC 309
GL50581_4180        CGCTATCGCATTGGGATCGTTTCCTGCATCTGCTATCCGAGCACTGTTCTCGACAACTAC 600
GL50803_113038      TTCTACCCCATCGGGATCATTTCCTGCATCTGCTACCCGAGTACTGTCCTCGACAACTAC 600
GL50581_4133        ----------------------CTTGTCATG--------GCCAAAGCACGTATCATACCT 258
GL50803_5810        ----------------------CTTGTCATG--------GCCAAGGCACGCATTATACCC 258
                                         :* .             *  *        *:

GL50581_2767        GAG-----------------------ATAACAAAGGAGAG---CGCAGG-AAAGCCGAA 242
GL50803_5258        GAC-----------------------ATAACAAAGGAGAC---TGCAGG-AAAGCCGAA 242
GL50581_352         GAGGGGCTTAATGTGACTGTTGGTGAGATAACAAGGGAACG---GGAGGACTTCGATGTG 366
GL50803_16522       GAGGGCCTCGACGTGACGATCGGCGAAATTACCAAGGAACG---GGGAGACTTTGACGTG 366
GL50581_4180        CAGTGCGTTCCGTTCACGTGCGCATCTCAGCCGGAGGACAGGATTGTAGCCTGCAGCAAC 660
GL50803_113038      AAGTGTGTTCCATTCACGTGTGCGTCCCAGCCGGAGGACAGGATCGTGGCCTGCAGCGAT 660
GL50581_4133        CAGGGCCAGG-AGCCCCAGGCCTTCGCGGACGTGTGGTGTG---ATCAGATATCTAAGTA 314
GL50803_5810        CAAGGTCAGG-AGCCCCAGGCCTTTGCAGATGTCTGGTGCG---ATCAGATATCCAAGTA 314
                     *                                 **:         .*  :     .:

GL50581_2767        C-CAGCCC----------------------------GCTGCTG-----GAGGTAGTCGTT 268
GL50803_5258        C-CCACCT----------------------------GCCGCGG-----GGGGAAGCCGTT 268
GL50581_352         CGTCAAGG----------------------------CGCGCAG-----GAGGCCATCTTT 393
GL50803_16522       CGCCAGGG----------------------------GCCACAA-----GAGGCAATCTTC 393
GL50581_4180        CACGGAATGTGCGAGCCAGATCGTGTCTGTTACTGCGACGATGGCTTTGAGGGCGACGTC 720
GL50803_113038      CGCGGGATCTGCGAGCCAGACAGAACATGCCGTTGTGACGAGGGCTTTGAGGGCGACGTC 720
GL50581_4133        CTTCAAGG----------------------------GTGGTGCCTCCGACCCAGACATG 345
GL50803_5810        CTTCAAGG----------------------------GTGGTGCGTCTGACCCGGACATG 345
                    *   .                                  .  .    *.    . * *
```

```
GL50581_2767        TTGCT-GCCAAAG-------CGCCGCAGGACGATGATG------------ACGATGAGT- 307
GL50803_5258        TTGCT-GCCAAGA-------CATCCCAAGACGAAGATGACAACGACGATAACGATGAGT- 319
GL50581_352         GATCTAGACGTAC-------CAGCCCACAAGGACGTCCGCATTCAGTGGATAGACTTTT- 445
GL50803_16522       GAGCTCGACGTGC-------CGGCCAACCAGGATGTCCGTGTCCAGTGGATAGACTATT- 445
GL50581_4180        TGTCAGTATGAAATCATTGACTGCCCTGAAGGAACGCGCTACTTCGAGGGCCAGTGTGTG 780
GL50803_113038      TGCCAGTACAAGATTATCGACTGCCCTGCAGGTACGCGCTACTTTGAGGGTCAGTGTGTG 780
GL50581_4133        GTGCTGCTTGTGG-------CTGAGATAGAAGATGTCG---CCTGGCGCATCAAAAATTA 395
GL50803_5810        GTGCTGCTTGTGG-------CCGAGATCGAGGATGTCG---CCTGGCGCATCAAGAATTG 395
                       *:    .:.       *  . .:  * *:                  ...  : *

GL50581_2767        AA---------------------------------------------------------- 309
GL50803_5258        AA---------------------------------------------------------- 321
GL50581_352         ATAGATAG---------------------------------------------------- 453
GL50803_16522       ACAGATAG---------------------------------------------------- 453
GL50581_4180        ACCGATGGCTGCATCTCTTCCGACGGACATGTGTGTGGGGGCCACGGGACTTGCACCAAC 840
GL50803_113038      ACCGATGGGTGCATCTCCCCCGACGGACACGTCTGCGGTGACCACGGGGTTTGCGCAAAC 840
GL50581_4133        A---------------------------------------------------------- 396
GL50803_5810        A---------------------------------------------------------- 396
                    *

GL50581_2767        ------------------------------------------------------------
GL50803_5258        ------------------------------------------------------------
GL50581_352         ------------------------------------------------------------
GL50803_16522       ------------------------------------------------------------
GL50581_4180        GACGCTTGCATTTGCAACGATGGCTACGAGATCATTGATAAGAACGCTTGCGTGTCGTCC 900
GL50803_113038      AACGCATGCACCTGCAACGAGGGGTACGAGCTCATCGATAAGAACGTCTGCATGCCATCC 900
GL50581_4133        ------------------------------------------------------------
GL50803_5810        ------------------------------------------------------------

GL50581_2767        ------------------------------------------------------------
GL50803_5258        ------------------------------------------------------------
GL50581_352         ------------------------------------------------------------
GL50803_16522       ------------------------------------------------------------
GL50581_4180        AATTGTGTTGTCGATGGAGAGGTCTGCCCGTATGGGACGTGCGAGCGCGACGGAATGTCC 960
GL50803_113038      AACTGCATCGTCGACGGAGAGGTGTGCCCGTACGGGACATGTGAGCGTGAAGGTATGTCC 960
GL50581_4133        ------------------------------------------------------------
GL50803_5810        ------------------------------------------------------------

GL50581_2767        ------------------------------------------------------------
GL50803_5258        ------------------------------------------------------------
GL50581_352         ------------------------------------------------------------
GL50803_16522       ------------------------------------------------------------
GL50581_4180        TGGGCATGCGCATGCCGTGCAGAGTACGAGCTCTACAACGGCAAGTGCTACCCTCACTCC 1020
GL50803_113038      TGGACGTGTGTGTGCCGCGCAGAGTATGTGCCCTACAATGGCAAGTGCTACCCCAATTCC 1020
GL50581_4133        ------------------------------------------------------------
GL50803_5810        ------------------------------------------------------------

GL50581_2767        ------------------------------------------------------------
GL50803_5258        ------------------------------------------------------------
GL50581_352         ------------------------------------------------------------
GL50803_16522       ------------------------------------------------------------
GL50581_4180        TGTGTTAGCGATGTCCTTCCAAATGGAGACCCCGAGCTCTGTTCAGGAGAGGGCCAGTGC 1080
GL50803_113038      TGTGTCACTGACATCCTTCCAAATGGAGATCCCAAGCTCTGCTCTGGACTGGGCCAATGC 1080
GL50581_4133        ------------------------------------------------------------
GL50803_5810        ------------------------------------------------------------

GL50581_2767        ------------------------------------------------------------
GL50803_5258        ------------------------------------------------------------
GL50581_352         ------------------------------------------------------------
GL50803_16522       ------------------------------------------------------------
GL50581_4180        GATATGGAAGCTCGTCGATGCACTTGCCTTTCTAACTATGCTGGAACGTATTGTCAGAAA 1140
GL50803_113038      GACATGGAGGCTCGCCGCTGCATCTGCCTCTCGATAACTACGGGAACATATTGCCAGAAG 1140
GL50581_4133        ------------------------------------------------------------
GL50803_5810        ------------------------------------------------------------
```

```
GL50581_2767        ------------------------------------------------------------
GL50803_5258        ------------------------------------------------------------
GL50581_352         ------------------------------------------------------------
GL50803_16522       ------------------------------------------------------------
GL50581_4180        TGTAGCGAAGAGGCTCTTCTACTCAATGATAGATGCGTCGCGCTTGCCTGTGTAAGTTAT 1200
GL50803_113038      TGCAGCGATGAGGCACTCCAGATCGGTGAAAGGTGCGTCTCGTTATCCTGCATAAGTTAT 1200
GL50581_4133        ------------------------------------------------------------
GL50803_5810        ------------------------------------------------------------

GL50581_2767        ------------------------------------------------------------
GL50803_5258        ------------------------------------------------------------
GL50581_352         ------------------------------------------------------------
GL50803_16522       ------------------------------------------------------------
GL50581_4180        GATCTCGATGGTAACCCTCTTGAGTGTGACGGACGCGGGGAATGTATAGCGTCCATGTTG 1260
GL50803_113038      GATGCCAATGGAAACCCTGTTGAGTGCAGCGGGCATGGAGGCTGCGTTGTTTCCTTTTCA 1260
GL50581_4133        ------------------------------------------------------------
GL50803_5810        ------------------------------------------------------------

GL50581_2767        ------------------------------------------------------------
GL50803_5258        ------------------------------------------------------------
GL50581_352         ------------------------------------------------------------
GL50803_16522       ------------------------------------------------------------
GL50581_4180        GCTGACTACCCTCAGGAGTATACATTTATCTGCAGTTGTTATGGTGGTTACTTTGCTCTC 1320
GL50803_113038      GCAGAACACCCTGAAGATTATAACTTTCTTTGCGACTGCGAAGATGGGTACCTTGCGTTA 1320
GL50581_4133        ------------------------------------------------------------
GL50803_5810        ------------------------------------------------------------

GL50581_2767        ------------------------------------------------------------
GL50803_5258        ------------------------------------------------------------
GL50581_352         ------------------------------------------------------------
GL50803_16522       ------------------------------------------------------------
GL50581_4180        GATACAAGCACATGCAGCCATGTCAATTGTTTGAATAACGATCTTTCCCTCAAGCCTATC 1380
GL50803_113038      GAAACAAGTACATGTACCCACTTCAAGTGCCTAAATCGCGATCTTTCCTTCAAACCCATC 1380
GL50581_4133        ------------------------------------------------------------
GL50803_5810        ------------------------------------------------------------

GL50581_2767        ------------------------------------------------------------
GL50803_5258        ------------------------------------------------------------
GL50581_352         ------------------------------------------------------------
GL50803_16522       ------------------------------------------------------------
GL50581_4180        TGCAGCGGAGTTGGATTCTGTCTCGGGCCTGAGGGCTGTGCGTGTCCAGAGGGCTATAGT 1440
GL50803_113038      TGTAGTGGCGTTGGCTTCTGTCTCGGACCCGAGGGTTGTGCATGTCCCGAGGGCTACAGT 1440
GL50581_4133        ------------------------------------------------------------
GL50803_5810        ------------------------------------------------------------

GL50581_2767        ------------------------------------------------------------
GL50803_5258        ------------------------------------------------------------
GL50581_352         ------------------------------------------------------------
GL50803_16522       ------------------------------------------------------------
GL50581_4180        GGACAAACGTGCGAACGCCGGGAATGTCCTGAAGGTGAGACAGCGGTCATGGCCTGGGGG 1500
GL50803_113038      GGTACGCGCTGTGAAAACCGGGAATGCCCTGAAGGCGAGACGGCAGTTCTGGCCTGGGGC 1500
GL50581_4133        ------------------------------------------------------------
GL50803_5810        ------------------------------------------------------------

GL50581_2767        ------------------------------------------------------------
GL50803_5258        ------------------------------------------------------------
GL50581_352         ------------------------------------------------------------
GL50803_16522       ------------------------------------------------------------
GL50581_4180        TGCGTTCCTGATGCCTGCGTCACGGAGTATTCTGATGGCGTTAAACGGGTATGCGCAGGC 1560
GL50803_113038      TGCGTCCCGGATGCGTGCGTTACTGAATATGCAGACGGTGTCCGGCGCGTTTGTGCCGGG 1560
GL50581_4133        ------------------------------------------------------------
GL50803_5810        ------------------------------------------------------------
```

```
GL50581_2767        ------------------------------------------------------------
GL50803_5258        ------------------------------------------------------------
GL50581_352         ------------------------------------------------------------
GL50803_16522       ------------------------------------------------------------
GL50581_4180        TTCGGGCACTGTATAACAGATAATGATGTTGCTCAGTGCCACTGCAACGGAGCCGCCAAG 1620
GL50803_113038      TTCGGACACTGCATAACAGAGGACGGAGTTGCTCGGTGCCGCTGCAACGGAGAGGCAGAA 1620
GL50581_4133        ------------------------------------------------------------
GL50803_5810        ------------------------------------------------------------

GL50581_2767        ------------------------------------------------------------
GL50803_5258        ------------------------------------------------------------
GL50581_352         ------------------------------------------------------------
GL50803_16522       ------------------------------------------------------------
GL50581_4180        CTCATCGACGGCGTGTGCACAGCACCGGCATGCATCGGAGCTGACGGAACTGTTTGCAAA 1680
GL50803_113038      CTCATTGACGGTGTTTGCACAACCCCGGCGTGCATTGGGGCCGATGGAACTGTATGCAAG 1680
GL50581_4133        ------------------------------------------------------------
GL50803_5810        ------------------------------------------------------------

GL50581_2767        ---------------------------------------------
GL50803_5258        ---------------------------------------------
GL50581_352         ---------------------------------------------
GL50803_16522       ---------------------------------------------
GL50581_4180        GGTAGAGGTACTTGTCGCGATGGGGCTTGTTTCTATGGTAACTAG 1725
GL50803_113038      GGAAGAGGGATCTGCCGTGACGGAGCTTGTTTCTATGGCAATTGA 1725
GL50581_4133        ---------------------------------------------
GL50803_5810        ---------------------------------------------
```

References

ABEL, E. S., DAVIDS, B. J., ROBLES, L. D., LOFLIN, C. E., GILLIN, F. D. & CHAKRABARTI, R. 2001. Possible roles of protein kinase A in cell motility and excystation of the early diverging eukaryote Giardia lamblia. *J Biol Chem,* 276**,** 10320-9.

ADAM, R. D. 1991. The biology of Giardia spp. *Microbiol Rev,* 55**,** 706-32.

ADAM, R. D. 2001. Biology of Giardia lamblia. *Clin Microbiol Rev,* 14**,** 447-75.

ADAM, R. D., DAHLSTROM, E. W., MARTENS, C. A., BRUNO, D. P., BARBIAN, K. D., RICKLEFS, S. M., HERNANDEZ, M. M., NARLA, N. P., PATEL, R. B., PORCELLA, S. F. & NASH, T. E. 2013. Genome sequencing of Giardia lamblia genotypes A2 and B isolates (DH and GS) and comparative analysis with the genomes of genotypes A1 and E (WB and Pig). *Genome Biol Evol,* 5**,** 2498-511.

ADAM, R. D., NIGAM, A., SESHADRI, V., MARTENS, C. A., FARNETH, G. A., MORRISON, H. G., NASH, T. E., PORCELLA, S. F. & PATEL, R. 2010. The Giardia lamblia vsp gene repertoire: characteristics, genomic organization, and evolution. *BMC Genomics,* 11**,** 424.

ADAM, R. D., YANG, Y. M. & NASH, T. E. 1992. The cysteine-rich protein gene family of Giardia lamblia: loss of the CRP170 gene in an antigenic variant. *Mol Cell Biol,* 12**,** 1194-201.

AEBERSOLD, R. & GOODLETT, D. R. 2001. Mass spectrometry in proteomics. *Chemical reviews,* 101**,** 269-95.

AEBERSOLD, R. & MANN, M. 2003. Mass spectrometry-based proteomics. *Nature,* 422**,** 198-207.

AGARWAL, S. & MAYER, L. 2009. Pathogenesis and treatment of gastrointestinal disease in antibody deficiency syndromes. *J Allergy Clin Immunol,* 124**,** 658-64.

AGRE, P., KING, L. S., YASUI, M., GUGGINO, W. B., OTTERSEN, O. P., FUJIYOSHI, Y., ENGEL, A. & NIELSEN, S. 2002. Aquaporin water channels--from atomic structure to clinical medicine. *J Physiol,* 542**,** 3-16.

AGUAYO-ORTIZ, R., MENDEZ-LUCIO, O., ROMO-MANCILLAS, A., CASTILLO, R., YEPEZ-MULIA, L., MEDINA-FRANCO, J. L. & HERNANDEZ-CAMPOS, A. 2013. Molecular basis for benzimidazole resistance from a novel beta-tubulin binding site model. *J Mol Graph Model,* 45**,** 26-37.

AL-NAIMI, S., WINPENNY, J., HUNTER, P. R. & TYLER, K. M. Unpublished data. A soluble mediator from Giardia Lamblia disrupts human gastrointestinal epithelial cell function.

ALDEEN, W. E., CARROLL, K., ROBISON, A., MORRISON, M. & HALE, D. 1998. Comparison of nine commercially available enzyme-linked immunosorbent assays for detection of Giardia lamblia in fecal specimens. *J Clin Microbiol,* 36**,** 1338-40.

ALMEIDA, A., POZIO, E. & CACCIO, S. M. 2010. Genotyping of Giardia duodenalis cysts by new real-time PCR assays for detection of mixed infections in human samples. *Appl Environ Microbiol,* 76**,** 1895-901.

ALUM, A., SBAI, B., ASAAD, H., RUBINO, J. R. & KHALID IJAZ, M. 2012. ECC-RT-PCR: a new method to determine the viability and infectivity of Giardia cysts. *Int J Infect Dis,* 16**,** e350-3.

ALVARADO, M. E. & WASSERMAN, M. 2010. Analysis of phosphorylated proteins and inhibition of kinase activity during Giardia intestinalis excystation. *Parasitol Int,* 59**,** 54-61.

ANDERSON, J. M. & VAN ITALLIE, C. M. 1995. Tight junctions and the molecular basis for regulation of paracellular permeability. *Am J Physiol,* 269**,** G467-75.

ANDREWS, B. J. & BJORVATN, B. 1994. Chemotherapy of Entamoeba histolytica: studies in vitro with bacitracin and its zinc salt. *Trans R Soc Trop Med Hyg,* 88**,** 98-100.

ANDREWS, B. J., MYLVAGANAM, H. & YULE, A. 1994. Sensitivity of Trichomonas vaginalis, Tritrichomonas foetus and Giardia intestinalis to bacitracin and its zinc salt in vitro. *Trans R Soc Trop Med Hyg,* 88**,** 704-6.

ANDREWS, B. J., PANITESCU, D., JIPA, G. H., VASILE-BUGARIN, A. C., VASILIU, R. P. & RONNEVIG, J. R. 1995. Chemotherapy for giardiasis: randomized clinical trial of bacitracin, bacitracin zinc, and a combination of bacitracin zinc with neomycin. *Am J Trop Med Hyg,* 52**,** 318-21.

ANKARKLEV, J., JERLSTROM-HULTQVIST, J., RINGQVIST, E., TROELL, K. & SVARD, S. G. 2010. Behind the smile: cell biology and disease mechanisms of Giardia species. *Nat Rev Microbiol,* 8**,** 413-22.

ARBO, A., PAVIA-RUZ, N. & SANTOS, J. I. 2006. Opsonic requirements for the respiratory burst of neutrophils against Giardia lamblia trophozoites. *Arch Med Res,* 37**,** 465-73.

BACKHED, F., LEY, R. E., SONNENBURG, J. L., PETERSON, D. A. & GORDON, J. I. 2005. Host-bacterial mutualism in the human intestine. *Science,* 307**,** 1915-20.

BALCIOGLU, I. C., KURT, O., LIMONCU, M. E., DINC, G., GUMUS, M., KILIMCIOGLU, A. A., KAYRAN, E. & OZBILGIN, A. 2007. Rural life, lower socioeconomic status and parasitic infections. *Parasitol Int,* 56**,** 129-33.

BALDI, F., BIANCO, M. A., NARDONE, G., PILOTTO, A. & ZAMPARO, E. 2009. Focus on acute diarrhoeal disease. *World J Gastroenterol,* 15**,** 3341-8.

BALLWEBER, L. R., XIAO, L., BOWMAN, D. D., KAHN, G. & CAMA, V. A. 2010. Giardiasis in dogs and cats: update on epidemiology and public health significance. *Trends Parasitol,* 26**,** 180-9.

BARAKAT, A. & HALAWA, E. F. 2013. Household costs of seeking outpatient care in Egyptian children with diarrhea: a cross-sectional study. *Pan Afr Med J,* 14**,** 42.

BARTLETT, A. V., ENGLENDER, S. J., JARVIS, B. A., LUDWIG, L., CARLSON, J. F. & TOPPING, J. P. 1991. Controlled trial of Giardia lamblia: control strategies in day care centers. *Am J Public Health,* 81**,** 1001-6.

BAYLESS, T. M. & CHRISTOPHER, N. L. 1969. Disaccharidase deficiency. *Am J Clin Nutr,* 22**,** 181-90.

BAZAN-TEJEDA, M. L., ARGUELLO-GARCIA, R., BERMUDEZ-CRUZ, R. M., ROBLES-FLORES, M. & ORTEGA-PIERRES, G. 2007. Protein kinase C isoforms from Giardia duodenalis: identification and functional characterization of a beta-like molecule during encystment. *Arch Microbiol,* 187**,** 55-66.

BEAUMONT, D. M. & JAMES, O. F. 1986. Unsuspected giardiasis as a cause of malnutrition and diarrhoea in the elderly. *Br Med J (Clin Res Ed),* 293**,** 554-5.

BEDNARSKA, M., BAJER, A., SINSKI, E., GIROUARD, A. S., TAMANG, L. & GRACZYK, T. K. 2007. Fluorescent in situ hybridization as a tool to retrospectively identify Cryptosporidium parvum and Giardia lamblia in samples from terrestrial mammalian wildlife. *Parasitol Res,* 100**,** 455-60.

BEDORE, J., LEASK, A. & SEGUIN, C. A. 2014. Targeting the extracellular matrix: Matricellular proteins regulate cell-extracellular matrix communication within distinct niches of the intervertebral disc. *Matrix Biol*.

BEHR, M. A., KOKOSKIN, E., GYORKOS, T. W., CEDILOTTE, L., FAUBERT, G. M. & MACLEAN, J. D. 1997. Laboratory diagnosis for Giardia lamblia infection: A comparison of microscopy, coprodiagnosis and serology. *Can J Infect Dis,* 8**,** 33-8.

BELL, A., MONAGHAN, P. & PAGE, A. P. 2006. Peptidyl-prolyl cis-trans isomerases (immunophilins) and their roles in parasite biochemistry, host-parasite interaction and antiparasitic drug action. *Int J Parasitol,* 36**,** 261-76.

BENCHIMOL, M. 2002. A new set of vesicles in Giardia lamblia. *Exp Parasitol,* 102**,** 30-7.

BENERE, E., T, V. A. N. A., COS, P. & MAES, L. 2011. Variation in growth and drug susceptibility among Giardia duodenalis assemblages A, B and E in axenic in vitro culture and in the gerbil model. *Parasitology,* 138**,** 1354-61.

BENERE, E., VAN ASSCHE, T., VAN GINNEKEN, C., PEULEN, O., COS, P. & MAES, L. 2012. Intestinal growth and pathology of Giardia duodenalis assemblage subtype A(I), A(II), B and E in the gerbil model. *Parasitology,* 139**,** 424-33.

BERKMAN, D. S., LESCANO, A. G., GILMAN, R. H., LOPEZ, S. L & BLACK, M. M. 2002. Effects of stunting, diarrhoeal disease, and parasitic infection during infancy on cognition in late childhood: a follow-up study. *Lancet,* 359**,** 564-71.

BERNAL, R. M., TOVAR, R., SANTOS, J. I. & MUNOZ, M. L. 1998. Possible role of calmodulin in excystation of Giardia lamblia. *Parasitol Res,* 84**,** 687-93.

BERNANDER, R., PALM, J. E. & SVARD, S. G. 2001. Genome ploidy in different stages of the Giardia lamblia life cycle. *Cell Microbiol,* 3**,** 55-62.

BERTRAND, I., GANTZER, C., CHESNOT, T. & SCHWARTZBROD, J. 2004. Improved specificity for Giardia lamblia cyst quantification in wastewater by development of a real-time PCR method. *J Microbiol Methods,* 57**,** 41-53.

BETANCOURT, W. Q. & ROSE, J. B. 2004. Drinking water treatment processes for removal of Cryptosporidium and Giardia. *Vet Parasitol,* 126**,** 219-34.

BHAT, S. A., SINGH, N., TRIVEDI, A., KANSAL, P., GUPTA, P. & KUMAR, A. 2012. The mechanism of redox sensing in Mycobacterium tuberculosis. *Free Radic Biol Med,* 53**,** 1625-41.

BINGLE, L. E., EASTLAKE, J. L., MBAILEY, M. & GIBSON, W. C. 2001. A novel GFP approach for the analysis of genetic exchange in trypanosomes allowing the in situ detection of mating events. *Microbiology,* 147**,** 3231-40.

BIRKY, C. W., JR. 1996. Heterozygosity, heteromorphy, and phylogenetic trees in asexual eukaryotes. *Genetics,* 144**,** 427-37.

BIRKY, C. W., JR. 2005. Sex: is giardia doing it in the dark? *Curr Biol,* 15**,** R56-8.

BLOBEL, G. & DOBBERSTEIN, B. 1975. Transfer of proteins across membranes. I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma. *J Cell Biol,* 67**,** 835-51.

BOREHAM, P. F., PHILLIPS, R. E. & SHEPHERD, R. W. 1985. A comparison of the in-vitro activity of some 5-nitroimidazoles and other compounds against Giardia intestinalis. *J Antimicrob Chemother,* 16**,** 589-95.

BORGSTROM, B. 1988. Mode of action of tetrahydrolipstatin: a derivative of the naturally occurring lipase inhibitor lipstatin. *Biochim Biophys Acta,* 962**,** 308-16.

BOTERO-GARCES, J. H., GARCIA-MONTOYA, G. M., GRISALES-PATINO, D., AGUIRRE-ACEVEDO, D. C. & ALVAREZ-URIBE, M. C. 2009. Giardia intestinalis and nutritional status in children participating in the complementary nutrition program, Antioquia, Colombia, May to October 2006. *Rev Inst Med Trop Sao Paulo,* 51**,** 155-62.

BRANDT, T., BREITENSTEIN, S., VON DER HARDT, H. & TUMMLER, B. 1995. DNA concentration and length in sputum of patients with cystic fibrosis during inhalation with recombinant human DNase. *Thorax,* 50**,** 880-2.

BRASSEUR, P. & FAVENNEC, L. 1995. Two cases of giardiasis unsuccessfully treated by albendazole. *Parasite,* 2**,** 422.

BREATHNACH, A. S., MCHUGH, T. D. & BUTCHER, P. D. 2010. Prevalence and clinical correlations of genetic subtypes of Giardia lamblia in an urban setting. *Epidemiol Infect,* 138**,** 1459-67.

BROWN, D. M., UPCROFT, J. A., EDWARDS, M. R. & UPCROFT, P. 1998. Anaerobic bacterial metabolism in the ancient eukaryote Giardia duodenalis. *Int J Parasitol,* 28**,** 149-64.

BROWN, T. 2002. Chapter 3: Transcriptomes and Proteomes. *Genomes, Second Edition.* Oxford: Wiley-Liss.

BURET, A. G. 2007. Mechanisms of epithelial dysfunction in giardiasis. *Gut,* 56**,** 316-7.

BURET, A. G. 2008. Pathophysiology of enteric infections with Giardia duodenalius. *Parasite,* 15**,** 261-5.

CACCIO, S. M., BECK, R., LALLE, M., MARINCULIC, A. & POZIO, E. 2008. Multilocus genotyping of Giardia duodenalis reveals striking differences between assemblages A and B. *Int J Parasitol,* 38**,** 1523-31.

CACCIO, S. M. & RYAN, U. 2008. Molecular epidemiology of giardiasis. *Mol Biochem Parasitol,* 160**,** 75-80.

CACCIO, S. M. & SPRONG, H. 2010. Giardia duodenalis: genetic recombination and its implications for taxonomy and molecular epidemiology. *Exp Parasitol,* 124**,** 107-12.

CACCIO, S. M., THOMPSON, R. C., MCLAUCHLIN, J. & SMITH, H. V. 2005. Unravelling Cryptosporidium and Giardia epidemiology. *Trends Parasitol,* 21**,** 430-7.

CARNABY, S., BUTCHER, P. D., SUMMERBELL, C. D., NAEEM, A. & FARTHING, M. J. 1995. Minisatellites corresponding to the human polycore probes 33.6 and 33.15 in the genome of the most 'primitive' known eukaryote Giardia lamblia. *Gene,* 166**,** 167-72.

CARNABY, S., KATELARIS, P. H., NAEEM, A. & FARTHING, M. J. 1994. Genotypic heterogeneity within Giardia lamblia isolates demonstrated by M13 DNA fingerprinting. *Infect Immun,* 62**,** 1875-80.

CARRANZA, P. G. & LUJAN, H. D. 2010. New insights regarding the biology of Giardia lamblia. *Microbes Infect,* 12**,** 71-80.

CAVALIER-SMITH, T. 2002. Origins of the machinery of recombination and sex. *Heredity (Edinb),* 88**,** 125-41.

CAVALIER-SMITH, T. 2003. Protist phylogeny and the high-level classification of Protozoa. *European Journal of Protistology,* 39**,** 338-348.

CDC. *Prevention and Control- Giardiasis* [Online]. http://www.cdc.gov/parasites/giardia/prevent.html. Available: http://www.cdc.gov/parasites/giardia/prevent.html.

CDC 2013. Notifiable Diseases and Mortality Tables for 2004-2013.

CEVALLOS, A., CARNABY, S., JAMES, M. & FARTHING, J. G. 1995. Small intestinal injury in a neonatal rat model of giardiasis is strain dependent. *Gastroenterology,* 109**,** 766-73.

CHANTRET, I., BARBAT, A., DUSSAULX, E., BRATTAIN, M. G. & ZWEIBAUM, A. 1988. Epithelial polarity, villin expression, and enterocytic differentiation of cultured human colon carcinoma cells: a survey of twenty cell lines. *Cancer Res,* 48**,** 1936-42.

CHIN, A. C., TEOH, D. A., SCOTT, K. G., MEDDINGS, J. B., MACNAUGHTON, W. K. & BURET, A. G. 2002. Strain-dependent induction of enterocyte apoptosis by Giardia lamblia disrupts epithelial barrier function in a caspase-3-dependent manner. *Infect Immun,* 70**,** 3673-80.

CHIQUET-EHRISMANN, R. & CHIQUET, M. 2003. Tenascins: regulation and putative functions during pathological stress. *J Pathol,* 200**,** 488-99.

CIFUENTES, E., GOMEZ, M., BLUMENTHAL, U., TELLEZ-ROJO, M. M., ROMIEU, I., RUIZ-PALACIOS, G. & RUIZ-VELAZCO, S. 2000. Risk factors for Giardia intestinalis infection in agricultural villages practicing wastewater irrigation in Mexico. *Am J Trop Med Hyg,* 62**,** 388-92.

CLARK, J. T. & HOLBERTON, D. V. 1988. Triton-labile antigens in flagella isolated from Giardia lamblia. *Parasitol Res,* 74**,** 415-23.

COBRIN, G. M. & ABREU, M. T. 2005. Defects in mucosal immunity leading to Crohn's disease. *Immunol Rev,* 206**,** 277-95.

CONDEELIS, J. 1995. Elongation factor 1 alpha, translation and the cytoskeleton. *Trends Biochem Sci,* 20**,** 169-70.

CONNER, S. D. & SCHMID, S. L. 2003. Regulated portals of entry into the cell. *Nature,* 422**,** 37-44.

CONNOLLY, P. F., JAGER, R. & FEARNHEAD, H. O. 2014. New roles for old enzymes: killer caspases as the engine of cell behavior changes. *Front Physiol,* 5**,** 149.

COOPER, M. A., ADAM, R. D., WOROBEY, M. & STERLING, C. R. 2007. Population genetics provides evidence for recombination in Giardia. *Curr Biol,* 17**,** 1984-8.

COSTACHE, C. A., COLOSI, I. A. & COLOSI, H. A. 2009. Immunochromatography versus microscopy for the identification of *Giardia lamblia* and *Cryptosporidium parvum* in human feces *Sci Parasitol,* 10**,** 26-31.

COTTON, J. A., BEATTY, J. K. & BURET, A. G. 2011. Host parasite interactions and pathophysiology in Giardia infections. *Int J Parasitol,* 41**,** 925-33.

COTTON, J. A., BHARGAVA, A., FERRAZ, J. G., YATES, R. M., BECK, P. L. & BURET, A. G. 2014. Giardia duodenalis cathepsin B proteases degrade intestinal epithelial interleukin-8 and attenuate interleukin-8-induced neutrophil chemotaxis. *Infect Immun*.

COVACIN, C., AUCOIN, D. P., ELLIOT, A. & THOMPSON, R. C. 2011. Genotypic characterisation of Giardia from domestic dogs in the USA. *Vet Parasitol,* 177**,** 28-32.

CRAIG, E. A., GAMBILL, B. D. & NELSON, R. J. 1993. Heat shock proteins: molecular chaperones of protein biogenesis. *Microbiol Rev,* 57**,** 402-14.

CROSSLEY, R. & HOLBERTON, D. 1985. Assembly of 2.5 nm filaments from giardin, a protein associated with cytoskeletal microtubules in Giardia. *J Cell Sci,* 78**,** 205-31.

D'ANCHINO, M., ORLANDO, D. & DE FEUDIS, L. 2002. Giardia lamblia infections become clinically evident by eliciting symptoms of irritable bowel syndrome. *J Infect,* 45**,** 169-72.

DACKS, J. B. & DOOLITTLE, W. F. 2002. Novel syntaxin gene sequences from Giardia, Trypanosoma and algae: implications for the ancient evolution of the eukaryotic endomembrane system. *J Cell Sci,* 115**,** 1635-42.

DARBON, A., PORTAL, A., GIRIER, L., PANTIN, J. & LECLAIRE, C. 1962. [Treatment of giardiasis (lambliasis) with metronidazole. Apropos of 100 cases]. *La Presse medicale,* 70**,** 15-6.

DAS, S., STEVENS, T., CASTILLO, C., VILLASENOR, A., ARREDONDO, H. & REDDY, K. 2002. Lipid metabolism in mucous-dwelling amitochondriate protozoa. *Int J Parasitol,* 32**,** 655-75.

DAVIDS, B. J., REINER, D. S., BIRKELAND, S. R., PREHEIM, S. P., CIPRIANO, M. J., MCARTHUR, A. G. & GILLIN, F. D. 2006. A new family of giardial cysteine-rich non-VSP protein genes and a novel cyst protein. *PLoS One,* 1**,** e44.

DAWSON, D. 2005. Foodborne protozoan parasites. *Int J Food Microbiol,* 103**,** 207-27.

DEAN, M., HAMON, Y. & CHIMINI, G. 2001. The human ATP-binding cassette (ABC) transporter superfamily. *J Lipid Res,* 42**,** 1007-17.

DENG, M. Q. & CLIVER, D. O. 1999. Rapid DNA extraction methods and new primers for randomly amplified polymorphic DNA analysis of Giardia duodenalis. *J Microbiol Methods,* 37**,** 193-200.

DENKEL, L. A., HORST, S. A., ROUF, S. F., KITOWSKI, V., BOHM, O. M., RHEN, M., JAGER, T. & BANGE, F. C. 2011. Methionine sulfoxide reductases are essential for virulence of Salmonella typhimurium. *PLoS One,* 6**,** e26974.

DI MATTEO, A., SCANDURRA, F. M., TESTA, F., FORTE, E., SARTI, P., BRUNORI, M. & GIUFFRE, A. 2008. The O2-scavenging flavodiiron protein in the human parasite Giardia intestinalis. *J Biol Chem,* 283**,** 4061-8.

DOBELL, C. 1920. The Discovery of the Intestinal Protozoa of Man. *Proc R Soc Med,* 13**,** 1-15.

DOMANOVA, O., BORBE, S., MUHLFELD, S., BECKER, M., KUBITZ, R., HAUSSINGER, D. & BERLAGE, T. 2011. Toponomics method for the automated quantification of membrane protein translocation. *BMC Bioinformatics,* 12**,** 370.

DORSCH, M. R. & VEAL, D. A. 2001. Oligonucleotide probes for specific detection of Giardia lamblia cysts by fluorescent in situ hybridization. *J Appl Microbiol,* 90**,** 836-42.

DRUBIN, D. G. & NELSON, W. J. 1996. Origins of cell polarity. *Cell,* 84**,** 335-44.

DUBOIS, K. N., ABODEELY, M., SAJID, M., ENGEL, J. C. & MCKERROW, J. H. 2006. Giardia lamblia cysteine proteases. *Parasitol Res,* 99**,** 313-6.

DUBOIS, K. N., ABODEELY, M., SAKANARI, J., CRAIK, C. S., LEE, M., MCKERROW, J. H. & SAJID, M. 2008. Identification of the major cysteine protease of Giardia and its role in encystation. *J Biol Chem,* 283**,** 18024-31.

DUREL, P., ROIRON, V., SIBOULET, A. & BOREL, L. J. 1960. Systemic treatment of human trichomoniasis with a derivative of nitro-imidazole, 8823 RP. *The British journal of venereal diseases,* 36**,** 21-6.

ECDC 2010a. - Annual Epidemiological Report on Communicable Diseases in Europe 2010.

ECDC. 2010b. *Factsheet for the general public - Giardia* [Online]. http://www.ecdc.europa.eu/en/healthtopics/giardiasis/factsheet-general-public/Pages/Factsheet_general_public.aspx. Available: http://www.ecdc.europa.eu/en/healthtopics/giardiasis/factsheet-general-public/Pages/Factsheet_general_public.aspx.

ECDC 2012. Annual epidemiological report reporting on 2010 surveillance ata and 2011 epidemic data.

ECDC 2013. Annual epidemiological report reporting on 2011 surveillance data and 2012 epidemic intelligence data.

ECKMANN, L., LAURENT, F., LANGFORD, T. D., HETSKO, M. L., SMITH, J. R., KAGNOFF, M. F. & GILLIN, F. D. 2000. Nitric oxide production by human intestinal epithelial cells and competition for arginine as potential determinants of host defense against the lumen-dwelling pathogen Giardia lamblia. *J Immunol,* 164**,** 1478-87.

EDLIND, T. D. 1989. Susceptibility of Giardia lamblia to aminoglycoside protein synthesis inhibitors: correlation with rRNA structure. *Antimicrob Agents Chemother,* 33**,** 484-8.

ELIGIO-GARCIA, L., CORTES-CAMPOS, A., COTA-GUAJARDO, S., GAXIOLA, S. & JIMENEZ-CARDOSO, E. 2008. Frequency of Giardia intestinalis assemblages isolated from dogs and humans in a community from Culiacan, Sinaloa, Mexico using beta-giardin restriction gene. *Vet Parasitol,* 156**,** 205-9.

ELLIS, J. G. T., DAVILA, M. & CHAKRABARTI, R. 2003. Potential involvement of extracellular signal-regulated kinase 1 and 2 in encystation of a primitive eukaryote, Giardia lamblia. Stage-specific activation and intracellular localization. *J Biol Chem,* 278**,** 1936-45.

ELMENDORF, H. G., DAWSON, S. C. & MCCAFFERY, J. M. 2003. The cytoskeleton of Giardia lamblia. *Int J Parasitol,* 33**,** 3-28.

ELNIFRO, E. M., ASHSHI, A. M., COOPER, R. J. & KLAPPER, P. E. 2000. Multiplex PCR: optimization and application in diagnostic virology. *Clin Microbiol Rev,* 13**,** 559-70.

ELSAFI, S. H., AL-MAQATI, T. N., HUSSEIN, M. I., ADAM, A. A., HASSAN, M. M. & AL ZAHRANI, E. M. 2013. Comparison of microscopy, rapid immunoassay, and molecular techniques for the detection of Giardia lamblia and Cryptosporidium parvum. *Parasitol Res,* 112**,** 1641-6.

ERLANDSEN, S. L., BEMRICK, W. J., SCHUPP, D. E., SHIELDS, J. M., JARROLL, E. L., SAUCH, J. F. & PAWLEY, J. B. 1990. High-resolution immunogold localization of Giardia cyst wall antigens using field emission SEM with secondary and backscatter electron imaging. *J Histochem Cytochem,* 38**,** 625-32.

ERLANDSEN, S. L., JARROLL, E., WALLIS, P. & VAN KEULEN, H. 2005. Development of species-specific rDNA probes for Giardia by multiple fluorescent in situ hybridization combined with immunocytochemical identification of cyst wall antigens. *J Histochem Cytochem,* 53**,** 917-27.

ESCOBEDO, A. A. & NUNEZ, F. A. 1999. Prevalence of intestinal parasites in Cuban acquired immunodeficiency syndrome (AIDS) patients. *Acta Trop,* 72**,** 125-30.

FARTHING, M. J. 1992. Giardia comes of age: progress in epidemiology, immunology and chemotherapy. *J Antimicrob Chemother,* 30**,** 563-6.

FARTHING, M. J., PEREIRA, M. E. & KEUSCH, G. T. 1986. Description and characterization of a surface lectin from Giardia lamblia. *Infect Immun,* 51**,** 661-7.

FASO, C., BISCHOF, S. & HEHL, A. B. 2013. The proteome landscape of Giardia lamblia encystation. *PLoS One,* 8**,** e83207.

FAUBERT, G. 2000. Immune response to Giardia duodenalis. *Clin Microbiol Rev,* 13**,** 35-54, table of contents.

FELIZIANI, C., MERINO, M. C., RIVERO, M. R., HELLMAN, U., PISTORESI-PALENCIA, M. C. & ROPOLO, A. S. 2011. Immunodominant proteins alpha-1 giardin and beta-giardin are expressed in both assemblages A and B of Giardia lamblia. *BMC Microbiol,* 11**,** 233.

FENG, Y. & XIAO, L. 2011. Zoonotic potential and molecular epidemiology of Giardia species and giardiasis. *Clin Microbiol Rev,* 24**,** 110-40.

FILICE, F. P. 1952. Studies on the cytology and life history of a Giardia from the laboratory rat. *In:* ZOOLOGY, U. O. C. P. I. (ed.) *Studies on the cytology and life history of a Giardia from the laboratory rat.* University of California Press: Berkeley.

FLANAGAN, P. A. 1992. Giardia--diagnosis, clinical course and epidemiology. A review. *Epidemiol Infect,* 109**,** 1-22.

FLUEGGE, K., SCHWEIER, O., SCHILTZ, E., BATSFORD, S. & BERNER, R. 2004. Identification and immunoreactivity of proteins released from Streptococcus agalactiae. *Eur J Clin Microbiol Infect Dis,* 23**,** 818-24.

FORD, B. J. 2005. The discovery of Giardia. *Microscope,* 53**,** 147-53.

FRANZEN, O., JERLSTROM-HULTQVIST, J., CASTRO, E., SHERWOOD, E., ANKARKLEV, J., REINER, D. S., PALM, D., ANDERSSON, J. O., ANDERSSON, B. & SVARD, S. G. 2009. Draft genome sequencing of giardia intestinalis assemblage B isolate GS: is human giardiasis caused by two different species? *PLoS Pathog,* 5**,** e1000560.

FRANZEN, O., JERLSTROM-HULTQVIST, J., EINARSSON, E., ANKARKLEV, J., FERELLA, M., ANDERSSON, B. & SVARD, S. G. 2013. Transcriptome profiling of Giardia intestinalis using strand-specific RNA-seq. *PLoS Comput Biol,* 9**,** e1003000.

GARCIA, L. S. & GARCIA, J. P. 2006. Detection of Giardia lamblia antigens in human fecal specimens by a solid-phase qualitative immunochromatographic assay. *J Clin Microbiol,* 44**,** 4587-8.

GARDNER, T. B. & HILL, D. R. 2001. Treatment of giardiasis. *Clinical microbiology reviews,* 14**,** 114-28.

GAUNT, M. W., YEO, M., FRAME, I. A., SLOTHARD, J. R., CARRASCO, H. J., TAYLOR, M. C., MENA, S. S., VEAZEY, P., MILES, G. A., ASCOTA, N., DE ARIAS, A. R. & MILES, M. A. 2003. Mechanism of genetic exchange in American trypanosomes. *Nature,* 421**,** 936-9.

GELANEW, T., LALLE, M., HAILU, A., POZIO, E. & CACCIO, S. M. 2007. Molecular characterization of human isolates of Giardia duodenalis from Ethiopia. *Acta Trop,* 102**,** 92-9.

GILBERT, H. F. 1997. Protein disulfide isomerase and assisted protein folding. *J Biol Chem,* 272**,** 29399-402.

GILLIN, F. D. & DIAMOND, L. S. 1981. Inhibition of clonal growth of Giardia lamblia and Entamoeba histolytica by metronidazole, quinacrine, and other antimicrobial agents. *J Antimicrob Chemother,* 8**,** 305-16.

GONI, P., MARTIN, B., VILLACAMPA, M., GARCIA, A., SERAL, C., CASTILLO, F. J. & CLAVEL, A. 2012. Evaluation of an immunochromatographic dip strip test for simultaneous detection of Cryptosporidium spp, Giardia duodenalis, and Entamoeba histolytica antigens in human faecal samples. *Eur J Clin Microbiol Infect Dis,* 31**,** 2077-82.

GORDTS, B., HEMELHOF, W., ASSELMAN, C. & BUTZLER, J. P. 1985. In vitro susceptibilities of 25 Giardia lamblia isolates of human origin to six commonly used antiprotozoal agents. *Antimicrob Agents Chemother,* 28**,** 378-80.

GOTTIG, N., ELIAS, E. V., QUIROGA, R., NORES, M. J., SOLARI, A. J., TOUZ, M. C. & LUJAN, H. D. 2006. Active and passive mechanisms drive secretory granule biogenesis during differentiation of the intestinal parasite Giardia lamblia. *J Biol Chem,* 281**,** 18156-66.

GOTTLIEB, M. 1989. The surface membrane 3'-nucleotidase/nuclease of trypanosomatid protozoa. *Parasitol Today,* 5**,** 257-60.

GRASSET, E., PINTO, M., DUSSAULX, E., ZWEIBAUM, A. & DESJEUX, J. F. 1984. Epithelial properties of human colonic carcinoma cell line Caco-2: electrical parameters. *Am J Physiol,* 247**,** C260-7.

GRAZIOLI, B., MATERA, G., LARATTA, C., SCHIPANI, G., GUARNIERI, G., SPINIELLO, E., IMENEO, M., AMOROSI, A., FOCA, A. & LUZZA, F. 2006. Giardia lamblia infection in patients with irritable bowel syndrome and dyspepsia: a prospective study. *World J Gastroenterol,* 12**,** 1941-4.

GUICHARD, A., CRUZ-MORENO, B., AGUILAR, B., VAN SORGE, N. M., KUANG, J., KURKCIYAN, A. A., WANG, Z., HANG, S., PINETON DE CHAMBRUN, G. P., MCCOLE, D. F., WATNICK, P., NIZET, V. & BIER, E. 2013. Cholera toxin disrupts barrier function by inhibiting exocyst-mediated trafficking of host proteins to intestinal cell junctions. *Cell Host Microbe,* 14**,** 294-305.

HADRYS, H., BALICK, M. & SCHIERWATER, B. 1992. Applications of random amplified polymorphic DNA (RAPD) in molecular ecology. *Mol Ecol,* 1**,** 55-63.

HAHN, J., SEEBER, F., KOLODZIEJ, H., IGNATIUS, R., LAUE, M., AEBISCHER, T. & KLOTZ, C. 2013. High sensitivity of Giardia duodenalis to tetrahydrolipstatin (orlistat) in vitro. *PloS one,* 8**,** e71597.

HALL, P. A., COATES, P. J., ANSARI, B. & HOPWOOD, D. 1994. Regulation of cell number in the mammalian gastrointestinal tract: the importance of apoptosis. *J Cell Sci,* 107 (Pt 12)**,** 3569-77.

HAMPL, V., HUG, L., LEIGH, J. W., DACKS, J. B., LANG, B. F., SIMPSON, A. G. & ROGER, A. J. 2009. Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups". *Proc Natl Acad Sci U S A,* 106**,** 3859-64.

HAQUE, R., ROY, S., SIDDIQUE, A., MONDAL, U., RAHMAN, S. M., MONDAL, D., HOUPT, E. & PETRI, W. A., JR. 2007. Multiplex real-time PCR assay for detection of Entamoeba histolytica, Giardia intestinalis, and Cryptosporidium spp. *Am J Trop Med Hyg,* 76**,** 713-7.

HARMS, H. K., MATOUK, E., TOURNIER, G., VON DER HARDT, H., WELLER, P. H., ROMANO, L., HEIJERMAN, H. G., FITZGERALD, M. X., RICHARD, D., STRANDVIK, B., KOLBE, J., KRAEMER, R. & MICHALSEN, H. 1998. Multicenter, open-label study of recombinant human DNase in cystic fibrosis patients with moderate lung disease. DNase International Study Group. *Pediatr Pulmonol,* 26**,** 155-61.

HARRIS, J. C., PLUMMER, S. & LLOYD, D. 2001. Antigiardial drugs. *Appl Microbiol Biotechnol,* 57**,** 614-9.

HEHL, A. B. & MARTI, M. 2004. Secretory protein trafficking in Giardia intestinalis. *Mol Microbiol,* 53**,** 19-28.

HEHL, A. B., MARTI, M. & KOHLER, P. 2000. Stage-specific expression and targeting of cyst wall protein-green fluorescent protein chimeras in Giardia. *Mol Biol Cell,* 11**,** 1789-800.

HEITMAN, J. 2006. Sexual reproduction and the evolution of microbial pathogens. *Curr Biol,* 16**,** R711-25.

HELED, J. & DRUMMOND, A. J. 2010. Bayesian inference of species trees from multilocus data. *Mol Biol Evol,* 27**,** 570-80.

HERBST, E. W., ARMBRUSTER, M., RUMP, J. A., BUSCHER, H. P. & PETER, H. H. 1994. Intestinal B cell defects in common variable immunodeficiency. *Clin Exp Immunol,* 95**,** 215-21.

HERMASZEWSKI, R. A. & WEBSTER, A. D. 1993. Primary hypogammaglobulinaemia: a survey of clinical manifestations and complications. *Q J Med,* 86**,** 31-42.

HEYWOOD, P. & MAGEE, P. T. 1976. Meiosis in protists. Some structural and physiological aspects of meiosis in algae, fungi, and protozoa. *Bacteriol Rev,* 40**,** 190-240.

HILTPOLD, A., FREY, M., HULSMEIER, A. & KOHLER, P. 2000. Glycosylation and palmitoylation are common modifications of giardia variant surface proteins. *Mol Biochem Parasitol,* 109**,** 61-5.

HODGES, K. & GILL, R. 2010. Infectious diarrhea: Cellular and molecular mechanisms. *Gut Microbes,* 1**,** 4-21.

HOLBERTON, D. V. 1981. Arrangement of subunits in microribbons from Giardia. *J Cell Sci,* 47**,** 167-85.

HOLM, L. M., JAHN, T. P., MOLLER, A. L., SCHJOERRING, J. K., FERRI, D., KLAERKE, D. A. & ZEUTHEN, T. 2005. NH3 and NH4+ permeability in aquaporin-expressing Xenopus oocytes. *Pflugers Arch,* 450**,** 415-28.

HOMAN, W. L. & MANK, T. G. 2001. Human giardiasis: genotype linked differences in clinical symptomatology. *Int J Parasitol,* 31**,** 822-6.

HOYNE, G. F., BOREHAM, P. F., PARSONS, P. G., WARD, C. & BIGGS, B. 1989. The effect of drugs on the cell cycle of Giardia intestinalis. *Parasitology,* 99 Pt 3**,** 333-9.

HUEY, C. S., MAHDY, M. A., AL-MEKHLAFI, H. M., NASR, N. A., LIM, Y. A., MAHMUD, R. & SURIN, J. 2013. Multilocus genotyping of Giardia duodenalis in Malaysia. *Infect Genet Evol,* 17**,** 269-76.

HUNTER, P. R. & THOMPSON, R. C. 2005. The zoonotic transmission of Giardia and Cryptosporidium. *Int J Parasitol,* 35**,** 1181-90.

IANELLI, C. J., DELELLIS, R. & THORLEY-LAWSON, D. A. 1998. CD48 binds to heparan sulfate on the surface of epithelial cells. *J Biol Chem,* 273**,** 23367-75.

INVS 2012. Institut de Veille Sanitaire -Bulletin epidemiologique hebdomadaire: Recommendations sanitaires pour les voyageurs, 2012.

JABAJI, Z., SEARS, C. M., BRINKLEY, G. J., LEI, N. Y., JOSHI, V. S., WANG, J., LEWIS, M., STELZNER, M., MARTIN, M. G. & DUNN, J. C. 2013. Use of collagen gel as an alternative extracellular matrix for the in vitro and in vivo growth of murine small intestinal epithelium. *Tissue Eng Part C Methods,* 19**,** 961-9.

JACKSON, A. P., OTTO, T. D., DARBY, A., RAMAPRASAD, A., XIA, D., ECHAIDE, I. E., FARBER, M., GAHLOT, S., GAMBLE, J., GUPTA, D., GUPTA, Y., JACKSON, L., MALANDRIN, L., MALAS, T. B., MOUSSA, E., NAIR, M., REID, A. J., SANDERS, M., SHARMA, J., TRACEY, A., QUAIL, M. A., WEIR, W., WASTLING, J. M., HALL, N., WILLADSEN, P., LINGELBACH, K., SHIELS, B., TAIT, A., BERRIMAN, M., ALLRED, D. R. & PAIN, A. 2014. The evolutionary dynamics of variant antigen genes in Babesia reveal a history of genomic innovation underlying host-parasite interaction. *Nucleic Acids Res*.

JASINSKI, M., STUKKENS, Y., DEGAND, H., PURNELLE, B., MARCHAND-BRYNAERT, J. & BOUTRY, M. 2001. A plant plasma membrane ATP binding cassette-type transporter is involved in antifungal terpenoid secretion. *Plant Cell,* 13**,** 1095-107.

JEFFREYS, A. J., WILSON, V. & THEIN, S. L. 1985. Hypervariable 'minisatellite' regions in human DNA. *Nature,* 314**,** 67-73.

JENKINS, M. C., O'BRIEN, C. N., MURPHY, C., SCHWARZ, R., MISKA, K., ROSENTHAL, B. & TROUT, J. M. 2009. Antibodies to the ventral disc protein delta-giardin prevent in vitro binding of Giardia lamblia trophozoites. *J Parasitol,* 95**,** 895-9.

JERLSTROM-HULTQVIST, J., FRANZEN, O., ANKARKLEV, J., XU, F., NOHYNKOVA, E., ANDERSSON, J. O., SVARD, S. G. & ANDERSSON, B. 2010. Genome analysis and comparative genomics of a Giardia intestinalis assemblage E isolate. *BMC Genomics,* 11**,** 543.

JEX, A. R., STANLEY, K. K., LO, W., LITTMAN, R., VERWEIJ, J. J., CAMPBELL, B. E., NOLAN, M. J., PANGASA, A., STEVENS, M. A., HAYDON, S. & GASSER, R. B. 2012. Detection of diarrhoeal pathogens in human faeces using an automated, robotic platform. *Mol Cell Probes,* 26**,** 11-5.

JOHNSTON, B. C., DONEN, R., POONI, A., POND, J., XIE, F., GIGLIA, L., KAM, A., BHAMBER, A., BAMI, K., PATEL, Y. & GUYATT, G. H. 2013. Conceptual framework for health-related quality of life assessment in acute gastroenteritis. *J Pediatr Gastroenterol Nutr,* 56**,** 280-9.

JOHNSTON, S. P., BALLARD, M. M., BEACH, M. J., CAUSER, L. & WILKINS, P. P. 2003. Evaluation of three commercial assays for detection of Giardia and Cryptosporidium organisms in fecal specimens. *J Clin Microbiol,* 41**,** 623-6.

JUNG, H. C., ECKMANN, L., YANG, S. K., PANJA, A., FIERER, J., MORZYCKA-WROBLEWSKA, E. & KAGNOFF, M. F. 1995. A distinct array of proinflammatory cytokines is expressed in human colon epithelial cells in response to bacterial invasion. *J Clin Invest,* 95**,** 55-65.

KAMITANI, T., KITO, K., NGUYEN, H. P., FUKUDA-KAMITANI, T. & YEH, E. T. 1998. Characterization of a second member of the sentrin family of ubiquitin-like proteins. *J Biol Chem,* 273**,** 11349-53.

KARANIS, P., KOURENTI, C. & SMITH, H. 2007. Waterborne transmission of protozoan parasites: a worldwide review of outbreaks and lessons learnt. *J Water Health,* 5**,** 1-38.

KATELARIS, P. H. & FARTHING, M. J. 1992. Diarrhoea and malabsorption in giardiasis: a multifactorial process? *Gut,* 33**,** 295-7.

KATIYAR, S. K., GORDON, V. R., MCLAUGHLIN, G. L. & EDLIND, T. D. 1994. Antiprotozoal activities of benzimidazoles and correlations with beta-tubulin sequence. *Antimicrob Agents Chemother,* 38**,** 2086-90.

KAVOUSI, S. 1979. Giardiasis in infancy and childhood: a prospective study of 160 cases with comparison of quinacrine (Atabrine) and metronidazole (Flagyl). *Am J Trop Med Hyg,* 28**,** 19-23.

KEELING, P. J., BURGER, G., DURNFORD, D. G., LANG, B. F., LEE, R. W., PEARLMAN, R. E., ROGER, A. J. & GRAY, M. W. 2005. The tree of eukaryotes. *Trends Ecol Evol,* 20**,** 670-6.

KEISTER, D. B. 1983. Axenic culture of Giardia lamblia in TYI-S-33 medium supplemented with bile. *Trans R Soc Trop Med Hyg,* 77**,** 487-8.

KELLOGG, J. A. & ELDER, C. J. 1999. Justification for use of a single trichrome stain as the sole means for routine detection of intestinal parasites in concentrated stool specimens. *J Clin Microbiol,* 37**,** 835-7.

KIEDROWSKI, M. R., CROSBY, H. A., HERNANDEZ, F. J., MALONE, C. L., MCNAMARA, J. O., 2ND & HORSWILL, A. R. 2014. Staphylococcus aureus Nuc2 Is a Functional, Surface-Attached Extracellular Nuclease. *PLoS One,* 9**,** e95574.

KIM, H. Y. & GLADYSHEV, V. N. 2004. Methionine sulfoxide reduction in mammals: characterization of methionine-R-sulfoxide reductases. *Mol Biol Cell,* 15**,** 1055-64.

KIM, K. T., MOK, M. T. & EDWARDS, M. R. 2005. Protein kinase B from Giardia intestinalis. *Biochem Biophys Res Commun,* 334**,** 333-41.

KNODLER, L. A., NOIVA, R., MEHTA, K., MCCAFFERY, J. M., ALEY, S. B., SVARD, S. G., NYSTUL, T. G., REINER, D. S., SILBERMAN, J. D. & GILLIN, F. D. 1999. Novel protein-disulfide isomerases from the early-diverging protist Giardia lamblia. *J Biol Chem,* 274**,** 29805-11.

KOEHLER, A. V., JEX, A. R., HAYDON, S. R., STEVENS, M. A. & GASSER, R. B. 2013. Giardia/giardiasis - A perspective on diagnostic and analytical tools. *Biotechnol Adv*.

KOFOID, C. A. & CHRISTIANSEN, E. B. 1915. On the Life-History of Giardia. *Proc Natl Acad Sci U S A,* 1**,** 547-52.

KORCHEV, Y. E., GORELIK, J., LAB, M. J., SVIDERSKAYA, E. V., JOHNSTON, C. L., COOMBES, C. R., VODYANOY, I. & EDWARDS, C. R. 2000. Cell volume measurement using scanning ion conductance microscopy. *Biophys J,* 78**,** 451-7.

KOSHY, S., ALIZADEH, P., TIMCHENKO, L. T. & BEETON, C. 2010. Quantitative measurement of GLUT4 translocation to the plasma membrane by flow cytometry. *J Vis Exp*.

KOTLER, B. M., KERSTETTER, J. E. & INSOGNA, K. L. 2013. Claudins, dietary milk proteins, and intestinal barrier regulation. *Nutr Rev,* 71**,** 60-5.

KOTLOFF, K. L., NATARO, J. P., BLACKWELDER, W. C., NASRIN, D., FARAG, T. H., PANCHALINGAM, S., WU, Y., SOW, S. O., SUR, D., BREIMAN, R. F., FARUQUE, A. S., ZAIDI, A. K., SAHA, D., ALONSO, P. L., TAMBOURA, B., SANOGO, D., ONWUCHEKWA, U., MANNA, B., RAMAMURTHY, T., KANUNGO, S., OCHIENG, J. B., OMORE, R., OUNDO, J. O., HOSSAIN, A., DAS, S. K., AHMED, S., QURESHI, S., QUADRI, F., ADEGBOLA, R. A., ANTONIO, M., HOSSAIN, M. J., AKINSOLA, A., MANDOMANDO, I., NHAMPOSSA, T., ACACIO, S., BISWAS, K., O'REILLY, C. E., MINTZ, E. D., BERKELEY, L. Y., MUHSEN, K., SOMMERFELT, H., ROBINS-BROWNE, R. M. & LEVINE, M. M. 2013. Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. *Lancet,* 382**,** 209-22.

KRANE, C. M. & GOLDSTEIN, D. L. 2007. Comparative functional analysis of aquaporins/glyceroporins in mammals and anurans. *Mamm Genome,* 18**,** 452-62.

KREUTNER, A. K., DEL BENE, V. E. & AMSTEY, M. S. 1981. Giardiasis in pregnancy. *Am J Obstet Gynecol,* 140**,** 895-901.

KUCERS, A., CROWE, S. M., GRAYSON, L. M. & HOY, J. F. 1997. Bacitracin and gramicidin. *In:* HOY, J. F. (ed.) *The use of antibiotics. A clinical review of antibacterial, antifungal, and antiviral drugs, 5th edition*. Oxford, United Kingdom: Butterworth-Heinemann.

KULAKOVA, L., SINGER, S. M., CONRAD, J. & NASH, T. E. 2006. Epigenetic mechanisms are involved in the control of Giardia lamblia antigenic variation. *Mol Microbiol,* 61**,** 1533-42.

LACEY, E. 1988. The role of the cytoskeletal protein, tubulin, in the mode of action and mechanism of drug resistance to benzimidazoles. *Int J Parasitol,* 18**,** 885-936.

LAEMMLI, U. K., BEGUIN, F. & GUJER-KELLENBERGER, G. 1970. A factor preventing the major head protein of bacteriophage T4 from random aggregation. *J Mol Biol,* 47**,** 69-85.

LANFREDI-RANGEL, A., ATTIAS, M., DE CARVALHO, T. M., KATTENBACH, W. M. & DE SOUZA, W. 1998. The peripheral vesicles of trophozoites of the primitive protozoan Giardia

lamblia may correspond to early and late endosomes and to lysosomes. *J Struct Biol,* 123**,** 225-35.

LANGFORD, T. D., SILBERMAN, J. D., WEILAND, M. E., SVARD, S. G., MCCAFFERY, J. M., SOGIN, M. L. & GILLIN, F. D. 2002. Giardia lamblia: identification and characterization of Rab and GDI proteins in a genome survey of the ER to Golgi endomembrane system. *Exp Parasitol,* 101**,** 13-24.

LASEK-NESSELQUIST, E., WELCH, D. M. & SOGIN, M. L. 2010. The identification of a new Giardia duodenalis assemblage in marine vertebrates and a preliminary analysis of G. duodenalis population biology in marine systems. *Int J Parasitol,* 40**,** 1063-74.

LASEK-NESSELQUIST, E., WELCH, D. M., THOMPSON, R. C., STEUART, R. F. & SOGIN, M. L. 2009. Genetic exchange within and between assemblages of Giardia duodenalis. *J Eukaryot Microbiol,* 56**,** 504-18.

LAU, A. H., LAM, N. P., PISCITELLI, S. C., WILKES, L. & DANZIGER, L. H. 1992. Clinical pharmacokinetics of metronidazole and other nitroimidazole anti-infectives. *Clin Pharmacokinet,* 23**,** 328-64.

LAUWAET, T., DAVIDS, B. J., TORRES-ESCOBAR, A., BIRKELAND, S. R., CIPRIANO, M. J., PREHEIM, S. P., PALM, D., SVARD, S. G., MCARTHUR, A. G. & GILLIN, F. D. 2007. Protein phosphatase 2A plays a crucial role in Giardia lamblia differentiation. *Mol Biochem Parasitol,* 152**,** 80-9.

LE BORGNE, R. & HOFLACK, B. 1998. Protein transport from the secretory to the endocytic pathway in mammalian cells. *Biochim Biophys Acta,* 1404**,** 195-209.

LEBBAD, M., MATTSSON, J. G., CHRISTENSSON, B., LJUNGSTROM, B., BACKHANS, A., ANDERSSON, J. O. & SVARD, S. G. 2010. From mouse to moose: multilocus genotyping of Giardia isolates from various animal species. *Veterinary parasitology,* 168**,** 231-9.

LEBWOHL, B., DECKELBAUM, R. J. & GREEN, P. H. 2003. Giardiasis. *Gastrointest Endosc,* 57**,** 906-13.

LECHEVALLIER, M. W., NORTON, W. D. & LEE, R. G. 1991. Giardia and Cryptosporidium spp. in filtered drinking water supplies. *Appl Environ Microbiol,* 57**,** 2617-21.

LEITSCH, D., BURGESS, A. G., DUNN, L. A., KRAUER, K. G., TAN, K., DUCHENE, M., UPCROFT, P., ECKMANN, L. & UPCROFT, J. A. 2011. Pyruvate:ferredoxin oxidoreductase and thioredoxin reductase are involved in 5-nitroimidazole activation while flavin metabolism is linked to 5-nitroimidazole resistance in Giardia lamblia. *J Antimicrob Chemother,* 66**,** 1756-65.

LEMOS, V., GRACZYK, T. K., ALVES, M., LOBO, M. L., SOUSA, M. C., ANTUNES, F. & MATOS, O. 2005. Identification and determination of the viability of Giardia lamblia cysts and Cryptosporidium parvum and Cryptosporidium hominis oocysts in human fecal and water supply samples by fluorescent in situ hybridization (FISH) and monoclonal antibodies. *Parasitol Res,* 98**,** 48-53.

LEV, B., WARD, H., KEUSCH, G. T. & PEREIRA, M. E. 1986. Lectin activation in Giardia lamblia by host protease: a novel host-parasite interaction. *Science,* 232**,** 71-3.

LI, Z., KULAKOVA, L., LI, L., GALKIN, A., ZHAO, Z., NASH, T. E., MARIANO, P. S., HERZBERG, O. & DUNAWAY-MARIANO, D. 2009. Mechanisms of catalysis and inhibition operative in the arginine deiminase from the human pathogen Giardia lamblia. *Bioorg Chem,* 37**,** 149-61.

LINGDAN, L., PENGTAO, G., WENCHAO, L., JIANHUA, L., JU, Y., CHENGWU, L., HE, L., GUOCAI, Z., WENZHI, R., YUJIANG, C. & XICHEN, Z. 2012. Differential dissolved protein expression throughout the life cycle of Giardia lamblia. *Exp Parasitol,* 132**,** 465-9.

LINTON, K. J. & HIGGINS, C. F. 1998. The Escherichia coli ATP-binding cassette (ABC) proteins. *Mol Microbiol,* 28**,** 5-13.

LIU, J., GRATZ, J., AMOUR, C., KIBIKI, G., BECKER, S., JANAKI, L., VERWEIJ, J. J., TANIUCHI, M., SOBUZ, S. U., HAQUE, R., HAVERSTICK, D. M. & HOUPT, E. R. 2013. A laboratory-developed TaqMan Array Card for simultaneous detection of 19 enteropathogens. *J Clin Microbiol,* 51**,** 472-80.

LIU, L. X. & WELLER, P. F. 1996. Antiparasitic drugs. *N Engl J Med,* 334**,** 1178-84.

LODISH, H., BERK, A., ZIPURSKY, J., MATSUDAIRA, P., BALTIMORE, D. & DARNELL, J. (eds.) 2000. *Molecular Cell biology, Fourth Edition,* New York: W. H. Freeman.

LOGDSON, J. M. J. 2008. Evolutionary genetics: sex happens in Giardia. *Curr Biol,* 18**,** R66-8.

LOPEZ-RUBIO, J. J., RIVIERE, L. & SCHERF, A. 2007. Shared epigenetic mechanisms control virulence factors in protozoan parasites. *Curr Opin Microbiol,* 10**,** 560-8.

LOPEZ, M. F., BERGGREN, K., CHERNOKALSKAYA, E., LAZAREV, A., ROBINSON, M. & PATTON, W. F. 2000. A comparison of silver stain and SYPRO Ruby Protein Gel Stain with respect to protein detection in two-dimensional gels and identification by peptide mass profiling. *Electrophoresis,* 21**,** 3673-83.

LOVERDE, P. T., OSMAN, A. & HINCK, A. 2007. Schistosoma mansoni: TGF-beta signaling pathways. *Exp Parasitol,* 117**,** 304-17.

LUCAS, M. L. 2010. Diarrhoeal disease through enterocyte secretion: a doctrine untroubled by proof. *Exp Physiol,* 95**,** 479-84.

LUJAN, H. D., MAROTTA, A., MOWATT, M. R., SCIAKY, N., LIPPINCOTT-SCHWARTZ, J. & NASH, T. E. 1995a. Developmental induction of Golgi structure and function in the primitive eukaryote Giardia lamblia. *J Biol Chem,* 270**,** 4612-8.

LUJAN, H. D., MOWATT, M. R., BYRD, L. G. & NASH, T. E. 1996. Cholesterol starvation induces differentiation of the intestinal parasite Giardia lamblia. *Proc Natl Acad Sci U S A,* 93**,** 7628-33.

LUJAN, H. D., MOWATT, M. R., CONRAD, J. T., BOWERS, B. & NASH, T. E. 1995b. Identification of a novel Giardia lamblia cyst wall protein with leucine-rich repeats. Implications for secretory granule formation and protein assembly into the cyst wall. *J Biol Chem,* 270**,** 29307-13.

LUJAN, H. D., MOWATT, M. R. & NASH, T. E. 1997. Mechanisms of Giardia lamblia differentiation into cysts. *Microbiol Mol Biol Rev,* 61**,** 294-304.

LUJAN, H. D., MOWATT, M. R. & NASH, T. E. 1998. The molecular mechanisms of giardia encystation. *Parasitol Today,* 14**,** 446-50.

LUJAN, H. D. & TOUZ, M. C. 2003. Protein trafficking in Giardia lamblia. *Cell Microbiol,* 5**,** 427-34.

LUO, J., TENG, M., ZHANG, G. P., LUN, Z. R., ZHOU, H. & QU, L. H. 2009. Evaluating the evolution of G. lamblia based on the small nucleolar RNAs identified from Archaea and unicellular eukaryotes. *Parasitol Res,* 104**,** 1543-6.

MALYS, N. & MCCARTHY, J. E. 2011. Translation initiation: variations in the mechanism can be anticipated. *Cell Mol Life Sci,* 68**,** 991-1003.

MARTI, M. & HEHL, A. B. 2003. Encystation-specific vesicles in Giardia: a primordial Golgi or just another secretory compartment? *Trends Parasitol,* 19**,** 440-6.

MARTI, M., LI, Y., SCHRANER, E. M., WILD, P., KOHLER, P. & HEHL, A. B. 2003. The secretory apparatus of an ancient eukaryote: protein sorting to separate export pathways occurs before formation of transient Golgi-like compartments. *Mol Biol Cell,* 14**,** 1433-47.

MARTINEZ, M., PALACIOS, J. M., IMPERIAL, J. & RUIZ-ARGUESO, T. 2004. Symbiotic autoregulation of nifA expression in Rhizobium leguminosarum bv. viciae. *J Bacteriol,* 186**,** 6586-94.

MATSUZAKI, T., TAJIKA, Y., ABLIMIT, A., AOKI, T., HAGIWARA, H. & TAKATA, K. 2004. Aquaporins in the digestive system. *Med Electron Microsc,* 37**,** 71-80.

MAYRHOFER, G., ANDREWS, R. H., EY, P. L. & CHILTON, N. B. 1995. Division of Giardia isolates from humans into two genetically distinct assemblages by electrophoretic analysis of enzymes encoded at 27 loci and comparison with Giardia muris. *Parasitology,* 111 (Pt 1)**,** 11-7.

MCCAFFERY, J. M., FAUBERT, G. M. & GILLIN, F. D. 1994. Giardia lamblia: traffic of a trophozoite variant surface protein and a major cyst wall epitope during growth, encystation, and antigenic switching. *Exp Parasitol,* 79**,** 236-49.

MEDZHITOV, R. & JANEWAY, C., JR. 2000. Innate immune recognition: mechanisms and pathways. *Immunol Rev,* 173**,** 89-97.

MELLMAN, I. & SIMONS, K. 1992. The Golgi complex: in vitro veritas? *Cell,* 68**,** 829-40.

MEYER, E. A. 1970. Isolation and axenic cultivation of Giardia trophozoites from the rabbit, chinchilla, and cat. *Exp Parasitol,* 27**,** 179-83.

MING, L. J. & EPPERSON, J. D. 2002. Metal binding and structure-activity relationship of the metalloantibiotic peptide bacitracin. *J Inorg Biochem,* 91**,** 46-58.

MONIS, P. T., ANDREWS, R. H., MAYRHOFER, G. & EY, P. L. 1999. Molecular systematics of the parasitic protozoan Giardia intestinalis. *Mol Biol Evol,* 16**,** 1135-44.

MONIS, P. T., ANDREWS, R. H., MAYRHOFER, G. & EY, P. L. 2003. Genetic diversity within the morphological species Giardia intestinalis and its relationship to host origin. *Infect Genet Evol,* 3**,** 29-38.

MONIS, P. T., CACCIO, S. M. & THOMPSON, R. C. 2009. Variation in Giardia: towards a taxonomic revision of the genus. *Trends Parasitol,* 25**,** 93-100.

MONIS, P. T. & THOMPSON, R. C. 2003. Cryptosporidium and Giardia-zoonoses: fact or fiction? *Infect Genet Evol,* 3**,** 233-44.

MORAIS, E. R., OLIVEIRA, K. C., MAGALHAES, L. G., MOREIRA, E. B., VERJOVSKI-ALMEIDA, S. & RODRIGUES, V. 2013. Effects of curcumin on the parasite Schistosoma mansoni: a transcriptomic approach. *Mol Biochem Parasitol,* 187**,** 91-7.

MOREIRA, E. D., JR., NASSRI, V. B., SANTOS, R. S., MATOS, J. F., DE CARVALHO, W. A., SILVANI, C. S. & SANTANA E SANT'ANA, C. 2005. Association of Helicobacter pylori infection and giardiasis: results from a study of surrogate markers for fecal exposure among children. *World J Gastroenterol,* 11**,** 2759-63.

MORGAN, D. O. 1997. Cyclin-dependent kinases: engines, clocks, and microprocessors. *Annu Rev Cell Dev Biol,* 13**,** 261-91.

MORKEN, M. H., VALEUR, J., NORIN, E., MIDTVEDT, T., NYSAETER, G. & BERSTAD, A. 2009. Antibiotic or bacterial therapy in post-giardiasis irritable bowel syndrome. *Scand J Gastroenterol,* 44**,** 1296-303.

MORRISON, H. G., MCARTHUR, A. G., GILLIN, F. D., ALEY, S. B., ADAM, R. D., OLSEN, G. J., BEST, A. A., CANDE, W. Z., CHEN, F., CIPRIANO, M. J., DAVIDS, B. J., DAWSON, S. C., ELMENDORF, H. G., HEHL, A. B., HOLDER, M. E., HUSE, S. M., KIM, U. U., LASEK-NESSELQUIST, E., MANNING, G., NIGAM, A., NIXON, J. E., PALM, D., PASSAMANECK, N. E., PRABHU, A., REICH, C. I., REINER, D. S., SAMUELSON, J., SVARD, S. G. & SOGIN, M. L. 2007. Genomic minimalism in the early diverging intestinal parasite Giardia lamblia. *Science,* 317**,** 1921-6.

MOWATT, M. R., LUJAN, H. D., COTTEN, D. B., BOWERS, B., YEE, J., NASH, T. E. & STIBBS, H. H. 1995. Developmentally regulated expression of a Giardia lamblia cyst wall protein gene. *Mol Microbiol,* 15**,** 955-63.

MPHB 2013. Manitoba Public Health Branch - Communicable disease management protocol: Giardiasis.

MULLER, J., WASTLING, J., SANDERSON, S., MULLER, N. & HEMPHILL, A. 2007. A novel Giardia lamblia nitroreductase, GlNR1, interacts with nitazoxanide and other thiazolides. *Antimicrob Agents Chemother,* 51**,** 1979-86.

MUREK, M., KOPIC, S. & GEIBEL, J. 2010. Evidence for intestinal chloride secretion. *Exp Physiol,* 95**,** 471-8.

NAGESHAN, R. K., ROY, N., RANADE, S. & TATU, U. 2014. Trans-spliced Heat Shock Protein 90 Modulates Encystation in Giardia lamblia. *PLoS Negl Trop Dis,* 8**,** e2829.

NANDAN, D. & REINER, N. E. 2005. Leishmania donovani engages in regulatory interference by targeting macrophage protein tyrosine phosphatase SHP-1. *Clin Immunol,* 114**,** 266-77.

NARAI, A., ARAI, S. & SHIMIZU, M. 1997. Rapid decrease in transepithelial electrical resistance of human intestinal Caco-2 cell monolayers by cytotoxic membrane perturbents. *Toxicol In Vitro,* 11**,** 347-54.

NASH, T. E. 2002. Surface antigenic variation in Giardia lamblia. *Mol Microbiol,* 45**,** 585-90.

NASH, T. E. 2013. Unraveling how Giardia infections cause disease. *J Clin Invest,* 123**,** 2346-7.

NATHNAC 2013. National Travel Health Network and Centre - Giardiasis.

NHS. *Giardiasis - Prevention* [Online]. http://www.nhs.uk/Conditions/Giardiasis/Pages/Prevention.aspx. Available: http://www.nhs.uk/Conditions/Giardiasis/Pages/Prevention.aspx.

NICHOLS, G., LANE, C., ASGARI, N., VERLANDER, N. Q. & CHARLETT, A. 2009. Rainfall and outbreaks of drinking water related disease and in England and Wales. *J Water Health,* 7**,** 1-8.

NILLIUS, D., MULLER, J. & MULLER, N. 2011. Nitroreductase (GlNR1) increases susceptibility of Giardia lamblia and Escherichia coli to nitro drugs. *J Antimicrob Chemother,* 66**,** 1029-35.

NKENFOU, C. N., NANA, C. T. & PAYNE, V. K. 2013. Intestinal parasitic infections in HIV infected and non-infected patients in a low HIV prevalence region, West-Cameroon. *PLoS One,* 8**,** e57914.

NOOR AZIAN, M. Y., SAN, Y. M., GAN, C. C., YUSRI, M. Y., NURULSYAMZAWATY, Y., ZUHAIZAM, A. H., MASLAWATY, M. N., NORPARINA, I. & VYTHILINGAM, I. 2007. Prevalence of intestinal protozoa in an aborigine community in Pahang, Malaysia. *Trop Biomed,* 24**,** 55-62.

NZMH. 2013. *New Zealand Ministry of Health - Giardia* [Online]. https://www.health.govt.nz/your-health/conditions-and-treatments/diseases-and-illnesses/food-and-water-borne-diseases/giardia. Available: https://www.health.govt.nz/your-health/conditions-and-treatments/diseases-and-illnesses/food-and-water-borne-diseases/giardia.

OKE, T. T., MOSKOVITZ, J. & WILLIAMS, D. L. 2009. Characterization of the methionine sulfoxide reductases of Schistosoma mansoni. *J Parasitol,* 95**,** 1421-8.

OKSENHENDLER, E., GERARD, L., FIESCHI, C., MALPHETTES, M., MOUILLOT, G., JAUSSAUD, R., VIALLARD, J. F., GARDEMBAS, M., GALICIER, L., SCHLEINITZ, N., SUAREZ, F., SOULAS-SPRAUEL, P., HACHULLA, E., JACCARD, A., GARDEUR, A., THEODOROU, I., RABIAN, C. & DEBRE, P. 2008. Infections in 252 patients with common variable immunodeficiency. *Clin Infect Dis,* 46**,** 1547-54.

OLDSTONE, M. B. 1998. Molecular mimicry and immune-mediated diseases. *FASEB J,* 12**,** 1255-65.

ORTEGA, Y. R. & ADAM, R. D. 1997. Giardia: overview and update. *Clin Infect Dis,* 25**,** 545-9; quiz 550.

OSTERHOLM, M. T., FORFANG, J. C., RISTINEN, T. L., DEAN, A. G., WASHBURN, J. W., GODES, J. R., RUDE, R. A. & MCCULLOUGH, J. G. 1981. An outbreak of foodborne giardiasis. *New England Journal of Medicine*

304**,** 24–28.

PAGET, T. A. & JAMES, S. L. 1994. The mucolytic activity of polyamines and mucosal invasion. *Biochem Soc Trans,* 22**,** 394S.

PAGET, T. A., JARROLL, E. L., MANNING, P., LINDMARK, D. G. & LLOYD, D. 1989. Respiration in the cysts and trophozoites of Giardia muris. *J Gen Microbiol,* 135**,** 145-54.

PALM, D., WEILAND, M., MCARTHUR, A. G., WINIECKA-KRUSNELL, J., CIPRIANO, M. J., BIRKELAND, S. R., PACOCHA, S. E., DAVIDS, B., GILLIN, F., LINDER, E. & SVARD, S. 2005. Developmental changes in the adhesive disk during Giardia differentiation. *Mol Biochem Parasitol,* 141**,** 199-207.

PALM, J. E., WEILAND, M. E., GRIFFITHS, W. J., LJUNGSTROM, I. & SVARD, S. G. 2003. Identification of immunoreactive proteins during acute human giardiasis. *J Infect Dis,* 187**,** 1849-59.

PARLATI, F., VARLAMOV, O., PAZ, K., MCNEW, J. A., HURTADO, D., SOLLNER, T. H. & ROTHMAN, J. E. 2002. Distinct SNARE complexes mediating membrane fusion in Golgi transport based on combinatorial specificity. *Proc Natl Acad Sci U S A,* 99**,** 5424-9.

PEATTIE, D. A. 1990. The giardins of Giardia lamblia: genes and proteins with promise. *Parasitol Today,* 6**,** 52-6.

PEATTIE, D. A., ALONSO, R. A., HEIN, A. & CAULFIELD, J. P. 1989. Ultrastructural localization of giardins to the edges of disk microribbons of Giardia lamblia and the nucleotide and deduced protein sequence of alpha giardin. *J Cell Biol,* 109**,** 2323-35.

PELAYO, L., FRAGA, J., NUNEZ, F. A., MENDOZA, D., TORRES, D. R. & FINLAY, C. M. 2003. Genetic characterization by random amplified polymorphic DNA analysis (RAPD) of 18 isolates of Giardia lamblia obtained from day care children. *Exp Parasitol,* 104**,** 162-6.

PEREZ, T., ALBORNOZ, J. & DOMINGUEZ, A. 1998. An evaluation of RAPD fragment reproducibility and nature. *Mol Ecol,* 7**,** 1347-57.

PERRY, D. A., MORRISON, H. G. & ADAM, R. D. 2011. Optical map of the genotype A1 WB C6 Giardia lamblia genome isolate. *Mol Biochem Parasitol,* 180**,** 112-4.

PFEFFER, S. R. 2001. Rab GTPases: specifying and deciphering organelle identity and function. *Trends Cell Biol,* 11**,** 487-91.

PHAC. *Public Health Agency Canada - GIARDIASIS* [Online]. http://dsol-smed.phac-aspc.gc.ca/dsol-smed/ndis/diseases/giar-eng.php.

PHE 2013. Epidemiological data: Gastrointestinal diseases from 2000 to 2012.

.

PHILIP, R., CAMPBELL, E. & WHEATLEY, D. N. 2003. Arginine deprivation, growth inhibition and tumour cell death: 2. Enzymatic degradation of arginine in normal and malignant cell cultures. *Br J Cancer,* 88**,** 613-23.

PILLAI, D. R. & KAIN, K. C. 1999. Immunochromatographic strip-based detection of Entamoeba histolytica-E. dispar and Giardia lamblia coproantigen. *J Clin Microbiol,* 37**,** 3017-9.

PINTO, M., ROBINE-LEON, S., APPAY, M., KEDINGER, M., TRIADOU, N., DUSSAULX, E., LACROIX, B., SIMON-ASSMANN, P., HAFFEN, K., FOGH, J. & ZWEIBAUM, A. 1983. Enterocyte-like Differenciation and Polarization of the Human Colon Carcinoma Cell Line Caco-2 in culture. *Biol Cell,* 47**,** 323-330.

PLUTZER, J., ONGERTH, J. & KARANIS, P. 2010. Giardia taxonomy, phylogeny and epidemiology: Facts and open questions. *Int J Hyg Environ Health,* 213**,** 321-33.

PORTER, J. D., GAFFNEY, C., HEYMANN, D. & PARKIN, W. 1990. Food-borne outbreak of Giardia lamblia. *Am J Public Health,* 80**,** 1259-60.

POXLEITNER, M. K., CARPENTER, M. L., MANCUSO, J. J., WANG, C. J., DAWSON, S. C. & CANDE, W. Z. 2008. Evidence for karyogamy and exchange of genetic material in the binucleate intestinal parasite Giardia intestinalis. *Science,* 319**,** 1530-3.

PRADO, M. S., STRINA, A., BARRETO, M. L., OLIVEIRA-ASSIS, A. M., PAZ, L. M. & CAIRNCROSS, S. 2003. Risk factors for infection with Giardia duodenalis in pre-school children in the city of Salvador, Brazil. *Epidemiol Infect,* 131**,** 899-906.

PRICE, E. R., ZYDOWSKY, L. D., JIN, M. J., BAKER, C. H., MCKEON, F. D. & WALSH, C. T. 1991. Human cyclophilin B: a second cyclophilin gene encodes a peptidyl-prolyl isomerase with a signal sequence. *Proc Natl Acad Sci U S A,* 88**,** 1903-7.

PRUCCA, C. G., RIVERO, F. D. & LUJAN, H. D. 2011. Regulation of antigenic variation in Giardia lamblia. *Annu Rev Microbiol,* 65**,** 611-30.

PRUCCA, C. G., SLAVIN, I., QUIROGA, R., ELIAS, E. V., RIVERO, F. D., SAURA, A., CARRANZA, P. G. & LUJAN, H. D. 2008. Antigenic variation in Giardia lamblia is regulated by RNA interference. *Nature,* 456**,** 750-4.

RAMESH, M. A., MALIK, S. B. & LOGSDON, J. M., JR. 2005. A phylogenomic inventory of meiotic genes; evidence for sex in Giardia and an early eukaryotic origin of meiosis. *Curr Biol,* 15**,** 185-91.

READ, C., WALTERS, J., ROBERTSON, I. D. & THOMPSON, R. C. 2002. Correlation between genotype of Giardia duodenalis and diarrhoea. *Int J Parasitol,* 32**,** 229-31.

REINER, D. S., DOUGLAS, H. & GILLIN, F. D. 1989. Identification and localization of cyst-specific antigens of Giardia lamblia. *Infect Immun,* 57**,** 963-8.

REINER, D. S., HETSKO, M. L., MESZAROS, J. G., SUN, C. H., MORRISON, H. G., BRUNTON, L. L. & GILLIN, F. D. 2003. Calcium signaling in excystation of the early diverging eukaryote, Giardia lamblia. *J Biol Chem,* 278**,** 2533-40.

REINER, D. S., MCCAFFERY, J. M. & GILLIN, F. D. 2001. Reversible interruption of Giardia lamblia cyst wall protein transport in a novel regulated secretory pathway. *Cell Microbiol,* 3**,** 459-72.

RENDTORFF, R. C. 1954. The experimental transmission of human intestinal protozoan parasites. II. Giardia lamblia cysts given in capsules. *Am J Hyg,* 59**,** 209-20.

REQUEJO, R., HURD, T. R., COSTA, N. J. & MURPHY, M. P. 2010. Cysteine residues exposed on protein surfaces are the dominant intramitochondrial thiol and may protect against oxidative damage. *FEBS J,* 277**,** 1465-80.

RICHARD, S. A., BLACK, R. E., GILMAN, R. H., GUERRANT, R. L., KANG, G., LANATA, C. F., MOLBAK, K., RASMUSSEN, Z. A., SACK, R. B., VALENTINER-BRANTH, P., CHECKLEY, W., CHILDHOOD, M. & INFECTION, N. 2013. Diarrhea in early childhood: short-term association with weight and long-term association with length. *Am J Epidemiol,* 178**,** 1129-38.

RIMHANEN-FINNE, R., ENEMARK, H. L., KOLEHMAINEN, J., TOROPAINEN, P. & HANNINEN, M. L. 2007. Evaluation of immunofluorescence microscopy and enzyme-linked immunosorbent assay in detection of Cryptosporidium and Giardia infections in asymptomatic dogs. *Vet Parasitol,* 145**,** 345-8.

RINGQVIST, E., AVESSON, L., SODERBOM, F. & SVARD, S. G. 2011. Transcriptional changes in Giardia during host-parasite interactions. *Int J Parasitol,* 41**,** 277-85.

RINGQVIST, E., PALM, J. E., SKARIN, H., HEHL, A. B., WEILAND, M., DAVIDS, B. J., REINER, D. S., GRIFFITHS, W. J., ECKMANN, L., GILLIN, F. D. & SVARD, S. G. 2008. Release of metabolic enzymes by Giardia in response to interaction with intestinal epithelial cells. *Mol Biochem Parasitol,* 159**,** 85-91.

ROBERTSON, L. J., FORBERG, T., HERMANSEN, L., HAMNES, I. S. & GJERDE, B. 2007. Giardia duodenalis cysts isolated from wild moose and reindeer in Norway: genetic characterization by PCR-rflp and sequence analysis at two genes. *J Wildl Dis,* 43**,** 576-85.

ROBERTSON, L. J. & GJERDE, B. K. 2006. Fate of Cryptosporidium oocysts and Giardia cysts in the Norwegian aquatic environment over winter. *Microb Ecol,* 52**,** 597-602.

ROBERTSON, L. J., HANEVIK, K., ESCOBEDO, A. A., MORCH, K. & LANGELAND, N. 2010. Giardiasis--why do the symptoms sometimes never stop? *Trends Parasitol,* 26**,** 75-82.

RODRIGUEZ-FUENTES, G. B., CEDILLO-RIVERA, R., FONSECA-LINAN, R., ARGUELLO-GARCIA, R., MUNOZ, O., ORTEGA-PIERRES, G. & YEPEZ-MULIA, L. 2006. Giardia duodenalis: analysis of secreted proteases upon trophozoite-epitheial cell interaction in vitro. *Mem Inst Oswaldo Cruz,* 101**,** 693-6.

ROSE, Z. B., GROVE, D. S. & SEAL, S. N. 1986. Mechanism of activation by anions of phosphoglycolate phosphatases from spinach and human red blood cells. *J Biol Chem,* 261**,** 10996-1002.

ROSSIGNOL, J. F. 2010. Cryptosporidium and Giardia: treatment options and prospects for new drugs. *Exp Parasitol,* 124**,** 45-53.

ROXSTROM-LINDQUIST, K., PALM, D., REINER, D., RINGQVIST, E. & SVARD, S. G. 2006. Giardia immunity--an update. *Trends Parasitol,* 22**,** 26-31.

ROXSTROM-LINDQUIST, K., RINGQVIST, E., PALM, D. & SVARD, S. 2005. Giardia lamblia-induced changes in gene expression in differentiated Caco-2 human intestinal epithelial cells. *Infect Immun,* 73**,** 8204-8.

ROY, N., NAGESHAN, R. K., RANADE, S. & TATU, U. 2012. Heat shock protein 90 from neglected protozoan parasites. *Biochim Biophys Acta,* 1823**,** 707-11.

SAJID, M. & MCKERROW, J. H. 2002. Cysteine proteases of parasitic organisms. *Mol Biochem Parasitol,* 120**,** 1-21.

SANDERS, N. N., EIJSINK, V. G., VAN DEN PANGAART, P. S., JOOST VAN NEERVEN, R. J., SIMONS, P. J., DE SMEDT, S. C. & DEMEESTER, J. 2007. Mucolytic activity of bacterial and human chitinases. *Biochim Biophys Acta,* 1770**,** 839-46.

SARAIYA, A. A. & WANG, C. C. 2008. snoRNA, a novel precursor of microRNA in Giardia lamblia. *PLoS Pathog,* 4**,** e1000224.

SAUCH, J. F. 1985. Use of immunofluorescence and phase-contrast microscopy for detection and identification of Giardia cysts in water samples. *Appl Environ Microbiol,* 50**,** 1434-8.

SAVAGE, D. C. 1977. Microbial ecology of the gastrointestinal tract. *Annu Rev Microbiol,* 31**,** 107-33.

SCHARF, R. J., DEBOER, M. D. & GUERRANT, R. L. 2014. Recent advances in understanding the long-term sequelae of childhood infectious diarrhea. *Curr Infect Dis Rep,* 16**,** 408.

SCHUURMAN, T., LANKAMP, P., VAN BELKUM, A., KOOISTRA-SMID, M. & VAN ZWET, A. 2007. Comparison of microscopy, real-time PCR and a rapid immunoassay for the detection of Giardia lamblia in human stool specimens. *Clin Microbiol Infect,* 13**,** 1186-91.

SCHWANHAUSSER, B., BUSSE, D., LI, N., DITTMAR, G., SCHUCHHARDT, J., WOLF, J., CHEN, W. & SELBACH, M. 2011. Global quantification of mammalian gene expression control. *Nature,* 473**,** 337-42.

SCORZA, A. V., BALLWEBER, L. R., TANGTRONGSUP, S., PANUSKA, C. & LAPPIN, M. R. 2012. Comparisons of mammalian Giardia duodenalis assemblages based on the beta-giardin, glutamate dehydrogenase and triose phosphate isomerase genes. *Vet Parasitol,* 189**,** 182-8.

SCOTT, C. C., VACCA, F. & GRUENBERG, J. 2014. Endosome maturation, transport and functions. *Semin Cell Dev Biol*.

SCOTT, K. G., LOGAN, M. R., KLAMMER, G. M., TEOH, D. A. & BURET, A. G. 2000. Jejunal brush border microvillous alterations in Giardia muris-infected mice: role of T lymphocytes and interleukin-6. *Infect Immun,* 68**,** 3412-8.

SCOTT, K. G., MEDDINGS, J. B., KIRK, D. R., LEES-MILLER, S. P. & BURET, A. G. 2002. Intestinal infection with Giardia spp. reduces epithelial barrier function in a myosin light chain kinase-dependent fashion. *Gastroenterology,* 123**,** 1179-90.

SCOTT, K. G., YU, L. C. & BURET, A. G. 2004. Role of CD8+ and CD4+ T lymphocytes in jejunal mucosal injury during murine giardiasis. *Infect Immun,* 72**,** 3536-42.

SEPER, A., FENGLER, V. H., ROIER, S., WOLINSKI, H., KOHLWEIN, S. D., BISHOP, A. L., CAMILLI, A., REIDL, J. & SCHILD, S. 2011. Extracellular nucleases and extracellular DNA play important roles in Vibrio cholerae biofilm formation. *Mol Microbiol,* 82**,** 1015-37.

SHARPE, H. J., STEVENS, T. J. & MUNRO, S. 2010. A comprehensive comparison of transmembrane domains reveals organelle-specific properties. *Cell,* 142**,** 158-69.

SHETTY, N. & PRABHU, T. 1988. Evaluation of faecal preservation and staining methods in the diagnosis of acute amoebiasis and giardiasis. *J Clin Pathol,* 41**,** 694-9.

SHIELDS, J. M., GLEIM, E. R. & BEACH, M. J. 2008. Prevalence of Cryptosporidium spp. and Giardia intestinalis in swimming pools, Atlanta, Georgia. *Emerg Infect Dis,* 14**,** 948-50.

SHIVELY, J. M., VAN KEULEN, G. & MEIJER, W. G. 1998. Something from almost nothing: carbon dioxide fixation in chemoautotrophs. *Annu Rev Microbiol,* 52**,** 191-230.

SIMON, C. E. 1922. A critique of the supposed rodent origin of human giardiasis.

SIMPSON, A. G., INAGAKI, Y. & ROGER, A. J. 2006. Comprehensive multigene phylogenies of excavate protists reveal the evolutionary positions of "primitive" eukaryotes. *Mol Biol Evol,* 23**,** 615-25.

SINGER, S. M., ELMENDORF, H. G., CONRAD, J. T. & NASH, T. E. 2001. Biological selection of variant-specific surface proteins in Giardia lamblia. *J Infect Dis,* 183**,** 119-24.

SKARIN, H., RINGQVIST, E., HELLMAN, U. & SVARD, S. G. 2011. Elongation factor 1-alpha is released into the culture medium during growth of Giardia intestinalis trophozoites. *Exp Parasitol,* 127**,** 804-10.

SLAVIN, I., SAURA, A., CARRANZA, P. G., TOUZ, M. C., NORES, M. J. & LUJAN, H. D. 2002. Dephosphorylation of cyst wall proteins by a secreted lysosomal acid phosphatase is essential for excystation of Giardia lamblia. *Mol Biochem Parasitol,* 122**,** 95-8.

SLEE, E. A., HARTE, M. T., KLUCK, R. M., WOLF, B. B., CASIANO, C. A., NEWMEYER, D. D., WANG, H. G., REED, J. C., NICHOLSON, D. W., ALNEMRI, E. S., GREEN, D. R. & MARTIN, S. J. 1999. Ordering the cytochrome c-initiated caspase cascade: hierarchical activation of caspases-2, -3, -6, -7, -8, and -10 in a caspase-9-dependent manner. *J Cell Biol,* 144**,** 281-92.

SMITH, H. V., CACCIO, S. M., COOK, N., NICHOLS, R. A. & TAIT, A. 2007. Cryptosporidium and Giardia as foodborne zoonoses. *Vet Parasitol,* 149**,** 29-40.

SMITH, P. D., GILLIN, F. D., SPIRA, W. M. & NASH, T. E. 1982. Chronic giardiasis: studies on drug sensitivity, toxin production, and host immune response. *Gastroenterology,* 83**,** 797-803.

SNOECK, V., GODDEERIS, B. & COX, E. 2005. The role of enterocytes in the intestinal barrier function and antigen uptake. *Microbes Infect,* 7**,** 997-1004.

SOGAARD, M., TANI, K., YE, R. R., GEROMANOS, S., TEMPST, P., KIRCHHAUSEN, T., ROTHMAN, J. E. & SOLLNER, T. 1994. A rab protein is required for the assembly of SNARE complexes in the docking of transport vesicles. *Cell,* 78**,** 937-48.

SOLAYMANI-MOHAMMADI, S. & SINGER, S. M. 2011. Host immunity and pathogen strain contribute to intestinal disaccharidase impairment following gut infection. *Journal of immunology,* 187**,** 3769-75.

SOLTYS, B. J., FALAH, M. & GUPTA, R. S. 1996. Identification of endoplasmic reticulum in the primitive eukaryote Giardia lamblia using cryoelectron microscopy and antibody to Bip. *J Cell Sci,* 109 (Pt 7)**,** 1909-17.

SPRONG, H., CACCIO, S. M. & VAN DER GIESSEN, J. W. 2009. Identification of zoonotic genotypes of Giardia duodenalis. *PLoS Negl Trop Dis,* 3**,** e558.

STARK, D., AL-QASSAB, S. E., BARRATT, J. L., STANLEY, K., ROBERTS, T., MARRIOTT, D., HARKNESS, J. & ELLIS, J. T. 2011. Evaluation of multiplex tandem real-time PCR for detection of Cryptosporidium spp., Dientamoeba fragilis, Entamoeba histolytica, and Giardia intestinalis in clinical stool samples. *J Clin Microbiol,* 49**,** 257-62.

STEUART, R. F. 2010. Proteomic analysis of Giardia: studies from the pre- and post-genomic era. *Experimental parasitology,* 124**,** 26-30.

STEUART, R. F., O'HANDLEY, R., LIPSCOMBE, R. J., LOCK, R. A. & THOMPSON, R. C. 2008. Alpha 2 giardin is an assemblage A-specific protein of human infective Giardia duodenalis. *Parasitology,* 135**,** 1621-7.

STRAND, E. A., ROBERTSON, L. J., HANEVIK, K., ALVSVAG, J. O., MORCH, K. & LANGELAND, N. 2008. Sensitivity of a Giardia antigen test in persistent giardiasis following an extensive outbreak. *Clin Microbiol Infect,* 14**,** 1069-71.

SULEMANA, A., PAGET, T. A. & JARROLL, E. L. 2014. Commitment to cyst formation in Giardia. *Microbiology,* 160**,** 330-9.

SULLIVAN, P. B., MARSH, M. N., PHILLIPS, M. B., DEWIT, O., NEALE, G., CEVALLOS, A. M., YAMSON, P. & FARTHING, M. J. 1991. Prevalence and treatment of giardiasis in chronic diarrhoea and malnutrition. *Arch Dis Child,* 66**,** 304-6.

SUN, C. H., MCCAFFERY, J. M., REINER, D. S. & GILLIN, F. D. 2003. Mining the Giardia lamblia genome for new cyst wall proteins. *J Biol Chem,* 278**,** 21701-8.

SUNDSTROM, P. & ALIAGA, G. R. 1994. A subset of proteins found in culture supernatants of Candida albicans includes the abundant, immunodominant, glycolytic enzyme enolase. *J Infect Dis,* 169**,** 452-6.

TAIPALE, M., JAROSZ, D. F. & LINDQUIST, S. 2010. HSP90 at the hub of protein homeostasis: emerging mechanistic insights. *Nat Rev Mol Cell Biol,* 11**,** 515-28.

TAIT, S. W. & GREEN, D. R. 2010. Mitochondria and cell death: outer membrane permeabilization and beyond. *Nat Rev Mol Cell Biol,* 11**,** 621-32.

TAKATA, K., MATSUZAKI, T. & TAJIKA, Y. 2004. Aquaporins: water channel proteins of the cell membrane. *Prog Histochem Cytochem,* 39**,** 1-83.

TAN, S., TOMPKINS, L. S. & AMIEVA, M. R. 2009. Helicobacter pylori usurps cell polarity to turn the cell surface into a replicative niche. *PLoS Pathog,* 5**,** e1000407.

TANIUCHI, M., VERWEIJ, J. J., NOOR, Z., SOBUZ, S. U., LIESHOUT, L., PETRI, W. A., JR., HAQUE, R. & HOUPT, E. R. 2011. High throughput multiplex PCR and probe-based detection with Luminex beads for seven intestinal parasites. *Am J Trop Med Hyg,* 84**,** 332-7.

TERESA PELLICER, M., FELISA NUNEZ, M., AGUILAR, J., BADIA, J. & BALDOMA, L. 2003. Role of 2-phosphoglycolate phosphatase of Escherichia coli in metabolism of the 2-phosphoglycolate formed in DNA repair. *J Bacteriol,* 185**,** 5815-21.

THOMPSON, R. C. 2002. Presidential address: rediscovering parasites using molecular tools--towards revising the taxonomy of Echinococcus, Giardia and Cryptosporidium. *Int J Parasitol,* 32**,** 493-6.

THOMPSON, R. C., REYNOLDSON, J. A. & MENDIS, A. H. 1993. Giardia and giardiasis. *Adv Parasitol,* 32**,** 71-160.

TIBAYRENC, M., KJELLBERG, F., ARNAUD, J., OURY, B., BRENIERE, S. F., DARDE, M. L. & AYALA, F. J. 1991. Are eukaryotic microorganisms clonal or sexual? A population genetics vantage. *Proc Natl Acad Sci U S A,* 88**,** 5129-33.

TOUZ, M. C., CONRAD, J. T. & NASH, T. E. 2005. A novel palmitoyl acyl transferase controls surface protein palmitoylation and cytotoxicity in Giardia lamblia. *Mol Microbiol,* 58**,** 999-1011.

TOUZ, M. C., LUJAN, H. D., HAYES, S. F. & NASH, T. E. 2003. Sorting of encystation-specific cysteine protease to lysosome-like peripheral vacuoles in Giardia lamblia requires a conserved tyrosine-based motif. *J Biol Chem,* 278**,** 6420-6.

TOUZ, M. C., NORES, M. J., SLAVIN, I., CARMONA, C., CONRAD, J. T., MOWATT, M. R., NASH, T. E., CORONEL, C. E. & LUJAN, H. D. 2002. The activity of a developmentally regulated cysteine proteinase is required for cyst wall formation in the primitive eukaryote Giardia lamblia. *J Biol Chem,* 277**,** 8474-81.

TOUZ, M. C., ROPOLO, A. S., RIVERO, M. R., VRANYCH, C. V., CONRAD, J. T., SVARD, S. G. & NASH, T. E. 2008. Arginine deiminase has multiple regulatory roles in the biology of Giardia lamblia. *J Cell Sci,* 121**,** 2930-8.

TOVAR, J., LEON-AVILA, G., SANCHEZ, L. B., SUTAK, R., TACHEZY, J., VAN DER GIEZEN, M., HERNANDEZ, M., MULLER, M. & LUCOCQ, J. M. 2003. Mitochondrial remnant organelles of Giardia function in iron-sulphur protein maturation. *Nature,* 426**,** 172-6.

TRAUB, R. J., INPANKAEW, T., REID, S. A., SUTTHIKORNCHAI, C., SUKTHANA, Y., ROBERTSON, I. D. & THOMPSON, R. C. 2009. Transmission cycles of Giardia duodenalis in dogs and humans in Temple communities in Bangkok--a critical evaluation of its prevalence using three diagnostic tests in the field in the absence of a gold standard. *Acta Trop,* 111**,** 125-32.

TRAUB, R. J., MONIS, P. T., ROBERTSON, I., IRWIN, P., MENCKE, N. & THOMPSON, R. C. 2004. Epidemiological and molecular evidence supports the zoonotic transmission of Giardia among humans and dogs living in the same community. *Parasitology,* 128**,** 253-62.

TROEGER, H., EPPLE, H. J., SCHNEIDER, T., WAHNSCHAFFE, U., ULLRICH, R., BURCHARD, G. D., JELINEK, T., ZEITZ, M., FROMM, M. & SCHULZKE, J. D. 2007. Effect of chronic Giardia lamblia infection on epithelial transport and barrier function in human duodenum. *Gut,* 56**,** 328-35.

TURK, V., TURK, B., GUNCAR, G., TURK, D. & KOS, J. 2002. Lysosomal cathepsins: structure, role in antigen processing and presentation, and cancer. *Adv Enzyme Regul,* 42**,** 285-303.

UMEDA, K., MATSUI, T., NAKAYAMA, M., FURUSE, K., SASAKI, H., FURUSE, M. & TSUKITA, S. 2004. Establishment and characterization of cultured epithelial cells lacking expression of ZO-1. *J Biol Chem,* 279**,** 44785-94.

UPCROFT, J. & UPCROFT, P. 1998. My favorite cell: Giardia. *Bioessays,* 20**,** 256-63.

UPCROFT, J. A., CAMPBELL, R. W., BENAKLI, K., UPCROFT, P. & VANELLE, P. 1999. Efficacy of new 5-nitroimidazoles against metronidazole-susceptible and -resistant Giardia, Trichomonas, and Entamoeba spp. *Antimicrob Agents Chemother,* 43**,** 73-6.

UPCROFT, P., MITCHELL, R. & BOREHAM, P. F. 1990. DNA fingerprinting of the intestinal parasite Giardia duodenalis with the M13 phage genome. *Int J Parasitol,* 20**,** 319-23.

UPCROFT, P. & UPCROFT, J. A. 2001. Drug targets and mechanisms of resistance in the anaerobic protozoa. *Clin Microbiol Rev,* 14**,** 150-64.

VAN DER GIESSEN, J. W., DE VRIES, A., ROOS, M., WIELINGA, P., KORTBEEK, L. M. & MANK, T. G. 2006. Genotyping of Giardia in Dutch patients and animals: a phylogenetic analysis of human and animal isolates. *Int J Parasitol,* 36**,** 849-58.

VARDAR, D., NORTH, C. L., SANCHEZ-IRIZARRY, C., ASTER, J. C. & BLACKLOW, S. C. 2003. Nuclear magnetic resonance structure of a prototype Lin12-Notch repeat module from human Notch1. *Biochemistry,* 42**,** 7061-7.

VENKATESAN, P. 1998. Albendazole. *J Antimicrob Chemother,* 41**,** 145-7.

VILLENEUVE, A. M. & HILLERS, K. J. 2001. Whence meiosis? *Cell,* 106**,** 647-50.

WALLIS, P. M., ERLANDSEN, S. L., ISAAC-RENTON, J. L., OLSON, M. E., ROBERTSON, W. J. & VAN KEULEN, H. 1996. Prevalence of Giardia cysts and Cryptosporidium oocysts and characterization of Giardia spp. isolated from drinking water in Canada. *Appl Environ Microbiol,* 62**,** 2789-97.

WALTERS, R. J., O'BRIEN, J. A., VALVERDE, M. A. & SEPULVEDA, F. V. 1992. Membrane conductance and cell volume changes evoked by vasoactive intestinal polypeptide and carbachol in small intestinal crypts. *Pflugers Arch,* 421**,** 598-605.

WAMPFLER, P. B., TOSEVSKI, V., NANNI, P., SPYCHER, C. & HEHL, A. B. 2014. Proteomics of Secretory and Endocytic Organelles in Giardia lamblia. *PLoS One,* 9**,** e94089.

WARD, W., ALVARADO, L., RAWLINGS, N. D., ENGEL, J. C., FRANKLIN, C. & MCKERROW, J. H. 1997. A primitive enzyme for a primitive cell: the protease required for excystation of Giardia. *Cell,* 89**,** 437-44.

WASTLING, J. M., XIA, D., SOHAL, A., CHAUSSEPIED, M., PAIN, A. & LANGSLEY, G. 2009. Proteomes and transcriptomes of the Apicomplexa--where's the message? *International journal for parasitology,* 39**,** 135-43.

WEBSTER, B. H. 1960. Furazolidone in the treatment of giardiasis. *Am J Dig Dis,* 5**,** 618-22.

WEGAYEHU, T., ADAMU, H. & PETROS, B. 2013. Prevalence of Giardia duodenalis and Cryptosporidium species infections among children and cattle in North Shewa Zone, Ethiopia. *BMC Infect Dis,* 13**,** 419.

WEI, C. J., TIAN, X. F., ADAM, R. D. & LU, S. Q. 2010. Giardia lamblia: intracellular localization of alpha8-giardin. *Exp Parasitol,* 126**,** 489-96.

WEIBEL, E. K., HADVARY, P., HOCHULI, E., KUPFER, E. & LENGSFELD, H. 1987. Lipstatin, an inhibitor of pancreatic lipase, produced by Streptomyces toxytricini. I. Producing organism, fermentation, isolation and biological activity. *J Antibiot (Tokyo),* 40**,** 1081-5.

WEILAND, M. E., PALM, J. E., GRIFFITHS, W. J., MCCAFFERY, J. M. & SVARD, S. G. 2003. Characterisation of alpha-1 giardin: an immunodominant Giardia lamblia annexin with glycosaminoglycan-binding activity. *Int J Parasitol,* 33**,** 1341-51.

WELCH, T. P. 2000. Risk of giardiasis from consumption of wilderness water in North America: a systematic review of epidemiologic data. *Int J Infect Dis,* 4**,** 100-3.

WELSH, J. & MCCLELLAND, M. 1990. Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res,* 18**,** 7213-8.

WHO 1996. -The world health report. .

WHO 2008. – Guidelines for drinking-water quality (Third edition incorporating first and second addenda).

WHO. 2013. *International travel and health - Giardiasis* [Online]. http://www.who.int/ith/diseases/giardiasis/en/. Available: http://www.who.int/ith/diseases/giardiasis/en/.

WILKINSON, B. & GILBERT, H. F. 2004. Protein disulfide isomerase. *Biochim Biophys Acta,* 1699**,** 35-44.

WILLIAMS, J. G., KUBELIK, A. R., LIVAK, K. J., RAFALSKI, J. A. & TINGEY, S. V. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res,* 18**,** 6531-5.

WOLFE, M. S. 1992. Giardiasis. *Clin Microbiol Rev,* 5**,** 93-100.

WRIGHT, E. M., HIRAYAMA, B. A. & LOO, D. F. 2007. Active sugar transport in health and disease. *J Intern Med,* 261**,** 32-43.

YU, L. C., FLYNN, A. N., TURNER, J. R. & BURET, A. G. 2005. SGLT-1-mediated glucose uptake protects intestinal epithelial cells against LPS-induced apoptosis and barrier defects: a novel cellular rescue mechanism? *FASEB J,* 19**,** 1822-35.

YU, L. C., HUANG, C. Y., KUO, W. T., SAYER, H., TURNER, J. R. & BURET, A. G. 2008. SGLT-1-mediated glucose uptake protects human intestinal epithelial cells against Giardia duodenalis-induced apoptosis. *Int J Parasitol,* 38**,** 923-34.

YU, L. C., TURNER, J. R. & BURET, A. G. 2006. LPS/CD14 activation triggers SGLT-1-mediated glucose uptake and cell rescue in intestinal epithelial cells via early apoptotic signals upstream of caspase-3. *Exp Cell Res,* 312**,** 3276-86.

ZEUTHEN, T. 2010. Water-transporting proteins. *J Membr Biol,* 234**,** 57-73.

ZHANG, W., ZHANG, X., WANG, R., LIU, A., SHEN, Y., LING, H., CAO, J., YANG, F., ZHANG, X. & ZHANG, L. 2012. Genetic characterizations of Giardia duodenalis in sheep and goats in Heilongjiang Province, China and possibility of zoonotic transmission. *PLoS Negl Trop Dis,* 6**,** e1826.