DECLARATION

I declare that the content of this thesis entitled "**Postgenomics analyses of species-specific *Cryptosporidium* genes**" was undertaken and completed by myself, unless otherwise acknowledged and has not been submitted in support of an application for another degree or qualification in this or any other university or institution.

Maha Bouzid, MSc., B.Sc. (Hons)
University of East Anglia
April 2010

ACKNOWLEDGMENTS

Firstly, I would like to thank my supervisors Dr. Kevin Tyler and Professor Paul Hunter, to whom I am greatly indebted. Dr. Kevin Tyler has supported and encouraged me on a regular basis throughout the PhD and his enthusiastic approach to science has been highly motivating and inspiring. I am extremely grateful to Professor Paul Hunter, who gave me the opportunity to do this thesis, for believing in my abilities and allowing me to pursue my goal. I particularly appreciate his kindness, guidance and expertise, which facilitated my transition to *Cryptosporidium* research. I also have to thank Paul for all the European trips to attend the Healthy water project meetings, which was great fun and excellent opportunity for networking. I would like to thank Professor Andrew Johnston for his understanding, encouragement and support when things went wrong. I would also like to acknowledge many members of the Biomedical Research Centre past and present for their help and advice: Dr. Claire Butler, Professor Edmundo Grisard, Dr. Roberto Pierini, Miss Rebecca Roberts, Miss Devina Divekar, Mrs Patricia Lunness, Dr. Penny Powell, Mr. Matthew Jefferson, Dr. Guy Wheeler, Mr. Momodou Jobe, Dr. Suha Al-Naimi, Dr. Yongping Bao and Mr. Lawrence Barrera-Briceno. I would like to express my appreciation to Ms Gilly Potter and acknowledge her prompt and continuous administrative support provided to me and to all Faculty of Health PhD students. My thanks are extended to members of the UK *Cryptosporidium* Reference Unit for the donation of material and technical support mainly Dr. Rachel Chalmers, Dr. Kristin Elwin, Dr. Steve Hadfield, Dr. Guy Robinson and Mr. Brian Campbell.

I would like to express my gratitude to my beloved parents. Thanks Mum and Dad for your continuous love, support and prayers. You always believed that I can do it, this thesis is dedicated to you both. I would like to thank my sisters, my brother, my cousins, my aunties and extended family members for their support and understanding that a career in science is a life-long process. Thanks to my friends for still being there after I've ignored them for such a long time. Finally, thanks to my dear Basil for being loving, caring, supportive and for bravely offering to proofread this thesis.

ABSTRACT

Cryptosporidium is an apicomplexan parasite causing diarrhoeal illness. Two species *C. parvum* and *C. hominis* are of public health relevance. *C. parvum* and *C. hominis* genome sequences showed only 3-5% sequence divergence. Putatively species-specific genes were identified *in silico*. The specificity of these genes was tested experimentally by PCR in a collection of *Cryptosporidium* clinical isolates and reference strains. 90% of the genes tested were common to both species. PCR product sequence analysis detected 78 SNPs, 78.3% (61) of which were species-specific. 64.2% of the SNPs were synonymous. The sequences of these novel genetic loci allowed the construction of a robust and novel multi-locus analysis, which clearly discriminated with high bootstrap values *Cryptosporidium* species and subtypes. Evidence was found of one *C. parvum* (Cops-1) and one *C. hominis* (Chos-1) specific gene. The potential of Cops-1 as species determinant and virulence factor was considered. Cops-1 is positioned telomerically and annotated as encoding a 50 kDa protein which is secreted, serine rich and containing internal repeats. The predicted protein has interesting features (signal peptide, transmembrane domain and myristoylation motif), which are highly suggestive of a potential role in virulence and host-parasite interaction. Subsequent analysis showed that Cops-1 has a truncated ortholog in *C. hominis*. The N terminus is conserved, suggesting that the characteristic features are likely to be maintained, if the ortholog is expressed. Interestingly, sera from *C. parvum*-infected patients recognized a 50 kDa protein in *C. parvum* but not *C. hominis* antigen preparations, consistent with the protein being antigenic for patients. An anti-Cops-1 monoclonal antibody stained *C. parvum* oocyst contents and sporozoite surface consistent with a role in host-cell interaction. This is the first report of a *C. parvum* specific protein and has important implications for diagnosis and typing of this neglected pathogen.

ABBREVIATIONS

AIDS	Acquired Immunodeficiency Syndrome
ATCC	American Type Culture Collection
BFTE	Bovine Fallopian Tube Epithelial cell line
BLAST	Basic local alignment search tool
BSA	Bovine serum albumin
Caco-2	Human colorectal adenocarcinoma cell line
cDNA	Complementary DNA
Chos-1	*C. hominis* specific gene
Cops-1	*C. parvum* specific gene
COWP	*Cryptosporidium* oocyst wall protein
CRU	*Cryptosporidium* Reference Unit
CSL	Circumsporozoite-like
DAPI	4',6-diamidino-2- phenylindole
DHFR	Dihydrofolate reductase
DIC	Differential interference contrast
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
DOP-PCR	Degenerate oligonucleotide primed PCR
dsDNA	Double stranded DNA
EIA	Enzyme immunoassay
ELISA	Enzyme-linked immunosorbent assay
EST	Expressed sequence tag
FACS	Fluorescence activated cell sorting
FISH	Fluorescent in-situ hybridization
FITC	Fluorescein isothiocyanate
FMD	Foot and mouth disease
GP60	Glycoprotein 60
GPI	Glycophosphatidylinositol
GSS	Genome sequence survey
HAART	Highly active antiretroviral therapy
HCT-8	Human ileocecal colorectal adenocarcinoma cell line
His-tag	Polyhistidine-tag
HIV	Human Immunodeficiency Virus
HPA	Health Protection Agency
HPLC	High performance liquid chromatography
HRP	Horseradish peroxidase
HSP	Heat shock protein
ICZN	International Commission of the Zoological nomenclature
IFA	Immunofluorescence assay
IgG	Immunoglobulin G
IMS	Immunomagnetic separation
IPTG	Isopropyl-beta-D-thiogalactopyranoside
ITS	Internal spacer

KLH	Keyhole limpet hemocyanin
LB	Luria-Bertani medium
LAMP	Loop-mediated isothermal amplification procedure
MDA	Multiple displacement amplification
MDCK	Madin-Darby Canine Kidney cell line
MLA	Multi-locus analysis
mRNA	Messenger RNA
NER	Nucleotide Excision Repair
NJ	Neighbor-Joining
PBS	Phosphate buffered saline
PCR	Polymerase Chain Reaction
PI	Propidium iodide
PVDF	Polyvinylidene fluoride
RFLP	Restriction Fragment Length Polymorphism
rRNA	Ribosomal RNA
RT	Reverse transcriptase
SAAP	Single amino acid polymorphism
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SNP	Single nucleotide polymorphism
SSCP	Single Strand Conformation Polymorphism
SSU rRNA	Small subunit ribosomal RNA
TBST	Tris-Buffered Saline Tween
Trap	Thrombospondin related adhesive protein
VSG	Variant surface protein
WGA	Whole genome amplification

CONTENTS

List of Figures

List of Tables

CHAPTER 1:

General introduction

1.1 Historical background

Cryptosporidium is thought to have been first described by Clarke in 1895 as "swarm spores lying upon the gastric epithelium of mice" (Clarke, 1895). In retrospect, these small organisms were probably the motile merozoites of *Cryptosporidium muris*, the validated species named and described in 1907 by the American parasitologist Tyzzer (1907) from the gastric epithelium of laboratory mice (*Mus musculus*). These small coccidian organisms were placed in a new genus *Cryptosporidium* meaning hidden sporocysts because unlike the previously known coccidia, the oocyst of this parasite did not have sporocysts surrounding the sporozoites (Current and Garcia, 1991). In 1910, Tyzzer described in detail the various life cycle stages of *C. muris* (Tyzzer, 1910) and in 1912, he described a smaller size protozoan parasite than *C. muris*, from the intestine of mice (Tyzzer, 1912). Based on morphology, life cycle and experimental infection, he proposed a new *Cryptosporidium* species: *C. parvum*. In addition, Tyzzer also reported the first description of *Cryptosporidium* in birds (Tyzzer, 1929). He considered it as *C. parvum*; however, this was not confirmed.

Following Tyzzer reports of *C. muris* and *C. parvum* (Tyzzer 1907, 1919, 1912), several studies relying primarily on oocyst's structural features resulted in the naming of an additional 19 *Cryptosporidium* species, most of which are not considered valid now (Current and Garcia, 1991). Slavin reported *Cryptosporidium* from turkeys (*Meleagridis gallopavo*) and called the species *C. meleagridis* (Slavin, 1955). This study, in which Slavin describes severe diarrhoea and mortality of the infected birds, is the first to demonstrate the pathogenicity of *Cryptosporidium* species. Other *Cryptosporidium* species include *C. wrairi*, described from the small intestine of guinea pigs (*Cavia porcellus*) (Vetterling *et al.,* 1971), *C. felis* isolated from the domestic cat (*Felis catus*) (Iseki, 1979), *C. serpentis* from snakes (*Elaphe guttata*) (Levine, 1980, Brownstein *et al.,* 1977) and *C. baileyi* discovered in the small and large intestine, the sinuses, trachea and conjunctiva of chicken (*Gallus gallus*) (Current *et al.,* 1986).

Recently described *Cryptosporidium* species include *C. saurophilum* isolated from lizards (*Eumeces schneideri*) (Koudela and Modry, 1998), *C. galli* described from chicken (*Gallus gallus*) (Pavlasek, 1999), *C. andersoni* isolated from cattle (*Bos taurus*) (Lindsay *et al.,* 2000), *C. canis* described from domestic dogs (*Canis familiaris*) (Fayer *et al.,* 2001) and *C. molnari* from sea bream (*Sparus aurata*) (Alvarez-Pellitero and Sitjà-Bobadilla, 2002). In 2002, Morgan-Ryan and colleagues (2002) proposed to rename *C. parvum* human genotype (genotype 1 or genotype H) and create a new *Cryptosporidium* species *C. hominis* based on biological and molecular data. New *Cryptosporidium* species have been described including *C. suis* isolated from domestic pig (*Sus scrofa*) (Ryan *et al.,* 2004), *C. scophthalmi* found in turbot (*Scophthalmus maximus*) (Alvarez-Pellitero *et al.,* 2004), *C. bovis* described in domestic cattle (*Bos taurus*) (Fayer *et al.,* 2005), *C. fayeri* isolated from Kangaroo (*Macropus rufus*) (Ryan *et al.,* 2008), *C. ryanae* found in cattle (Bos taurus) (Fayer *et al.,* 2008), *C. macropodum* described in kangaroos (*Macropus giganteus*) (Power and Ryan, 2008) and *C. xiaoi* isolated from sheep (*Ovis aries*) (Fayer and Santín, 2009). These species are currently considered valid by Xiao and Fayer (2008) as summarized in Table 1.1.

Until 1970, *Cryptosporidium* species were not considered economically or medically important, but the veterinary importance of *Cryptosporidium* was highlighted by the association of *C. parvum* with bovine diarrhoea (Panciera *et al.,* 1971). Since then, several sporadic cases and outbreak reports were published and *C. parvum* is now regarded as an important cause of neonatal diarrhoea in calves and lambs associated with economic loses (O'Handley and Olson, 2006, De Graaf *et al.,* 1999, Current and Garcia, 1991, Moore and Zeman, 1991, Holland, 1990, Tzipori *et al.,* 1980, Meuten *et al.,* 1974). Another species, *C. baileyi,* is now recognized as an important cause of respiratory disease in poultry (O'Donoghue, 1995, Goodwin, 1989, Current and Snyder, 1988, Blagburn *et al.,* 1987).

Species	Original description	Type host
*Cryptosporidium andersoni**	Lindsay *et al.* (2000)	*Bos taurus* (domestic cattle)
Cryptosporidium baileyi	Current *et al.* (1986)	*Gallus gallus* (chicken)
Cryptosporidium bovis	Fayer *et al.* (2005)	*Bos taurus* (domestic cattle)
*Cryptosporidium canis**	Fayer *et al.* (2001)	*Canis familiaris* (domestic dog)
Cryptosporidium fayeri	Ryan *et al.* (2008)	*Macropus rufus* (red kangaroo)
*Cryptosporidium felis**	Iseki (1979)	*Felis catus* (domestic cat)
Cryptosporidium galli	Pavlasek (1999)	*Gallus gallus* (chicken)
*Cryptosporidium hominis**	Morgan-Ryan *et al.* (2002)	*Homo sapiens* (human)
Cryptosporidium macropodum	Power and Ryan (2008)	*Macropus giganteus* (grey kangaroo)
*Cryptosporidium meleagridis**	Slavin (1955)	*Meleagris gallopavo* (turkey)
Cryptosporidium molnari	Alvarez-Pellitero and Sitja-Bobadilla (2002)	*Sparus aurata* (gilthead sea bream) *Dicentrarchus labrax* (European seabass)
*Cryptosporidium muris**	Tyzzer (1910)	*Mus musculus* (house mouse)
*Cryptosporidium parvum**	Tyzzer (1912)	*Mus musculus* (house mouse)
Cryptosporidium scophthalmi	Alvarez-Pellitero *et al.* (2004)	*Scophthalmi maximus* (turbot)
Cryptosporidium serpentis	Levine (1980) Brownstein *et al.*, (1977)	*Elaphe guttata* (corn snake) *Elaphe subocularis* (rat snake) *Sanzinia madagascarensus* (Madagascar boa)
*Cryptosporidium suis**	Ryan *et al.* (2004)	*Sus scrofa* (domestic pig)
Cryptosporidium varanii	Pavlasek *et al.* (1995)	*Varanus prasinus* (Emerald monitor)
Cryptosporidium wrairi	Vetterling *et al.* (1971)	*Cavia porcellus* (guinea pig)

Table 1.1: Valid *Cryptosporidium* species, their original host and details of the original report (Reproduced from Xiao and Fayer (2008), with modification). (*) species found to infect humans.

The first cases of human cryptosporidiosis were reported in 1976 and associated with severe watery diarrhoea (Meisel *et al.,* 1976, Nime *et al.,* 1976). The public health significance of cryptosporidiosis became apparent when *Cryptosporidium* was recognized as a common cause of acute diarrhoea in immunocompetent individuals (Jokipii *et al.,* 1983, Current *et al.,* 1983, Tzipori *et al.,* 1983) and when chronic infections were associated with several mortalities in patients infected with human immunodeficiency virus (HIV) who had developed acquired immunodeficiency syndrome (AIDS) (O'Donoghue, 1995, Crawford and Vermund, 1988, Fayer and Ungar., 1986, Soave *et al.,* 1984, Current *et al.,* 1983, Ma and Soave, 1983, Forgacs *et al.,*1983).

The first report of a waterborne cryptosporidiosis outbreak was by D'Antonio and colleagues (1985). Since then, *Cryptosporidium* has been associated with several waterborne outbreaks. The importance of *Cryptosporidium* as a water borne pathogen was highlighted by the massive outbreak in Milwaukee (Wisconsin, USA) in 1993 affecting 403,000 persons (Mac Kenzie *et al.,* 1994). However, this estimate was disputed as retrospective community-based studies of diarrhoeal disease were shown to overestimate the incidence of illness (Hunter and Syed, 2001). The estimated cost of the Milwaukee outbreak was over 90 million US dollars (Corso *et al.,* 2003). Karanis and colleagues reviewed over 150 waterborne *Cryptosporidium* outbreaks worldwide (Karanis *et al.,* 2007), demonstrating the high prevalence and the widespread distribution of *Cryptosporidium*. The investigation of these outbreaks allowed an improved understanding of the epidemiology of *Cryptosporidium,* particularly identification of risk factors and transmission routes and provided better insight into the public health impact of waterborne cryptosporidiosis.

1.2 Taxonomy

When Tyzzer first identified *C. muris* he was uncertain of its systematic position, "its possession of an organ of attachment and of iodophilic granules, it resembles the gregarines", but "in its morphology, in the lack of motion in the adult, and in sexual dimorphism it resembles the coccidia" (Tyzzer, 1907). However, when he further characterized the parasite and officially proposed the new genus and species *Cryptosporidium muris*, it was placed in the class Sporozoa, subclass Telosporidia, order Coccidiomorpha, suborder Coccidia and within the family Eimeridae (Tyzzer, 1910). No flagella were demonstrated with the microgametes of *C. muris* but all the other requirements of the family were met, especially after the definition of the group had been modified to include extracellular Coccidia (Robinson, 2005, Tyzzer, 1910). Despite this classification, Tyzzer stated once more "it would appear that the mode of life of this species is essentially that of a gregarine as it undergoes the greater part of its development either attached to the surface of the epithelium or free in the lumen of the gastric glands" (Tyzzer, 1910).

The classification of *Cryptosporidium* has changed since then, the most current classification published by the Society of Protozoologists, is by Lee and colleagues (2000 a) (see Robinson, 2005):

Phylum **Apicomplexa** Levine, 1970
Class **Conoidasida** Levine, 1988
Subclass **Coccidiasina** Leuckart, 1879
Order **Eucoccidiorida** Leger and Duboscq, 1910
Suborder **Eimeriorina** Leger, 1911
Family **Cryptosporidiidae** Leger, 1911
Genus ***Cryptosporidium*** Tyzzer, 1907

This classification was based on the following biological characteristics of each group (Current and Garcia, 1991):

- Apicomplexa: Invasive forms have apical complex with polar rings, rhoptries, micronemes, conoid and sub-pellicular microtubules.
- Sporozoasida: locomotion of invasive forms by body flexion gliding or undulation.
- Coccidiasina: life cycle with merogony, gametogony and sporogony.
- Eucoccidiorida: merogony present, found in vertebrate hosts.
- Eimeriorina: male and female gametes develop independently.
- Cryptosporidiidae: homoxenous (one host life cycle), with developmental stages just under the membrane of the host cell. Oocysts without sporocysts and with four sporozoites. Microgametes with flagella.

However, the taxonomy of *Cryptosporidium* remains controversial until today, particularly the classification within the subclass Coccidiasina. Carreno and colleagues (1999) showed, based on SSU rRNA gene sequences, that *Cryptosporidium* species (*C. baileyi*, *C. parvum*, *C. wrairi*, *C. serpentis*, *C. muris*) form a monophyletic clade that is a sister group to the gregarines, parasites of invertebrates, belonging to the subclass Gregarinasina. This gregarine-*Cryptosporidium* clade was separated from the coccidia clade including *Sarcocystis*, *Toxoplasma* and *Eimeria* and supported by 95% bootstrap value. Other studies provided further support to the hypothesis that *Cryptosporidium* lineage is separate from Coccidia and is an early emerging lineage among the Apicomplexa (Kuo *et al.,* 2008, Leander *et al.* 2003, Zhu *et al.* 2000 a, Morrison and Ellis, 1997, Barta *et al.,* 1991).

These phylogenetic data were supported by biological evidence. In fact, despite *C. parvum* being a typical Coccidium in its morphology (apical complex, dense bodies, micronemes, rhoptries) and lifestyle, there are some fundamental differences between this parasitic protist and the eucoccidia (Zhu *et al.,* 2000 a).

These include: (i) the extracytoplasmic but intracellular location of *C. parvum* in a parasitophorous vacuole just beneath the enterocyte apical membrane; (ii) the presence of an acristate, ribosome-studded mitochondrion posterior to the nucleus (Riordan *et al.,* 1999); (iii) plant-like polyamine biosynthesis by decarboxylation of arginine rather than ornithine (Keithly *et al.,* 1997); (iv) the apparent lack of a plastid or plastid genome (Zhu *et al.,* 2000 b); (v) insensitivity to most anticoccidial drugs (Coombs, 1999, Woods *et al.,* 1996) and (vi) sporulation of oocysts within the intestine resulting in enterocyte reinvasion and prolonged life-threatening infection in immunocompromised patients. On the other hand, *Cryptosporidium* has common biological features with the gregarines including a monoxenous life cycle, oocysts with four sporozoites, a usual location in the host gastrointestinal tract and extracellular gamonts or trophozoites (Barta and Thompson, 2006). The last characteristic was established recently, when Hijjawi and colleagues (2002) described for the first time developmental stages in the life cycle of *C. andersoni* and *C. parvum* (extracellular trophozoites/gamont stages), which have similar characteristics to the gregarines. These results have subsequently been reproduced and validated in cell culture (Rosales *et al.,* 2005). The similarity to the gregarines was further supported by the cross reactivity of an anti-*Cryptosporidium* monoclonal antibody with the sporocysts of the gregarine *Monocystis* species (Bull *et al.,* 1998). Barta and Thompson (2006) compiled molecular and biological evidence for distancing *Cryptosporidium* species from the Coccidia conceptually, biologically and taxonomically.

The taxonomy at the species level of *Cryptosporidium* has also changed over time. Since the first description by Tyzzer of *C. muris* and *C. parvum*, *Cryptosporidium* was commonly confused with members of the coccidian genus *Sarcocystis* as they have similar morphological characteristics (thin walled oocysts that rupture releasing free sporocysts, each containing four sporozoites) (Xiao *et al.,* 2004). This ambiguity was resolved when the criterion of possession of unique attachment organelle was used as the key feature to define the family "Cryptosporidiidae**"** and the genus "*Cryptosporidium*"(Xiao *et al.,* 2004). After

the recognition of clear differences between *Cryptosporidium* and *Sarcocystis*, the erroneous concept of strict host specificity was applied to *Cryptosporidium* spp., which led to the description of several new species. Subsequently, cross-transmission studies showed that *Cryptosporidium* isolates from different animals can be transmitted from one host species to another and several of the proposed new species were synonymised (Xiao *et al.,* 2004). One disadvantage of adopting this methodology was the widespread use of the name *C. parvum* for *Cryptosporidium* parasites from all kinds of mammals including humans. More recently, the development of molecular characterization tools helped to clarify *Cryptosporidium* taxonomy and validate the existence of multiple species (Xiao *et al.,* 2004). These molecular data, supported by multiple parameters including morphology, developmental biology, host specificity and histopathology allowed the definition of new species (Morgan-Ryan *et al.,* 2002, Alvarez-Pellitero and Sitja-Bobadilla, 2002, Fayer *et al.,* 2001, Lindsay *et al.,* 2000). In fact, the suggested criteria for naming of a new *Cryptosporidium* species are i) morphometric measurement of the oocysts, ii) genetic analysis of common loci, iii) demonstration of natural and/or experimental host specificity and iv) compliance with the rules of the International Commission of the Zoological nomenclature (ICZN) (valid taxonomic description and name-bearing type established from syntypes (two or more specimens) (Xiao *et al.,* 2004, Egyed *et al.,* 2003). The establishment of these tailored species definition criteria for *Cryptosporidium* is important because of the difficulty in fulfilling the classical definition of species as groups of interbreeding natural populations reproductively isolated from other groups (Mayr, 1942). The evaluation of the interbreeding potential is tricky because it is difficult to conduct genetic crossing studies in *Cryptosporidium,* although few studies showed the occurrence of intraspecies sexual recombination (Feng *et al.,* 2002, Mallon *et al.,* 2003 b). In addition, *Cryptosporidium* has a large bias toward a clonal population structure, as mating normally occurs between siblings (high numbers of genetically similar parasites in localized areas) (Awad-El-Kariem, 1999, Gibbons *et al.,* 1998, Sulaiman *et al.,* 2001).

Xiao and colleagues (2004) reviewed the taxonomy of *Cryptosporidium* and concluded that only 13 *Cryptosporidium* species are valid and associated with a particular host range. These include *C. andersoni* (cattle), *C. baileyi* (chicken and some other birds), *C. canis* (dogs), *C. felis* (cats), *C. galli* (birds), *C. hominis* (humans), *C. meleagridis* (birds and humans), *C. molnari* (fish), *C. muris* (rodents and some other mammals), *C. parvum* (ruminants and humans), *C. wrairi* (guinea pigs), *C. saurophilum* (lizards and snakes) and *C. serpentis* (snakes and lizards). Since then, Xiao and Fayer (2008) updated the valid number of species to 18 to take into consideration the newly validated and newly discovered *Cryptosporidium* species as shown in Table 1.1. In addition, the authors described 40 *Cryptosporidium* genotypes, which are genetically distinct populations for which there is insufficient biological information to assign to species status. The genotypes are usually named after the host they were isolated from (such as rabbit genotype, cervine genotype, goose genotype, skunk genotype, deer genotype). However, this practice is confusing as the genotype may not exclusively infect that animal or even that the named species is the major host (Robinson, 2005). As biological and molecular data increase, many of these genotypes are expected to be named as valid species. This was the case of *C. parvum* genotype 1 now known as *C. hominis*, dog genotype called *C. canis* and pig genotype 1 considered *C. suis* (Morgan-Ryan *et al.,* 2002, Fayer *et al.,* 2001, Ryan *et al.,* 2004).

1.3 Life cycle

The *Cryptosporidium* life cycle can be divided into six major developmental stages (Current and Garcia, 1991): excystation (the release of infective sporozoites), merogony (the asexual multiplication within host cells), gametogony (the formation of micro- and macro-gametes), fertilization (the union of micro- and macro-gametes), oocyst wall formation (to produce the environmentally resistant stage responsible for the transmission of the infection from one host to another) and sporogony (the formation of infective sporozoites).

Cryptosporidium species have a monoxenous life cycle completed within the gastrointestinal tract of a single host. During the whole cycle, the different forms are confined to the apical surfaces of the intestine epithelial cells (enterocytes). The infective forms attach to the apical surfaces by a poorly understood process and become internalized within an intracellular but extracytoplasmic compartment separated from the cytoplasm by an electron dense layer that appears to be predominantly of host origin (Barta and Thompson, 2006). The parasite is located within a parasitophorous vacuole, where it is protected from the hostile gut environment and is supplied with energy and nutrients by the host cell through a feeder organelle, which is unique among apicomplexan parasites (Tzipori and Ward 2002).

The life cycle begins with the ingestion of the sporulated oocysts by the susceptible host; the oocysts undergo excystation and release four infective sporozoites (Figure 1.1). The excystation of the oocysts has been reported to be triggered upon ingestion by various factors including reducing conditions, carbon dioxide, temperature, pancreatic enzymes and bile salts (O'Donoghue, 1995, Robertson *et al.,* 1993, Sundermann *et al.,* 1987, Reduker and Speer, 1985, Fayer and Leek, 1984). The excystation allows the emergence of the four infectious sporozoites through a suture in the oocyst wall (Reduker *et al.,* 1985). The released sporozoites glide over the intestinal cell releasing material from the apical complex (Okhuysen and Chappell, 2002). Gliding motility allows the zoites to migrate across the surface of host cells and to actively invade them (Wetzel *et al.,* 2005). This form of motility is conserved among different apicomplexan parasites. *C. parvum* sporozoites undergo circular and helical gliding movements, which are actin-myosin-tubulin dependent mechanisms (Wetzel *et al.,* 2005, Chen *et al.,* 2004). During gliding motility, sporozoites deposit trails of proteins, which are involved in attachment and invasion of host cells (Wanyiri and Ward, 2006).

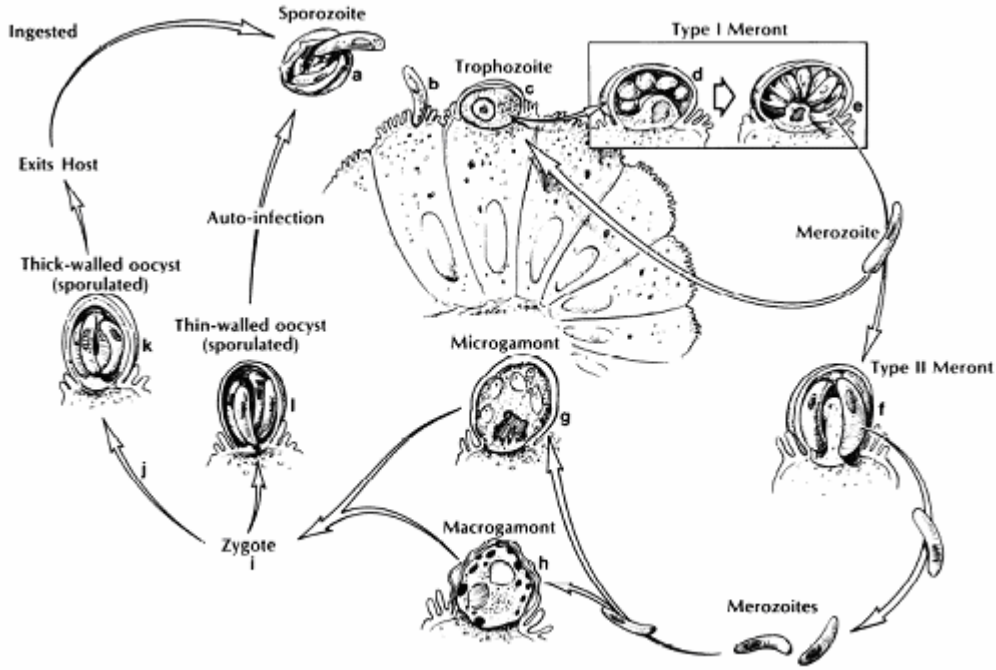


Figure 1.1: Schematic representation of *Cryptosporidium* life cycle stages.

(a) The oocyst excysts in the lumen of the intestine, releasing the infective sporozoites. (b) The sporozoite attaches to the surface of the enterocyte. (c) The sporozoite penetrates into host cells and develops into trophozoites within a parasitophorous vacuole. (d) and (e) The trophozoites undergo merogony to form merozoites. (f) After being released from type I meront, the invasive merozoites enter adjacent host cells to form additional type I meronts (recycling of type I meronts) or to form type II meronts. (g) and (h) Type II meronts enter host cells to form the sexual stages microgamonts and macrogamonts. (i) The zygote, product of fertilization of the macrogamont by the microgametes (released from microgamont). (j) The majority (80%) of the zygotes develop into environmentally resistant, thick-walled oocysts. (k) The thick walled oocysts undergo sporogony to form sporulated oocysts containing four sporozoites, which will be released in the faeces and allow transmission of the infection. (l) A smaller proportion (20%) of the zygotes have a unit membrane surrounding the four sporozoites, these are the thin-walled oocysts, which are the auto-infective forms that can maintain the parasites in the host without repeated exposure.

(Reproduced from Current and Garcia, 1991 with modifications)

In addition, *Cryptosporidium* sporozoites possess an apical complex composed of micronemes, a single rhoptry and dense granules (Tetley *et al.,* 1998). These secretory organelles are also present in closely related parasites such as *Toxoplasma*, *Plasmodium* and *Eimeria* and are involved in host-cell attachment and invasion (Boulter-Bitzer *et al.,* 2007, Okhuysen and Chappell, 2002). The pellicle mediates movement and attachment to the host cell membrane via molecules exposed on the surface that drive movement of the zoite, while the rhoptries and micronemes enable the zoite to adhere to and invade the cell, inducing the cell membrane to enclose the parasite in the parasitophorous vacuole (Boulter-Bitzer *et al.,* 2007, Bonnin *et al.,* 1993, Petersen *et al.,* 1992b, Tomley and Soldati, 2001).

Each sporozoite develops into a spherical trophozoite, which undergoes merogony and forms a type I meront containing eight merozoites (O'Donoghue, 1995). These merozoites are released and attach again to the surface of an epithelial cell where they undergo merogony once more and either form a further type I meront or a type II meront (Figure 1.1). A type II meront contains four merozoites. These merozoites, when released, attach again to the epithelium but instead of developing into further meronts, they initiate gametogony (Robinson, 2005). Individual merozoites produce either microgamonts or macrogamonts (Smith and Rose, 1998, Göbel and Brändler, 1982). Each microgamont undertakes nuclear division and differentiates to form up to 16 microgametes, which when released from the parasitophorous vacuole, locate and fertilise a unicellular macrogametocyte that has developed from a macrogamont (Figure 1.1). The product of fertilization, the zygote, undergoes two asexual cycles of sporogony to produce an oocyst with either a thick wall or a thin wall, containing four sporozoites (Current and Reese, 1986).

The thick-walled oocysts are released into the lumen of the intestine and excreted from the host in the faeces and are immediately infective allowing the spread of the infection to other susceptible hosts (Smith and Rose, 1998). In

addition, unlike other coccidian parasites, *Cryptosporidium* is able to autoinfect the same host. The auto-infection occurs through the thin-walled oocysts, which excyst once separated from the epithelium and the cycle starts again (Sińsk and Behnke, 2004, Current and Reese, 1986). The autoinfection and the recycling of type I meronts provide an explanation for the persistent chronic infection and the low infective dose (DuPont *et al.,* 1995, Okhuysen *et al.,* 1999, Jokipii and Jokipii, 1986).

1.4 *Cryptosporidium* culture

Attempts to culture *Cryptosporidium in vitro* have not been particularly successful, often resulting in an incomplete cycle or only short term maintenance of the infection. However, there have been reports of successful *in vitro* culture of *Cryptosporidium*. Gut and colleagues (1991) reported the culture of *C. parvum* and the observation of the different life stages (trophozoites, meronts, microgametocytes, and macrogametocytes) in Madin-Darby canine kidney (MDCK) cells. *C. parvum* culture was also successful using RL95-2 (human endometrial carcinoma) cell line (Rasmussen *et al.,* 1993), HCT-8 (human ileocecal colorectal adenocarcinoma) cells (Upton *et al.,* 1995, Meloni and Thompson, 1996), BFTE (bovine fallopian tube epithelial) cells (Yang *et al.,* 1996), Caco-2 (human colorectal adenocarcinoma) cell line (Arrowood *et al.,* 1994), HT29.74 (human enterocyte) cell line (Elliot *et al.,* 1997) and BS-C-1 (African green monkey kidney) cell line (Deng and Cliver, 1998). The cell lines used to culture *Cryptosporidium* were reviewed by Carey and colleagues (2004). Upton and colleagues (1994) compared the efficiency of 11 cell lines to culture *Cryptosporidium* and showed that HCT-8 supported a high number of parasite developmental stages. This cell line was considered an accurate alternative to animal infectivity models (Rochelle *et al.,* 2002). Furthermore, the use of surface-sterilized oocysts (by pre-treatment with sodium hypochlorite) seems to improve the efficiency of cell monolayers infection (Upton *et al.,* 2004 b). Hijjawi and colleagues (2001) reported for the first time the long-term maintenance of the

life cycle stages of *C. hominis* and *C. parvum* in HCT-8 cell line. In addition, *C. andersoni* was also successfully cultured in HCT-8 cells (Hijjawi *et al.,* 2002).

Later, Hijjawi and colleagues (2004) reported the continuous development of *C. parvum* in cell-free medium through all life cycle stages. These results suggest that *Cryptosporidium* may not be an obligate intracellular parasite as previously stated. The implication of these findings are that the greatly reduced biochemical repertoire retained by *Cryptosporidium* species after the loss of both a functional mitochondrion and an apicoplast only requires a nutrient-rich environment to salvage its metabolic needs (Hijjawi *et al.,* 2004). However, attempts to reproduce cell-free culture of *Cryptosporidium* were unsuccessful (Girouard *et al.,* 2006, Karanis *et al.,* 2008, Petry *et al.,* 2009). In addition, Woods and Upton (2007) reported that the developmental stages described by Hijjawi and colleagues (2004) correspond to budding yeasts, host cells and fungal conidia. Therefore, the ability of *Cryptosporidium* to develop extracellularly is still controversial (Xiao and Fayer, 2008). Since then, the same group showed that the extracellular developmental stages are indeed specific of *Cryptosporidium* using antibody staining (anti-sporozoite and anti-oocyst wall antibodies) and a *Cryptosporidium* specific rRNA oligonucleotide probe for fluorescent in-situ hybridization (FISH) in combination with non-specific dyes (Boxell *et al.,* 2008). Hijjawi (2010) reviewed the factors that can influence the success of *Cryptosporidium* cell-free culture (excystation, pre-treatment of oocysts, physical and chemical conditions, host cell type and maturity and culture media supplementation and formulation) and proposed improved methods for visualizing life cycle stages in cell-free culture.

1.5 Epidemiology

Cryptosporidiosis is a common cause of gastroenteritis worldwide. In England and Wales, an average of 4,500 laboratory identified cases is reported each year (1998-2008) according to the Health Protection Agency (HPA) (www.hpa.org.uk). The two *Cryptosporidium* species of public health importance are *C. hominis* and *C. parvum*. *C. hominis* is generally restricted to human infections and *C. parvum* is zoonotic and mostly associated with human and ruminant infections (Roberston and Gjerde, 2007). A long-term survey at the national level conducted by Chalmers and colleagues (2009 c) showed that the majority of human infective *Cryptosporidium* species were either *C. parvum* or *C. hominis*. *Cryptosporidium* is recognized mainly as a waterborne parasite, Karanis and colleagues (2007) reviewed over 150 worldwide waterborne outbreaks.

The epidemiology of *Cryptosporidium* has mainly been investigated at the genus level (Robertson *et al.,* 2002, Roy *et al.,* 2004, Horman *et al.,* 2004, Becher *et al.,* 2004, Wallis *et al.,* 1996, Okafor *et al.,* 1996, Nimri and Batchoun, 1994) and no attempt was made to distinguish between *Cryptosporidium* species or genotypes in case control studies until Hunter and colleagues (2004 b) showed different risk factors for *C. hominis* and *C. parvum.* Subsequently, several epidemiological investigations with speciation (Blanco *et al.,* 2009, Mueller-Doblies *et al.,* 2008, Leoni *et al.,* 2006) and even subtyping (Brook *et al.,* 2009, Xiao *et al.,* 2009, Geurden *et al.,* 2009, Zintl *et al.,* 2009, Chalmers *et al.,* 2008, Hunter *et al.,* 2007) have been reported. Identification of isolates at the species level has improved our understanding of the epidemiology and risk factors associated with these two *Cryptosporidium* species and has been valuable in outbreak investigations (Chalmers, 2008 b). In recent years, the epidemiology of *Cryptosporidium* has been extensively reviewed (Xiao, 2010, Yoder and Beach, 2010, Xiao and Fayer, 2008, Tzipori and Widmer, 2008, Xiao and Feng, 2008, Hunter and Thompson, 2005, Caccio, 2005, Ramirez *et al.,* 2004, Hunter and Nichols, 2002, Fayer *et al.,* 2000, Meinhardt *et al.,* 1996).

1.5.1 Geographical and seasonal distribution

Epidemiologic surveys indicate that human *Cryptosporidium* infection is distributed worldwide in urban and rural populations (Meinhardt *et al.,* 1996). Geographic differences exist in the disease burdens attributable to *C. hominis* and *C. parvum*. *C. hominis* has been shown to be more prevalent in North and South America, Australia and Africa suggesting that the transmission cycle of the anthroponotic parasite is of greatest importance in these regions (Kosek *et al.,* 2001). However, *C. parvum* is the predominant species in UK, Northern Ireland, France, Switzerland, Portugal, Slovenia, Czech Republic and New Zealand (Xiao and Feng, 2008, Boulter-Bitzer *et al.,* 2007, Ong *et al.,* 1999).

Seasonal differences in the distribution of *C. parvum* and *C. hominis* have also been reported. In the UK and New Zealand, the spring increase in the reported cryptosporidiosis cases is mostly due to *C. parvum* whereas the autumn increase is largely due to *C. hominis* (Leoni *et al.,* 2006, Learmonth *et al.,* 2004, Hunter *et al.,* 2004 b, McLauchlin *et al.,* 2000). This suggests that seasonal differences between the specific transmission routes might exist: it was speculated that the increase in *C. parvum* in the spring was due to lambing, calving and farm runoff from spring rains and the autumn *C. hominis* peak was likely the result of increased recreational water activities and international travel during late summer and early autumn (Xiao and Feng, 2008, Goh *et al.,* 2004, Hunter *et al.,* 2004 b).

1.5.2 Transmission

Several direct and indirect transmission routes of *Cryptosporidium* have been identified. Direct transmission occurs by the faecal-oral route through infected subjects, this includes animal to animal, animal to human (zoonotic) and human to human (anthroponotic) transmission (Xiao and Feng, 2008, Hunter and Thompson, 2005, Casemore *et al.,* 1985). Human to human contamination is

well-described, particularly in secondary cases, in outbreak settings (Glaberman *et al.,* 2002) and in day care centres and hospitals (Guerrant, 1997). The epidemiological evidence for zoonotic transmission of cryptosporidiosis is from studies associating cattle with outbreaks in veterinary students and animal researchers, who had contacts with infected young calves and in children visiting farms (Xiao and Feng, 2008, Kiang *et al.,* 2006, Smith *et al.,* 2004, Preiser *et al.,* 2003). Indirect transmission involves contact with *Cryptosporidium* faecally contaminated material, including water, food, clothes and footwear. Indirect transmission also occurs through environmental contamination, usually involving the release of untreated sewage by accident or as overflow following heavy rain events (Jiang *et al.,* 2005, Xiao *et al.,* 2000, Hayes *et al.,* 1989). Another mode of transmission via inhalation of oocysts was reported in immunocompromised patients and in children (Egger *et al.,* 1990, Harari *et al.,* 1986, Ma *et al.,* 1984) and was supported by experimental intranasal infection of piglets (Tzipori and Ward, 2002). The symptoms associated with this route are respiratory (laryngotracheitis) and could be accompanied by mild diarrhoea.

The oocyst is responsible for the dissemination of the infection. It is of small size (from 3.8 by 4.6 µm to 6.3 by 8.4 µm, depending on species) of spheroid or ellipsoid shape and is notoriously robust (Roberston and Gjerde 2007). Oocysts are excreted into the environment by the infected hosts and remain viable for a long time in moist and relatively cold environments (six months between 0 and 20°C, three months at 25-30°C, two months at -5°C, one week at 35°C, 8h at -20°C, while freezing below -20°C causes immediate killing of oocysts (Fayer *et al.,* 2000). Relatively little is presently known about the chemical composition of oocysts and how this composition is associated with survival ability (Millar *et al.*, 2002). The infective dose is low; Dupont *et al.* (1995) reported an ID_{50} of 132 oocysts (*C. parvum* Iowa) in healthy volunteers. Another human volunteer study using three *Cryptosporidium* strains showed the ID_{50} is variable between strains and the reported ID_{50} values were 9, 87 and 1042 for TAMU, Iowa and UCP strains, respectively (Okhuysen *et al.,* 1999).

Infected hosts act as reservoirs, allowing the perpetuation of the infection. Human and dairy effluents are probably the most important sources of environment and surface water contamination. The duration of oocyst shedding has been documented in humans and varies from 6.9 days to two months after the cessation of symptoms which normally last between 2-26 days in immunocompetent individuals (Jokipii and Jokipii, 1986). It is now recognised that the infectious dose as well as the duration and severity of illness are influenced by several factors including the species and strain of the parasite, as well as the age and immune status of the host (Robinson, 2005).

Oocysts are incompletely retained by sand filtration and are highly resistant to conventional water disinfection used in the water industry (Fayer *et al.* 1998). In addition, it has been demonstrated that the oocysts maintain their infectivity for several months in both salt and fresh water. The vulnerability of the water supply to *Cryptosporidium* oocysts has heightened biodefense concerns (Striepen and Kissinger, 2004). *Cryptosporidium* is now considered one of the most ubiquitous and difficult to control agent of waterborne disease. Therefore, and due to the relative ease of intentional contamination of water supplies, *Cryptosporidium* has been listed as a Category B priority pathogen for Bioterrorism by the Centre for Disease Control and Prevention (CDC) and National Institutes of Health (NIH) in the United States (http://www.bt.cdc.gov/agent/agentlist.asp#categorybdiseases).

Cryptosporidium has been incriminated in many waterborne outbreaks. Smith and Rose (1998) documented 19 outbreaks in UK and USA between 1984 and 1996, affecting an estimated number of 427,100 individuals. Karanis and colleagues (2007) updated the number of waterborne cryptosporidiosis outbreaks and highlighted that *Cryptosporidium* is responsible for 50% of waterborne outbreaks associated with protozoan parasites. *Cryptosporidium* oocysts also survive for long periods on different foodstuffs. Several studies

(Smith *et al.*, 2007, Millar *et al.*, 2002) highlighted the importance of foodborne cryptosporidiosis and documented related outbreaks.

1.5.3 *Cryptosporidium* risk factors

As discussed above, most of the data on transmission pathways for *Cryptosporidium* come from reports of outbreaks, the majority of which are waterborne (Karanis *et al.,* 2007). However, outbreaks represent only a small proportion of recorded cryptosporidiosis cases. The transmission routes for endemic diseases may or may not be the same as in outbreak settings. This has been evaluated in case control studies, the majority of which were reviewed by Yoder and Beach (2010) and Hunter and Thompson (2005). The key risk factors are ingestion of contaminated drinking or recreational water, contact with infected persons or animals, travel to disease endemic areas, contact with children < 6 years old (especially but not exclusively with diarrhoea) as shown in Table 1.2. These investigations also showed a consistent negative association with eating raw vegetables (Hunter *et al.*, 2004, Roy *et al.,* 2004, Robertson *et al.,* 2002). The authors suggested that repeated exposure to small numbers of *Cryptosporidium* oocysts on raw vegetables provide a protective immunity. This hypothesis was supported by a study by Frost and colleagues (2005), who reported that higher levels of anti-*Cryptosporidium* antibodies were associated with reduced reporting of diarrhoea in a subsequent diary-based study. The authors suggested that the lower levels of illness in people with increased antibodies may be due to enhanced immunity from repeated exposure to oocysts in drinking water.

Reference	Location	No. of cases	No. of controls	Significant risk factors	Odds ratios (95% CI)
Robertson *et al.*, 2002	Melbourne	201	795	Eating uncooked carrots	0.6 (0.4–0.9)
				Swimming in public pool	2.7 (1.9–3.8)
				Children<6 at home with diarrhoea	7.4 (4.0–13.8)
				Persons>5 at home with diarrhoea	1.8 (1.1–2.9)
				Animal contact in home	0.6 (0.4–0.8)
				Calf contact away from home	2.9 (1.5–5.7)
				Drink unboiled water from river, lake or dam	1.5 (0.8–2.7)
Robertson *et al.*, 2002	Adelaide	134	536	Eating uncooked carrots	0.6 (0.4–0.9)
				Swimming in public pool	1.2 (0.8–1.9)
				Children<6 at home with diarrhoea	8.6 (4.8–15.6)
				Persons>5 at home with diarrhoea	3.7 (2.2–6.2)
				Animal contact in home	0.6 (0.4–0.9)
				Calf contact away from home	5.1 (1.5–17.3)
				Drink unboiled water from river, lake or dam	3.1 (1.5–6.5)
Roy *et al.*, 2004	USA	282	490	Contact with persons (>2 to 11 yr old) with diarrhoea	3.0 (1.5–6.2)
				Contact with calves or cows	3.5 (1.8–6.8)
				International travel	7.7 (2.7–22.0)
				Freshwater swimming	1.9 (1.0–3.5)
				Eating raw vegetables	0.5 (0.3–0.7)
				Chronic medical condition	2.2 (1.2–4.0)
Hunter *et al.*, 2004 b	Wales and England	427	427	Travel outside of UK	5.7 (2.9–11.2)
				Case contact	4.6 (2.4–8.7)
				Touch cattle	3.9 (1.4–10.0)
				Toileting child<5 years	1.9 (1.1–3.2)
				No. of glasses unboiled water	1.1 (1.0–1.3)
				Eat ice cream	0.5 (0.3–0.7)
				Eat raw vegetables	0.5 (0.3–0.8)
				Eat tomatoes	0.6 (0.4–1.0)
Goh *et al.*, 2004	Cumbria, England	152	466	Drinking cold unboiled tap water	1.4 (1.14-1.71)
				Visiting farms	2.02 (1.04-3.9)
Pintar *et al.*, 2009	Ontario, Canada	36	803	Swimming in a lake or a river	2.9 (1.2-7.4)
				Drinking municipal water	2.4 (1.04-5.7)
				Having a family member with diarrhoeal illness	2.9 (1.3-6.4)
Valderrama *et al.*, 2009	Colorado, USA	47	92	Swallowing untreated water from lake, river or stream	8.0 (1.3-48.1)
				Exposure to recreational water	4.6 (1.4-14.6)
				Contact with child or diapers	3.8 (1.5-9.6)

Table 1.2: *Cryptosporidium* risk factors as determined by case control studies in developed nations reported since the early 1990s. Table compiled from data obtained from the following publications: Hunter and Thompson, 2005, Pintar *et al.,* 2009, Valderrama *et al.,* 2009, Goh *et al.*, 2004.

In immunocompromised subjects, mainly those with HIV infection, several studies investigated the specific risk factors for this high risk group and found that drinking untreated tap water, exposure to pets and animals, unsafe homosexual activity and use of public toilets are associated with *Cryptosporidium* infection (Dwivedi *et al.,* 2007, Colford *et al.,* 2005, Hellard *et al.,* 2003).

As has already been discussed, *C. hominis* and *C. parvum* have different host range and transmission cycles, and so epidemiological studies that looked at risk factors for both species combined, would emphasise risk factors common to both species and downplay risk factors unique to one or the other species (Hunter and Thompson, 2005). Therefore, a case control study discriminating *C. hominis* from *C. parvum* was performed in immunocompetent subjects by Hunter and colleagues (2004 b) and allowed the identification of species-specific risk factors. The major risk factor for *C. parvum* was touching or handling farm animals, whilst for *C. hominis* the main risk factors were travel outside the UK and nappy changing.

1.6 Clinical symptoms

Cryptosporidium is recognized worldwide as a common cause of infectious gastroenteritis, which characteristically results in watery diarrhoea that may sometimes be profuse and prolonged (Meinhardt *et al.,* 1996, Current and Garcia, 1991). Diarrhoea is generally the symptom leading to *Cryptosporidium* diagnosis. Other less common clinical features include abdominal pain, nausea, vomiting and mild fever. Occasionally, non-specific symptoms such as myalgia, weakness, malaise, headache and anorexia occur (Current and Garcia, 1991). The severity, persistence and ultimate outcome of the infection are typically dependent on a variety of parasite characteristics and host factors. Host factors include both the immune status and frequency of exposure of the infected individual, however, little is known regarding the pathogenic characteristics of the parasite (Meinhardt *et al.,* 1996). The severity of a *Cryptosporidium* infection

can vary from asymptomatic shedding of the oocysts to a severe and life-threatening disease. Immunocompetent individuals experience a short-term illness with complete spontaneous recovery. However, in immunocompromised patients, cryptosporidiosis symptoms are persistent and may become life-threatening (Chen *et al.,* 2002, O'Donoghue, 1995). Thus, only 1% of the immunocompetent patients require hospitalization, with very little risk of mortality. In immunocompromised patients, cryptosporidiosis is associated with a high rate of mortality (50%) (Junarek, 1995). In addition, *Cryptosporidium* infection causes atypical manifestations in immunocompromised patients, such as atypical gastrointestinal disease, biliary tract disease, respiratory tract disease and pancreatitis (Hunter and Nichols, 2002). *Cryptosporidium* species are significant sources of gastrointestinal infection particularly in developing nations, where cryptosporidiosis is common in children and it is frequently associated with persistent diarrhoea, malnutrition and stunted growth (Guerrant, 1997).

C. hominis and *C. parvum* account for more than 90% of the human cases of cryptosporidiosis (Cacciò, 2005). A retrospective study conducted by Hunter and colleagues (2004) showed differences between infections caused by *C. hominis* and *C. parvum* in immunocompetent subjects. In fact, *C. hominis* infection was associated with non-intestinal sequelae (joint pain, eye pain, recurrent headache and fatigue), which were not reported in people infected with *C. parvum*. This aspect further highlights the usefulness of speciation in *Cryptosporidium* diagnosis.

1.7 Treatment

The major health problem associated with cryptosporidiosis, mainly in immunocompromised patients, has been the lack of effective cure. Despite many attempts to identify antigens that will stimulate a protective immune response, no immunotherapeutics or vaccines are presently approved for prevention or

treatment of cryptosporidiosis in animals or humans (Sunnotel *et al.,* 2006, Fayer, 2004, Tzipori, 1998, Tzipori and Griffiths, 1998).

Many compounds were screened for anti-cryptosporidial activity and the majority were ineffective. Among the most commonly used treatments against cryptosporidiosis are paramomycin, azithromycin and nitazoxanide (Gargala, 2008). However, they have varying success, often with promising results in immunocompetent but poor efficacy in impaired immune system patients (Amadi *et al.,* 2009, Amadi *et al.,* 2002, Hewitt *et al.,* 2000). Nitazoxanide significantly shortened the duration of diarrhoea and decreased mortality in adults and in malnourished children (Gargala, 2008, Cohen 2005). However, it is not effective without an appropriate immune response. Therefore, in AIDS patients, combination therapy restoring immunity along with antimicrobial treatment of *Cryptosporidium* infection is necessary to control the infection (Gargala, 2008, Smith and Corcoran, 2004). Since then, highly active antiretroviral therapy (HAART) has been used to successfully cure cryptosporidia and microsporidia in HIV patients (Miao *et al.,* 2000). This success was related to increasing the patient's immune system by suppressing the virus and boosting the levels of CD4 lymphocytes (Schmidt *et al.,* 2001) and to the presence of protease inhibitors thought to have a direct effect on cryptosporidial infection (Mele *et al.,* 2003). Therefore, HAART is currently used for therapeutic and prophylactic treatment of cryptosporidiosis in persons at risk, especially AIDS patients (Sunnotel *et al.,* 2006, Willemot and Klein, 2004). However, treatment of underlying immunosuppression with antiretrovirals is not effective in immunocompromised patients without HIV (Abubakar *et al.,* 2007). Therefore, discovering an effective treatment for cryptosporidiosis is still highly desirable. To promote the development of new therapeutic and prophylactic treatment, further research is required to understand parasite metabolism, the invasion process and host-parasite interactions (Smith *et al.* 2005).

1.8 Prevention and control

Preventative measures are the most effective methods to control cryptosporidiosis in the absence of effective treatment. Measures such as limited human access to the countryside, containment of animals and restriction of livestock movement for trade or to pastures, which operated during the foot and mouth disease (FMD) epidemic in the UK in 2001, correlated with a marked decline (36.5%) in human cases of cryptosporidiosis caused by *C. parvum* (Smerdon *et al.,* 2003, Hunter *et al.,* 2003).

Ramirez and colleagues (2004) suggested measures to decrease the spread of the oocysts in the environment and to prevent water supply contamination including good hygiene, good cattle husbandry, diagnosis of cryptosporidiosis in domestic animals and household decontamination. These measures are relevant to *C. parvum* infection. Awareness of specific risk factors and taking relevant actions such as hand washing after nappy changing should likewise reduce *C. hominis* infection (Hunter *et al.,* 2004 b). For foodborne cryptosporidiosis, Millar and colleagues (2002) highlighted the importance of the ingestion of contaminated water and foodstuffs as a transmission route and issued recommendations to the food industry to reduce the likelihood of contamination in the human food chain. In addition, during the high prevalence season (as described in 15.3), water companies should reinforce the monitoring of water sources to ensure early detection of oocysts. In addition, a mobilization of medical staff and increased awareness of both medical staff and general public is likely to assist the prevention and control of cryptosporidiosis.

1.9 Detection

The detection of *Cryptosporidium* is important, not only for identifying the cause of disease but also in epidemiological surveys, disease surveillance and drinking water monitoring (Robinson, 2005). Several microscopic, molecular and

immunological methods have been developed to detect *Cryptosporidium* from different sources such as faeces, water, food and environmental samples. *Cryptosporidium* diagnosis is based upon oocysts visualization or detection of *Cryptosporidium* DNA or antigens.

1.9.1 Conventional methods

The conventional detection methods involve the microscopic identification of oocysts of 4-6 µm x 5-6 µm (Robinson, 2005, Fayer *et al.,* 2000). This method is widely used due to its relatively low cost. Differential staining using the modified Ziehl-Neelsen acid fast technique (Chalmers *et al.,* 2002, Nimri and Batchoun, 1994, Garcia *et al.,* 1983, Henriksen and Pohlenz, 1981) is the most commonly used technique to stain *Cryptosporidium* oocysts. In addition, other staining methods were also used such as Giemsa (Fischer, 1982), Auramine (Ratnam *et al.,* 1985), Kinyoun (Ma and Soave, 1983) and safranin-methylene blue stain (Baxby *et al.,* 1984). Staining techniques have poor specificity due to the presence of oocyst-like structures in faecal debris and because they stain other protozoan parasites such as *Isospora* and *Cyclospora* (Sunnotel *et al.,* 2006). Thus, immunological and molecular techniques have replaced staining for diagnostic of cryptosporidiosis.

1.9.2 Immunological techniques

Immunological and molecular techniques usually require further purification of oocysts using density gradient (Clark, 1999), saturated-salt solution centrifugation (Elwin *et al.,* 2001) or immunomagnetic separation (IMS) (Rochelle *et al.,* 1999, Bukhari *et al.,* 1998, Deng *et al.,* 1997).

The Immunofluorescence assay (IFA) is widely used for the detection of *Cryptosporidium* oocysts due to its high sensitivity and specificity (Deng and Cliver, 1998, Xiao *et al.,* 1993, Rusnak *et al.,* 1989, Garcia *et al.,* 1987, Stibbs and

Ongerth, 1986). IFA is based on the microscopic screening after staining with monoclonal antibodies tagged with fluorescein isothiocyanate (FITC) and several commercial IFA kits for *Cryptosporidium* detection are available. Immunofluorescence stains are often used in combination with a DNA strain mainly 4',6-diamidino-2- phenylindole (DAPI) to allow the identification of up to four sporozoites in the oocyst (Robinson, 2005). Internal structures can also be observed using differential interference contrast (DIC) microscopy. Several studies have compared staining protocols for the detection of *Cryptosporidium* oocysts (Quilez *et al.,* 1996, Alles *et al.,* 1995, MacPherson and McQueen, 1993, Mtambo *et al.,* 1992, Garcia *et al.,* 1992, Rusnak *et al.,* 1989, Arrowood and Sterling, 1989, Garcia *et al.,* 1987, Stibbs and Ongerth, 1986, Baxby *et al.,* 1984, Garcia *et al.,* 1983), the majority of which demonstrated that immunofluorescence staining techniques have not only superior sensitivity, especially when small number of oocysts were present in the samples, but also improved specificity due to the use of *Cryptosporidium* specific antibodies.

In addition, Enzyme immunoassays (EIA) have been used to detect *Cryptosporidium* oocysts. The first EIA developed for the detection of *Cryptosporidium* in faecal smears were described in 1990 (Robert *et al.,* 1990, Ungar, 1990, Anusz *et al.,* 1990). The threshold of detection of these techniques was $>10^6$ and $3x10^5$ oocysts per ml, so their sensitivity was much lower than IFA (10^3 oocysts per ml) (Marks *et al.,* 2004, Garcia and Shimizu, 1997). Since these first generation assays, several EIA kits were developed and are commercially available. Ignatius and colleagues (1997) compared three EIA kits and showed that they exhibit lower sensitivities than the routinely used microscopic methods.

1.9.3 Molecular detection techniques

Several molecular methods have been developed for the detection of *Cryptosporidium.* The majority of these methods involve the amplification of

Cryptosporidium nucleic acid, usually by polymerase chain reaction (PCR) (Bouzid *et al.,* 2008). PCR-based methods can often detect much lower parasite numbers than other detection techniques. Single oocyst detection has been reported by several authors and for different PCR systems as reviewed by Wiedenmann and colleagues (1998). In addition to the high sensitivity, the use of genus or species-specific primers has increased the assay specificity. Due to this high sensitivity and specificity, PCR has been extensively used for *Cryptosporidium* detection (Clark, 1999, Fischer *et al.,* 1998, Fayer *et al.,* 2000). Several genetic targets were used, the most popular targets include COWP (*Cryptosporidium* oocyst wall protein), HSP70 (Heat shock protein 70), SSU rRNA (Small subunit rRNA), ITS1-2 (Internal spacer 1), 18S rRNA (18S ribosomal RNA), GP60 (Glycoprotein 60), Trap1 and 2 (Thrombospondin related adhesive protein of *Cryptosporidium* 1 and 2) and DHFR (dihydrofolate reductase) genes. The different PCR protocols used have been widely reviewed (Sunnotel *et al.,* 2006, Carey *et al.,* 2004, Cacciò, 2005, Chappell and Okhuysen, 2002, Morgan and Thompson, 1998, Quintero-Betancourt *et al.,* 2002). One major limitation of PCR is its susceptibility to inhibition by a variety of substances, which can be present not only in stool samples but also in environmental materials and food (Wiedenmann *et al.,* 1998). These inhibitors can be removed using sucrose flotation, Percoll density gradient centrifugation, fluorescence activated cell sorting (FACS) and immunomagnetic separation (IMS) followed by DNA or RNA extraction methods (Wiedenmann *et al.,* 1998).

To increase the PCR specificity and sensitivity, several PCR variants have been developed such as nested PCR (Kostrzynska *et al.,* 1999, Balatbat *et al.,* 1996, Zhu *et al.,* 1998, Sturbaum *et al.,* 2001) and real-time PCR assays (Higgins *et al.,* 2001, MacDonald *et al.,* 2002, Limor *et al.,* 2002, Fontaine and Guillot, 2002) for oocysts detection in water (including sewage) with a sensitivity as low as one oocyst (Guy *et al.,* 2003).

1.9.4 Oocysts viability assessment

For an improved assessment of the human risk related to acquiring cryptosporidiosis from detected oocysts, viability techniques have been developed aiming to differentiate between viable and non viable oocysts in various sample types (Sunnotel *et al.,* 2006, Quintero-Betancourt *et al.,* 2002, Robertson and Gjerde, 2007). These methods include the inclusion or exclusion of vital dyes by the oocysts, such as 4'-6-diamidino-2-phenylindole (DAPI) and propidium iodide (PI) (Schupp and Erlandsen, 1987). Sporozoite nuclei, which take up DAPI but fail to stain with PI, are viable, while nuclear material that stains with both fluorochromes is not viable. Other nucleic acid stains (SYTO-9, hexidium) were also used to access oocyst viability (Belosevic *et al.,* 1997). Another method to assess oocyst viability is the *in vitro* excystation method (Campbell *et al.,* 1992). Nucleic acid staining and *in vitro* excystation methods are used in many laboratories and have similar performances, but they tend to overestimate oocyst viability (Black *et al.,* 1996). In addition, RT-PCR has been used to assess oocyst viability by reverse transcription of mRNA targeting several markers (heat shock protein, β-tubulin, amyloglucosidase) (Stinear *et al.,* 1996, Widmer *et al.,* 1999, Jenkins *et al.,* 2000). Other methods allowing viability studies are cell culture (Slifko *et al.,* 1997, Rochelle *et al.,* 1997, Di Giovanni *et al.,* 1999), mouse infection (Korich *et al.,* 1990) and fluorescent *in situ* hybridization (FISH) (Vesey *et al.,* 1998). Among the methods described, animal infectivity and cell culture are the most reliable; animal infectivity being the gold-standard and cell culture frequently considered as providing equivalent results (Robertson and Gjerde, 2007).

It is worth bearing in mind that the presence of oocysts in the water supply, whether alive or dead, still represents an undeniable risk of human infection to the recipient community. In addition, whenever *Cryptosporidium* oocysts are detected, an investigation of the source and means of water contamination should be carried out to prevent future contamination of the water supply.

1.10 Molecular characterization of *Cryptosporidium* species

As discussed above, the detection of *Cryptosporidium* is important for diagnosis, epidemiological surveys, disease surveillance and drinking water monitoring. The identification of *Cryptosporidium* isolates at the species level enabled an improved understanding of the public health importance and revealed differences in the transmission and epidemiology of human infective *Cryptosporidium* species (Xiao and Ryan, 2004). In addition, the characterization of *Cryptosporidium* isolates to the subtype level is desirable as it provides information about the population structure and the pathogenicity of isolates and allows tracking of contamination sources during outbreaks (Chalmers *et al.,* 2008, Hunter *et al.,* 2007).

1.10.1 Genotyping

The vast majority of the molecular methods currently used to differentiate between *Cryptosporidium* species and/or genotypes are based on PCR, followed by Restriction Fragment Length Polymorphism (RFLP) (Jiang and Xiao, 2003). PCR-RFLP of the COWP gene differentiates *C. parvum*, *C hominis* and *C. wrairi* (Spano *et al.,* 1997), PCR-RFLP of DHFR gene distinguishes *C. parvum* and *C. hominis* (Gibbons *et al.,* 1998) and PCR-RFLP of the SSU rRNA discriminates *C. parvum*, *C. muris*, and *C. baileyi* (Leng *et al.,* 1996). PCR-RFLP is a popular genotyping technique due to the ease of use in modestly equipped laboratories. However, it can be difficult to interpret PCR-RFLP results when the isolate gives an unusual pattern. Sulaiman and colleagues (1999) reviewed the sensitivity and specificity of 11 PCR-RFLP techniques for *Cryptosporidium* genotyping. The genotyping techniques targeted SSU rRNA (Awad-El-Kariem *et al.,* 1994, Leng *et al.,* 1996, Xiao *et al.,* 1999), ITS-1 (Carraway *et al.,* 1996), COWP (Spano *et al.,* 1997), DHFR (Gibbons *et al.,* 1998), Poly-T (Carraway *et al.,* 1997), TRAP-C1 (Spano *et al.,* 1998), TRAP-C2 (Sulaiman *et al.,* 1998) and two undefined genomic regions (Bonnin *et al.,* 1996, Morgan *et al.,* 1997). PCR-RFLP was used to test

DNA of *C. parvum*, *C. hominis, C. muris*, and *C. serpentis*. The authors reported that SSU rRNA and DHFR genes are the most accurate targets for *Cryptosporidium* genotyping (Sulaiman *et al.,* 1999).

Jiang and Xiao (2003) compared the performance of ten commonly used genotyping techniques for the detection and differentiation of seven human-pathogenic *Cryptosporidium* spp. (*C. hominis*, *C. parvum*, *C. meleagridis*, *C. felis*, *C. canis*, *C. muris* and *Cryptosporidium* pig genotype I). The evaluated methods were based on the amplification and RFLP and/or sequencing of the following genes: COWP (Pedraza-Diaz *et al.,* 2001, Homan *et al.,* 1999), HSP70 (Gobet and Toze, 2001), TRAP-C1 (Spano *et al.,* 1998), TRAP-C2 (Elwin *et al.,* 2001), DHFR (Gibbons *et al.,* 1998), SSU rRNA (Sturbaum *et al.,*2001, Ward *et al.,* 2002, Xiao *et al.,* 2001 b) and an undefined genomic region (Guyot *et al.,* 2002). The authors found that all three SSU rRNA techniques amplified DNA from all *Cryptosporidium* species tested, but only two techniques (Xiao *et al.,* 2001 b, Ward *et al.,* 2000) allowed the discrimination of all species. The SSU rRNA gene is the only genetic locus so far characterized, which is able to detect and differentiate between all human pathogenic *Cryptosporidium* species.

1.10.2 Subtyping

Subtyping methods have been developed to examine the relationships between *Cryptosporidium* isolates. Several genetic loci were investigated:

1.10.2.1 The GP60 locus

The GP60 locus is currently the most commonly used genetic locus for *Cryptosporidium* subtyping (GP60, synonymous with gp15 or gp15/45/60) (Alves *et al.,* 2003 a, Peng *et al.,* 2003, Peng *et al.,* 2003 b, Zhou *et al.,* 2003, Wu *et al.,* 2003, Glaberman *et al.,* 2002, Leav *et al.,* 2002, Sulaiman *et al.,* 2001, Peng *et al.,* 2001, Strong *et al.,* 2000).

The GP60 locus encodes a 60 kDa surface-expressed glycoprotein. The 60 kDa precursor protein is cleaved into two subunits, gp15 and gp45 by a furin-like protease activity that is calcium-dependent (Wanyiri *et al.,* 2007). Gp15 and gp45 are localised in the apical region of sporozoites and merozoites and are present in the protein trails of gliding zoites, implying a functional role in parasite attachment, invasion and motility (Power *et al.,* 2009, O'Connor *et al.,* 2007, O'connor *et al.,* 2007 b, Wanyiri and Ward, 2006, Alves *et al.,* 2003 a, Sestak *et al.,* 2002, Cevallos *et al.,* 2000, Priest *et al.,* 2000). A monoclonal antibody against gp15 neutralized infectivity *in vitro* and passively protected against the disease *in vivo* (Strong *et al.,* 2000, Gut and Nelson, 1994). GP60 N and C terminal peptides include the hypothetical signal sequence and the GPI anchor attachment site are highly conserved among all *Cryptosporidium* isolates, suggesting that both features are important (Strong et a., 2000). The rest of the gene has a high degree of polymorphism, which is far greater than any other *Cryptosporidium* protein coding locus examined to date (Leav *et al.,* 2002, Sulaiman *et al.,* 2001, Strong *et al.,* 2000). GP60 locus contains a homopolymer serine tract encoded by a trinucleotide repeat, which is often hypervariable and manifests numerous single-nucleotide and single-amino-acid polymorphisms (SNPs and SAAPs) (Strong *et al.,* 2000).

A sub-genotyping technique based on PCR and sequencing of GP60 was developed and used to examine population structure and transmission dynamics of *C. parvum* and *C. hominis* (Strong *et al.,* 2000, Peng *et al.,* 2001, Peng *et al.,* 2003, Alves *et al.,* 2003 a, Alves *et al.,* 2006, Xiao and Ryan, 2004, Sulaiman *et al.,* 2005, Mallon *et al.,* 2003, Feltus *et al.,* 2006, Trotz-Williams *et al.,* 2006, Thompson *et al.,* 2007). GP60 sequencing allowed the classification of isolates into subtype families defined by the number of trinucleotide repeats (TCA/TCG) coding for the amino acid serine according to the nomenclature described by Sulaiman and colleagues (2005). For example, in the allele family IIa, one subtype had 15 copies of the TCA repeat and 1 copy of the TCG repeat, whereas the other subtype had 15 copies of the TCA repeat and two copies of the TCG repeat.

Therefore, the two subtypes were designated IIaA15G1R1 and IIaA15G2R1, respectively. In the subtype name IIaA15G1R1, IIa indicates that the subtype belongs to allele family IIa, A15 indicates that the subtype has 15 copies of the TCA repeat, and G1 indicates that the subtype has one copy of the TCG repeat. Because some subtypes have one copy of the sequence ACATCA immediately after the trinucleotide repeats whereas others have two copies of the sequence, R1 and R2 are used to differentiate these two types (Sulaiman *et al.,* 2005).

GP60 sequencing allowed the identification of several subtype families and numerous subtypes within each family. *C hominis* subtype families were called Ia, Ib, Ic, Id, Ie, If and *C. parvum* families were called IIa, IIb, IIc, IId, IIe. The subtype IbA10G2 seems to be the most common *C. hominis* subtype in UK, US, Peru, Australia and South Africa, while IIaA15G2R1 is the predominant *C. parvum* subtype in Europe, US, Canada and Australia (O'Brien *et al.,* 2007). Subtyping of *C. parvum* isolates showed that IIa and IId families were reported from humans and animals while IIb and IIc families were only found in humans (Alves *et al.,* 2006). Subtyping of a great number of isolates from different geographic locations never reported the presence of the IIc family in any animal species (Leav *et al.* 2002, Xiao and Ryan 2004, Xiao *et al.,* 2004). Thus, the IIc subtype family is considered anthroponotic (Alves *et al.,* 2006). The case control study conducted by Hunter and colleagues (2007) confirmed these findings. Interestingly, some GP60 subtypes were associated with strain virulence. Cama and colleagues (2007) reported differences in clinical manifestations (association with diarrhoea and/or vomiting) among *Cryptosporidium* species and subtypes. In addition to GP60 sequencing, PCR-RFLP and PCR-RFLP-SSCP (Single Strand Conformation Polymorphism) of GP60 have also been used for *Cryptosporidium* subtyping (Wu *et al.,* 2003).

Despite the extensive use of the GP60 marker, subsequent studies showed that the level of discrimination is not in accordance with multi-locus subtyping and GP60 genotype by itself is difficult to reconcile with the concept of subtype

defined as a genetically distinct population within a species (Widmer, 2009). Thus, Widmer stated that GP60 may not a reliable marker of *C. parvum* and *C. hominis* population structure. In addition, Hunter and colleagues (2008) stated that it is still unclear if the identification of different subtypes in outbreak settings represented different lineages or evolution of strains during the outbreaks. The ambiguities surrounding GP60 subtyping need to be clarified in the near future for an improved understanding of the epidemiology and population structure of *Cryptosporidium* isolates.

1.10.2.2 The rDNA Unit

The rDNA unit has been used for *Cryptosporidium* subtyping. The rDNA unit consists of five sequential loci: 5' small subunit rRNA (SSU rRNA, synonymous with 18S rRNA), internal transcribed spacer 1 (ITS1), 5.8S rRNA, internal transcribed spacer 2 (ITS2), and large subunit rRNA (LSU rRNA) 3' (Robinson, 2005). The rDNA unit in *C. parvum* has been described by Le Blancq and colleagues (1997). It has five copies per genome, shared over at least three chromosomes. The sequence analysis of the ITS1-5.8S rRNA-ITS2 region did not have sufficient discriminating power and only allowed genotyping of *Cryptosporidium* isolates (Morgan *et al.,* 1999). Sequence analysis of the SSU showed higher level of sequence divergence, but the subtyping potential of this locus for epidemiological purposes is questionable due to the limited variation within *C. hominis* and *C. parvum* isolates and to the intra-isolate heterogeneity attributed to the multiple copy number of the gene (Carraway *et al.,* 1994, Glaberman *et al.,* 2001, Sulaiman *et al.,* 2001, Gibbons-Matthews and Prescott, 2003).

1.10.2.3 Microsatellite and Minisatellite loci

Mini and microsatellites have been extensively used for *Cryptosporidium* subtyping, with 22 microsatellites and three minisatellites (MS1, MS5, MS12)

targets being described (Robinson *et al.,* 2005). ML1 was the most widely used microsatellite marker and allowed identification of two *C. hominis* alleles and five *C. parvum* alleles (Mallon *et al.,* 2003 b, Enemark *et al.,* 2002, Huetink *et al.,* 2001, Caccio *et al.,* 2000). In addition, sequence and fragment size analysis of the ML2 microsatellite has also been described (Caccio *et al.,* 2001, Alves *et al.,* 2003 b) and allowed identification of nine alleles (one *C. hominis* and eight *C. parvum*). ML2 seems to provide a higher level of discrimination than ML1 (Caccio *et al.,* 2001). Feng and colleagues (2000) described a method based on fragment size and sequence analysis of nine different microsatellite loci (4E12, Cp273, 12C07, 2G04, 6B03, 5B12, 1G09, 1F07, 7E1C) and Widmer and colleagues (2000) used 14 microsatellite loci for *C. hominis* and *C. parvum* subtyping and showed high level of discrimination.

1.10.2.4 The HSP70, the β-tubulin, TRAP-C2 and Poly-T loci

Several studies used the HSP70 locus for *Cryptosporidium* subtyping from clinical sources and environmental waters (LeChevalier *et al.,* 2003, Sulaiman *et al.,* 2001, Di Giovanni *et al.,* 1999, Peng *et al.,* 2003 b). The HSP70 locus was found to have several nucleotide substitutions and allowed the identification of high number of subtypes from a relatively few isolates.

Other genetic loci were also used for subtyping purposes, including β-tubulin, TRAP-C2 and Poly-T. However, they showed low levels of discrimination (Sulaiman *et al.,* 2001, Widmer *et al.,* 1998, Rochelle *et al.,* 2000).

1.10.2.5 Multi-locus subtyping

In addition to using an individual locus for subtyping, a combination of several loci was used to provide greater discriminatory power. Several investigations used multi-locus subtyping and showed the importance of multi-locus fingerprinting (Mallon *et al.,* 2003, Mallon *et al.,* 2003 b, Feng *et al.,* 2002, Caccio

et al., 2001, Widmer *et al.,* 2000). Sulaiman and colleagues (2001) used sequence analysis of one microsatellite locus (MS1) and the TRAP-C2, Poly-T, SSU rRNA, HSP70 and GP60 genes to increase the discriminatory power when subtyping 62 *C. hominis* isolates. Similarly, Glaberman and colleagues (2001) used DNA sequences of the SSU rRNA, GP60 and HSP70 loci on 11 *C. meleagridis* samples and showed higher discrimination using the multi-locus approach. Mallon and colleagues (2003, 2003 b) described a multi-locus approach using four microsatellite (TP14, MS9, ML1 and GP15) and three minisatellite (MS5, MS12 and MS1) loci to subtype human and animal isolates of *Cryptosporidium* using sequence and fragment size analysis. A total of 58 multi-locus subtype fingerprints was identified, which included seven *C. hominis* subtypes, 48 *C. parvum* subtypes, two *C. parvum* monkey genotype subtypes and one *C. meleagridis* subtype. The multi-locus approach has shown its usefulness in increasing the amount of information that can be obtained from each isolate and in turn has provided a higher level of discrimination between isolates.

Mallon and colleagues (2003) reported that *C. hominis* isolates are primarily of closely related multi-locus genotypes, suggesting a clonal population, while, *C. parvum* has a panmictic population structure. Subsequently, Peng and colleagues (2003 b) reported that in *C. hominis* isolates from Malawi, linkage disequilibrium analysis of HSP70 and GP60 loci showed possible intraspecific recombination. This result suggests that *C. hominis* can also have a panmictic population structure. Hunter and colleagues (2007) used three microsatellite loci (ML1, ML2, and GP60) to subtype *Cryptosporidium* clinical isolates and showed good discriminatory power. In addition, the authors reported low diversity of *C. hominis* strains as previously reported (Mallon *et al.,* 2003, 2003 b). This is supported by the theory of adaptive polymorphism, predicting that species occupying broad ecological niches are likely to be more diverse on the genotypic and phenotypic level when compared to species with a narrow host range (Hunter and Fraser, 1990). Several investigators currently acknowledge that the population structure of *C. hominis* and *C. parvum* is more complex than

previously suggested and that both clonal and panmictic population structure are possible, depending on the geographic distribution (Beck *et al.,* 2009, Jex *et al.,* 2008, Morrison *et al.,* 2008, Tanriverdi *et al.,* 2008, Tait *et al.,* 2004, Xiao and Ryan, 2004).

The ideal number of genetic loci that enables optimal degree of discrimination between *Cryptosporidium* species and subtypes is still ambiguous. Similarly, the identification of genetic loci that would accurately represent *Cryptosporidium* species population structure is still needed. In the actual situation, the identification of novel genetic loci is desirable and should be facilitated by the vast amount of genetic data generated from genome projects.

1.11 Virulence

1.11.1 Virulence factors and virulence determinants

Virulence is defined as the ability of a microorganism to cause disease (Alonso-Monge *et al.,* 2003, Poulin and Combes, 1999, Ebert and Herre, 1996, Garnick, 1992, Bloch, 1950, Pike and Mackenzie, 1940, Mellon, 1926, Smiley and Pearse, 1926). Virulence and pathogenicity are often used interchangeably, but virulence may also be used to indicate the degree of pathogenicity. Virulence is commonly used to define the likelihood of an infected person getting an illness and, if so, to express the severity of the symptoms. Therefore, virulence could be considered as an increased risk of infection and/or an increased severity of illness. Woolhouse and colleagues (2002) defined virulence as “the direct or indirect reduction in host fitness attributable to pathogen infection, often measured as pathogen induced host mortality”. Pathogen-centred views of virulence assert that pathogens are distinguished from non-pathogens by their expression of virulence factors. Although this concept appears to apply to some microbes that cause disease in normal hosts, it does not apply to most microbes that cause disease primarily in immunocompromised hosts (Casadevall and Pirofski, 2001).

The authors acknowledged that virulence, despite being a microbial characteristic, can only be expressed in a susceptible host and depends on the context and nature of host-microbe interaction. They therefore considered that virulence is directly linked to host damage (Casadevall and Pirofski, 2001).

The view that virulence is a single characteristic is difficult to reconcile with the fact that host-pathogen interaction is continuous and subject to changes on the basis of host, microbial and exogenous factors (Casadevall and Pirofski, 2001). In fact, virulence is often multi-factorial, involving a complex interplay between the parasite and the host. Various host factors, including age, sex and the status of the immune system affect the outcome of the host-parasite interaction (Okhuysen and Chappell, 2002). In addition, the genotypic and phenotypic characteristics of the parasite define intrinsic diversity in isolate pathogenicity and virulence (Fayer *et al.,* 2009). Therefore, an integrated view of microbial pathogenesis and virulence accounting for the contribution of both host and pathogen factors is considered more accurate (Woolhouse *et al.,* 2002, Casadevall and Pirofski, 1999).

Each of the microbiological attributes that contribute to virulence can in general be linked to specific structural elements or biochemical compounds within the organism, which are generally termed virulence factors. Although the terms “virulence determinants” and “virulence factors” are widely used to describe traits contributing to pathogenicity, a subtle distinction exists between the two terms. Virulence factors are microbial traits that promote host damage (Poulin and Combes, 1999) and more precisely, *a virulence factor is a gene product necessary but not sufficient to cause disease*. While, virulence determinants are the factors present in a microorganism that are responsible for the relative capacity of a parasite to cause damage in a host (Okhuysen and Chappell, 2002). A more precise definition of a virulence determinant could be *a gene enabling an organism to colonize the host successfully and which may then result in host pathology*. An operational definition of a virulence determinant has been

proposed as a gene belonging to a pathogen whose inactivation or deletion leads to a decrease in virulence of the pathogen and whose genetic reintroduction restores virulence (Alonso-Monge *et al.,* 2003).

On a practical note, Edberg (2009) stated "For a microbe to generate disease, a number of sequential virulence factors must be active. While clearly the genes that code for virulence must be present in the microbe, disease generation is a phenotypic phenomenon". This clearly suggests that the presence of a virulence gene does not mean that it will be active and that the organism will be virulent. This further complicates the picture and should be taken into account when investigating the presence of virulence genes in different species and isolates. To support his theory, Edberg (2009) gives several examples, particularly, *Cryptosporidium* spp., for which several Nucleotide Excision Repair (NER) genes were described (Rochelle *et al.,* 2004), however, these genes do not seem to be active as UV was shown to be a reliable disinfectant for *Cryptosporidium* causing irreversible DNA damage.

When considering how each individual virulence factor contributes to an overall virulence phenotype, it is crucial to identify a measure of virulence (McClelland *et al.,* 2006, Alonso-Monge *et al.,* 2003, Poulin and Combes, 1999). The ability of a microbe to cause disease in an animal model, which is central to Koch's postulate, has been the cornerstone of the measurement of virulence, but this relies on the availability of a susceptible experimental animal model (Casadevall and Pirofski, 2001). Some commonly used measures of virulence are mortality, microbial burden on tissue, lifetime reproductive success of infected hosts versus uninfected hosts (McClelland *et al.,* 2006, Alonso-Monge *et al.,* 2003, Poulin and Combes, 1999). Additional measurements of virulence could include different measures of the host damage and the immune response.

1.11.2 *Cryptosporidium* virulence factors

Several studies have tried to determine the factors responsible for the initiation, establishment and perpetuation of *Cryptosporidium* infection. *Cryptosporidium* is a relatively non-invasive parasite that establishes itself in a membrane-bound compartment on the apical surface of the intestinal epithelium (Okhuysen and Chappell, 2002). Nevertheless, it causes significant abnormalities in the absorptive and secretory functions of the gut. This damage could be the result of direct injury to the host epithelial cell or indirectly through the effect of inflammatory cells and cytokines recruited to the site of infection (Okhuysen and Chappell, 2002).

For *Cryptosporidium*, if virulence factors are considered to be the processes and substances by which the parasite initiates and maintains disease in the host, these factors can affect the host at any time during the life cycle from the time the parasite enters the body until it is killed or completes the cycle and leaves (Fayer *et al.,* 2009). To date *Cryptosporidium* specific virulence factors have not been characterised to the point of unequivocally establishing their roles in causing damage or proving that deletions result in a decrease of virulence (Okhuysen and Chappell, 2002). This is mainly due to the fact that, unlike other apicomplexan parasites such as *Toxoplasma* and *Plasmodium* spp., no transient or stable transfection systems have been developed for *Cryptosporidium*. Therefore, genes cannot be knocked out or knocked down, thus making it impossible to conclusively demonstrate the function of the protein encoded by these genes (Wanyiri and Ward, 2006).

Putative virulence factors for *Cryptosporidium* have been identified as genes involved in the initial interaction processes of *Cryptosporidium* oocysts and sporozoites with host epithelial cells including excystation, gliding motility, attachment, invasion, parasitophorous vacuole formation, intracellular maintenance and host cell damage (Fayer *et al.,* 2009, Wanyiri and Ward, 2006):

1.11.2.1 Adherence factors

A critical initial step in establishing infection is parasite attachment to host cells. Two classes of proteins namely mucin-like glycoproteins and thrombospondin-related adhesive proteins have been characterized (Wanyiri and Ward, 2006) and showed to mediate adhesion as summarized in Table 1.3:

- CSL (circumsporozoite-like glycoprotein) of ~ 1300 kDa was described by Riggs and colleagues (1997) and is associated with the apical complex of sporozoites and merozoites (Schaefer *et al.,* 2000, Langer and Riggs, 1999). CSL is released as a soluble glycoprotein and contains a ligand that binds specifically to a receptor on the surface of human and bovine intestinal epithelial cells (Langer and Riggs, 1999). The zoite ligand was shown to be involved in attachment and invasion (Langer *et al.,* 2001). Monoclonal antibodies to CSL elicited changes in sporozoites and merozoites, similar to the malarial circumsporozoite precipitate (CSP) reaction and caused complete neutralization of sporozoite infectivity (Riggs *et al.,* 1997).

- Gp900 is a large glycoprotein identified by immunoprecipitation of sporozoite extracts with hyperimmune bovine colostrum (Petersen *et al.,* 1992 b). This large mucin-like glycoprotein is located in micronemes and at the surface of invasive merozoites and sporozoites. Gp900 is deposited in trails during gliding motility and is known to mediate invasion (Bonnin *et al.,* 2001, Barnes *et al.,* 1998). The deduced aminoacid sequence of Gp900 has a signal peptide and a transmembrane domain (Barnes *et al.,* 1998). Specific antibodies to gp900 can competitively inhibit infection *in vitro* (Barnes *et al.,* 1998, Petersen *et al.,* 1997).

- Sporozoite and merozoite cell surface protein: gp15/40/60 complex (Cpgp40/15): Strong and colleagues (2000) reported that gp15/40/60

mRNA is translated into a ~ 60 kDa glycoprotein precursor during the intracellular stages of *C. parvum* life cycle. This precursor is proteolytically processed shortly after synthesis to generate 15 and 45 kDa glycoproteins. Independently, Cevallos and colleagues (2000) cloned and sequenced the same gene from *C. parvum* genomic DNA and called it Cpgp40/15. Gp40 is localized at the surface and apical region of the parasite and is shed from the surface, while gp15 is on the surface of sporozoites and is shed in trails during gliding movement (Boulter-Bitzer *et al.,* 2007, Cevallos *et al.*, 2000). In *C. parvum*, gp15 is attached to the membrane via a glycophosphatidylinositol (GPI) anchor (Priest *et al.,* 2001). Both gp40 and gp15 display O-linked-α-N-acetylgalactosamine (α-GalNAc), which are thought to be involved in invasion and attachment since lectins that recognize these determinants block sporozoite attachment (Cevallos *et al.,* 2000 b, Winter *et al.,* 2000, Gut and Nelson, 1994).

- P23 is a 23 kDa sporozoite surface protein that is antigenically conserved across geographically diverse isolates (Perryman *et al.,* 1996) and is deposited in trails during the initial stages of the infection (Arrowood *et al.,* 1991). P23 elicits antibody response in animals and humans exposed to *C. parvum* (Reperant *et al.,* 1994, Riggs *et al.,* 1994, Tilley *et al.,* 1993, Mead *et al.,* 1988). P23 has neutralization-sensitive epitopes and monoclonal antibodies were found to significantly reduce infection in mice and protect calves against cryptosporidiosis (Boulter-Bitzer *et al.,* 2007).

Cpgp40/15 and P23 were successfully expressed in the related apicomplexa *Toxoplasma gondii* as recombinant proteins with post-translational modifications similar to those of the native protein, which would be impossible using bacterial expression systems (O'Connor *et al.,* 2003, Shirafuji *et al.,* 2005). The recombinant proteins showed

appropriate localization and glycosylation. The expression of *Cryptosporidium* proteins in this heterologous expression system is a major advance and would assist an improved understanding of the functional role of several *Cryptosporidium* proteins.

- TRAP-C1 (thrombospondin-related adhesive protein *Cryptosporidium* 1) is a 76 kDa protein localized on the apical pole of sporozoites (Spano *et al.,* 1998 b). It showed sequence and structural homology to members of the thrombospondin family adhesive proteins in other apicomplexan parasites (*Plasmodium* spp., *Toxoplasma gondii*, *Eimeria tenella* and *Neospora* spp. (Boulter-Bitzer *et al.,* 2007). TRAP and structurally related proteins are involved in parasite gliding motility and cell penetration (Boulter-Bitzer *et al.,* 2007). Putignani and colleagues (2008) characterized "CpTSP8", one of the 12 *C. parvum* thrombospondin-related proteins (CpTSP2- CpTSP12) family identified by bioinformatic tools. The authors showed that CpTSP8 is located at the apical complex of sporozoites and merozoites and is translocated onto the parasite surface, as it is typical of micronemal proteins (MICs). Therefore, CpTSP8 was renamed CpMIC1. MIC proteins have been shown to be essential in host-cell attachment/invasion and gliding motility (Soldati *et al.,* 2001).

GP900, gp40, gp15, Cpa135, Cp2, P23 and TRAP-C1 (Table 1.3) have or are predicted to have mucin-type *O* –glycosylation, suggesting that this type of post-translational modification is common in proteins involved in attachment and invasion (Wanyiri and Ward, 2006).

Virulence factor gene/protein	**Putative function**	**Reference**
Serine protease	Excystation	Forney *et al.,* 1996a, 1996b
Aminopeptidase	Excystation	Okhuysen *et al.,* 1994
Circumsporozoite like Glycoprotein (CSL)	Adhesion	Riggs *et al.,* 1997
Glycoprotein 900 (gp900)	Adhesion	Petersen *et al.,* 1992 b
GP60/40/15	Adhesion	Cevallos *et al.,* 2000, 2000 b, Strong *et al.,* 2000
P23	Adhesion, locomotion	Arrowood *et al.,* 1991 Perryman *et al.,* 1996
TRAP-C1	Adhesion, locomotion	Spano *et al.,* 1998
Cp47	Adhesion	Nesterenko *et al.,* 1999
CPS-500	Adhesion, locomotion	Riggs *et al.,* 1989 Bjorneby *et al.,* 1990
Cp2	Invasion, membrane integrity	O'Hara *et al.,* 2004
Cpa135	Invasion	Tosini *et al.,* 2004
Secretory phospholipase	Invasion, intracellular establishment	Pollok *et al.,* 2003
Hemolysin H4	Membrane lysis	Steele *et al.,* 1995
Cryptosporidium ATP binding cassette (CpABC)	Transport, nutrient transport	Perkins *et al.,* 1999
C. parvum ATPase2 (CpATPase2)	Biomembrane heavy metal transporter	LaGier *et al.,* 2001
C. parvum ATPase3 (CpATPase3)	Biomembrane ion or phospholipid transporter	LaGier *et al.,* 2002
HSP70	Stress protection	Khramtsov *et al.,* 1995
HSP90	Stress protection	Woods *et al.,* 1999
Type I polyketide synthase (CpPKS1)	Unknown function	Zhu *et al.,* 2002
Cysteine protease	Immune/cytokine modulation	Nesterenko *et al.,* 1995
Acetyl co synthetase	Fatty acid metabolism	Camero *et al.,* 2003
C. parvum Subtiline like serine protease (CpSUB)	Invasion	Wanyiri *et al.,* 2009
C. parvum micronemal protein (CpMIC1)	Adhesion, locomotion	Putignani *et al.,* 2008
C. parvum Mucin like (CpMuc)	Invasion	O'Connor *et al.,* 2009

Table 1.3: *Cryptosporidium* putative virulence factors identified by immunological and molecular methods. Table compiled from data obtained from the following publications: Fayer *et al*, 2009, Wanyiri *et al.,* 2009, O'Connor *et al.,* 2009, Putignani *et al.,* 2008, Boulter-Bitzer *et al.,* 2007, Okhuysen and Chappell, 2002.

Because mucin-like proteins were shown to be important for *Cryptosporidium* host-parasite interactions, O'Connor and colleagues (2009) undertook data mining of the *Cryptosporidium* genome databases to identify other mucin-like genes. They discovered a single locus of seven small mucin sequences (CpMuc1-7), which were expressed throughout the intracellular development stages. Specific antibodies inhibited infection *in vitro*, which is consistent with a role in host-cell invasion.

Proteolytic processing of surface and apical complex proteins by parasite proteases has be shown to be required for invasion of host cells and for egress from them. Further effort was focused on identification of *Cryptosporidium* proteases. Wanyiri and colleagues (2009) characterized a *C. parvum* subtiline like serine protease (CpSUB) and showed that this protein is likely to be responsible for the processing of gp40/15 (Table 1.3).

As detailed above, *Cryptosporidium* has several distinct molecules to mediate attachment to and invasion of the epithelial intestinal cells. These proteins have features in common with other apicomplexans proteins implicated in mediating host-cell interaction (Tzipori and Ward, 2002). In addition, proteins and glycoproteins expressed on the surface of the invasive *C. parvum* sporozoite and merozoite stages and shed in trails by gliding zoites are thought to play essential roles in parasite motility and in parasite attachment to and invasion of host epithelial cells (Strong *et al.,* 2000). The relative contribution of each individual molecule remains to be determined. It is likely that by using a large number of seemingly redundant adhesive molecules, the parasite can maximize the opportunity for cell attachment across a broad range of potential hosts. It is also possible that quantitative or qualitative differences in these glycoproteins may confer selectivity for host attachment (Okhuysen and Chappell, 2002).

1.11.2.2 Cellular damage

Cell damage in enterocyte monolayers has been documented through disruption of tight cell junctions, loss of barrier function, release of lactate dehydrogenase and increased cell death (Adams *et al.,* 1994). The mechanisms causing cellular damage during *Cryptosporidium* infection remain unknown, however, several molecules can cause direct tissue damage such as phospholipases, proteases and haemolysins (Okhuysen and Chappell, 2002).

Proteases have been described to have important functions in a parasite's life cycle, such as mediating protein degradation, invasion of host tissues and evasion of host immunity (Shenai *et al.,* 2000, Que and Reed, 2000, Shin *et al.,* 2001). Distinct types of protease activities have been identified in *Cryptosporidium* sporozoites such as aminopeptidase, cysteine protease and serine protease implicated in the excystation process (Forney *et al.,* 1996 a, Okhuysen *et al.,* 1996, Nesterenko *et al.,* 1995). The identification of functional proteases in sporozoites during excystation and the prevention of infection in the presence of protease inhibitors suggest that proteases are important in the initial stages of *Cryptosporidium* infection (Okhuysen and Chappell, 2002).

A haemolysin H4 has been identified by screening a *C. parvum* expression library on sheep blood agar (Steele *et al.,* 1995). H4 has sequence similarity to the haemolysin of enterohemorrhagic *E. coli* 0157 H7. The function of H4 is unknown but its ability to disrupt cell membranes suggest a role in cellular invasion and/or the disruption of vacuolar membranes would allow merozoites to exit the parasitophorous vacuole and spread to adjacent cells (Okhuysen and Chappell, 2002). Another *Cryptosporidium* protein of interest is an ABC transporter gene (CpABC) localized in the electron dense feeding organelle of the parasitophorous vacuole (Perkins *et al.,* 1999). Interestingly, these genetic elements share structural similarities with bacterial genes, which are critical in producing secretory diarrhoea (Okhuysen and Chappell, 2002).

1.11.2.3 Heat shock proteins

Heat shock proteins (HSPs) are a family of large conserved proteins. They are usually defined by their apparent molecular weight in sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), with HSP90, HSP70 and HSP65 being common families (Miller *et al.,* 2000). Synthesis of HSPs, especially HSP70, increases dramatically under stressful conditions (sudden shifts in temperature, decreased availability of nutrients, immune attack). HSPs function as intra-cellular chaperones for other proteins, which play an important role in protein-protein interactions and facilitate transport, folding, assembly, biosynthesis and secretion of newly formed proteins (Okhuysen and Chappell, 2002).

In *Cryptosporidium* two HSPs (HSP70 and HSP90) have been described (Woods *et al.,* 1999, Khramtsov *et al.,* 1995). Considerable polymorphism in the HSP70 gene has been identified and was used for genotyping purposes (Morgan *et al.,* 2000, Sulaiman *et al.,* 2000). However, HSPs are under selective pressure and their high degree of polymorphism might not reflect the genetic relationships between isolates or subtypes. In the closely related apicomplexa *T. gondii*, it has been demonstrated that quantitative and qualitative differences in the HSP expression are directly related to parasite virulence (Miller *et al.,* 2000, Lyons and Johnson, 1998). In fact, high levels of expression of HSP70 were detected in virulent strains grown in mice, but only little expression of HSP70 was observed in avirulent strains (Lyons and Johnson, 1995). The relationship between the level of HSP expression and *Cryptosporidium* virulence warrants further investigation.

1.11.3 Host factors and *Cryptosporidium* virulence

Age is a major factor in the epidemiology of infection, with children being most at risk of cryptosporidiosis as shown by the cumulative data of cryptosporidiosis cases detected between 1989 and 2008 (Figure 1.2). The illness in children is not necessarily more or less severe than in older age groups and infection may even

be sub-clinical. Indeed asymptomatic carriage in children has been suggested as being an important reservoir for *C. hominis* infection (Hunter *et al.* 2004 b). As discussed above, the high incidence of cryptosporidiosis in children probably reflects lack of immunity due to few prior exposures (Yoder and Beach, 2010, Hunter and Thompson, 2005, Guerrant, 1997).

Perhaps not surprising given that many of the earliest identified cases of human cryptosporidiosis were in immune suppressed individuals (Current and Garcia, 1991), immune suppression is an important human specific risk factor for increased virulence. The severity of illness may be much greater in immune suppressed compared to immune compromised patients (Chalmers and Davis, 2010, Hunter and Nichols, 2002, Chen *et al.,* 2002, O'Donoghue, 1995, Guerrant, 1997). In such subjects, cryptosporidiosis is no longer self limiting and can be life-threatening (Chen *et al.,* 2002, O'Donoghue, 1995, Junarek, 1995). However, not all forms of immune suppression lead to increased disease severity of cryptosporidiosis. The main risk seems to be immune suppressive disorders that impact on T cell function, the most obvious being HIV/AIDS (Hunter and Nichols, 2002). In AIDS patients, the most severe disease occurs in people with a CD4 count of less than 50. There is also evidence that patients with solid organ transplantation and with other malignancies may also be at increased risk, though the weight of evidence is less strong probably reflecting less common infections in these groups (Hunter and Nichols 2002).

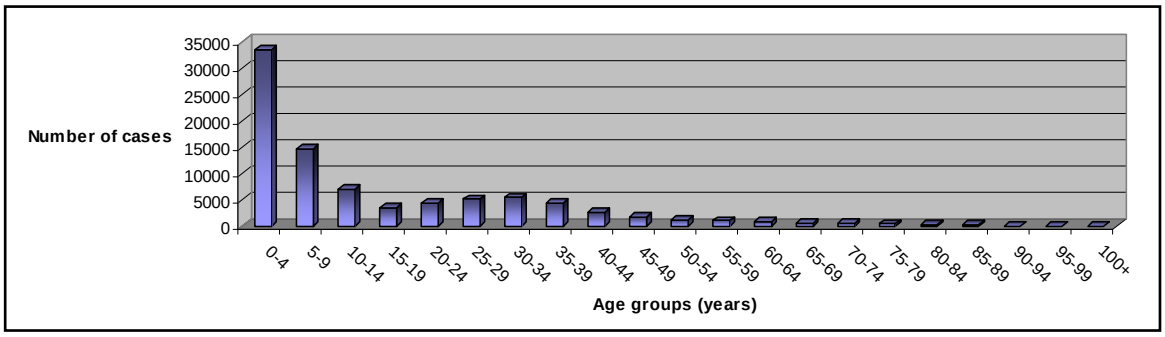


Figure 1.2: Total number of cryptosporidiosis cases detected in England and Wales by age group between 1989 and 2008 as reported by the Health protection Agency (www.hpa.org.uk).

Of particular relevance in developing country settings is the issue of malnutrition in children. There is good evidence from animal studies that cryptosporidiosis exacerbates malnutrition and that cryptosporidiosis is more severe in malnourished mice (Coutinho *et al.,* 2008). In human situations, the evidence also supports this synergistic relationship (Mor and Tzipori 2008, Hunter and Nichols 2002). However, the exact mechanisms behind the increased severity of illness in the malnourished child are not yet fully clear. One explanation is that in the malnourished child, the damage to the gut mucosa with a new *Cryptosporidium* infection would lead to further impaired nutrient absorption and stunting that would not be such an issue in the well nourished child (Mor and Tzipori 2008).

1.11.4 *Cryptosporidium* species-specific virulence factors

Cryptosporidium putative virulence factors detailed in 1.11.2, which mediate sporozoite attachment and invasion of the epithelial intestinal cells and which inactivation can prevent infection *in vitro* and *in vivo*, are common to all *Cryptosporidium* spp. These virulence factors were mostly identified in *C. parvum*. Among 13 valid *Cryptosporidium* species, only eight infect human (*C. hominis*, *C. parvum*, *C. canis*, *C. felis*, *C. muris*, *C. andersoni* and *C. meleagridis*) and among 40 *Cryptosporidium* genotypes, only two infect human: pig and cervine genotypes (Fayer *et al.,* 2009). The virulence factors in other human infective *Cryptosporidium* species require investigation.

In addition to inter-species differences in virulence, there were reports of intra-species and inter-isolate virulence variability (Xiao *et al.,* 2000, Okhuysen *et al.,* 1999, Widmer *et al.,* 1998). The genetic determinants of virulence and host range for *Cryptosporidium* species and genotypes and the specific mechanisms that shape this biological diversity are not elucidated (Morgan *et al.,* 1999, Sulaiman *et al.,* 2000). The discovery of such factors is desirable as it would

advance our understanding of the pathogenesis, host tropism, evolution and epidemiology of *Cryptosporidium* species.

1.11.5 Virulence factors and species determinants

The consideration of genetic determinants of host tropism as virulence factors is legitimate as these genes would enable the parasite to adhere and invade epithelial cells from different host species. However, the relationships between the host range and the virulence of human infective *Cryptosporidium* species is not clear. The Red Queen hypothesis postulates that species continually evolve but do not become better adapted (Venditti *et al.,* 2010, Van Valen, 1973), therefore the interactions between hosts and parasites lead to constant natural selection for adaptation and counter adaptation (Lively and Dybdahl, 2000). Antagonistic coevolution between hosts and parasites can involve rapid fluctuations of genotype frequencies and recombinations in the host, which may be advantageous and can quickly produce disproportionately fit offspring (Engelstädter and Bonhoeffer, 2009). In fact, successful parasite species should evolve to become less virulent overtime and therefore, only the maladapted novel parasites are harmful (Ebert, 1994). Considering the host range for *C. hominis* and *C. parvum*, the latter is thought to be more virulent allowing it to infect a wide range of species, while the former is expected to be less virulent as postulated by the theory of host-parasite co-evolution and virulence attenuation. However, Hunter and colleagues (2004) showed that *C. hominis* is more virulent. Based on similar observations, Woolhouse and colleagues (2002) stated that host-parasite co-evolution is in principal a powerful determinant of the biology and genetics of infection and disease, which has proven difficult to demonstrate rigorously in practice. A theory on the evolution of the virulence states that pathogenicity can be maintained when it is a direct or indirect consequence of the parasite's exploitation of the host (Ebert, 1994). Therefore, the parasite is expected to balance parasite reproduction against host survival, such that parasite transmission is maximised (Ebert, 1994, Anderson and May, 1984).

The identification of genetic determinants of host tropism in human infective *Cryptosporidium* species would improve our understanding of evolution, virulence and epidemiology of *Cryptosporidium* species. In addition, these virulence factors could be used as potential drug target and vaccine candidate (Casadevall and Pirofski, 2001, Brubaker, 1985).

1.12 Genomics of *Cryptosporidium* species

1.12.1 Genome organisation

Several studies tried to reveal the characteristics of *Cryptosporidium* genome prior the sequencing era. Karyotypic analyses suggested that *Cryptosporidium* contains eight chromosomes, ranging in size from 0.945 to 2.2 Mb, giving a total haploid genome size of approximately 10 Mb (Blunt *et al.,* 1997, Hays *et al.,* 1995). In addition, *C. parvum* was shown to have two small extrachromosomal cytoplasmic virus-like double-stranded RNAs (of 1,786 and 1,374 nucleotides, respectively) (Khramtsov *et al.,* 1997). The RNAs have a single open reading frame each, which encodes a putative RNA-dependent RNA polymerase and a protein with limited homology to mammalian protein kinases, respectively (Clark, 1999). Le Blancq and colleagues (1997) investigated *C. parvum* rRNA gene organization and reported that the small and large rRNA subunits are 1.7 and 3.6 kb, respectively, plus a 151 bp putative 5.8S rRNA. It was also demonstrated that *Cryptosporidium* has mitochondrial biosynthesis genes (LaGier *et al.,* 2003) and unlike other apicomplexan, it lacks an apicoplast (Zhu *et al.,* 2000 a).

1.12.2 Genome sequencing projects

As a pathogen of public health relevance, *Cryptosporidium* was included in genome sequencing projects. Two reference strains served as genome representatives: *C. parvum* Iowa and *C. hominis* TU502. Genome sequences showed similar genome sizes of 9.11 and 9.16 Mb, respectively (Abrahamsen *et al*, 2004, Xu *et al*, 2004). Genome analysis revealed extremely streamlined

metabolic pathways and lack of many cellular structures and metabolic pathways found in other apicomplexans (Wanyiri and Ward, 2006). Energy metabolism is largely from glycolysis and both aerobic and anaerobic metabolisms are available, thus conferring environmental flexibility (Barta and Thompson, 2006). Limited biosynthetic capabilities and minimal metabolism were reported, suggesting a large dependence on nutrient acquisition from the host (Rider and Zhu, 2010). *Cryptosporidium* has genes associated with apical complex organelles despite the fact that they lack an apicoplast and possess a degenerate mitochondrion that has lost its genome (Abrahamsen *et al*, 2004, Xu *et al*, 2004).

A comprehensive genome database, CryptoDB (www.CryptoDB.org), serves as a public interface to *Cryptosporidium* genome sequences (Puiu *et al.,* 2004). This website offers access to sophisticated tools, which enable the identification of genes based on text, sequence similarity and motif queries (Striepen and Kissinger, 2004). Expressed sequence tag (EST) and genome sequence survey (GSS) DNA-sequencing projects are in progress, together with the genome sequencing of *C. muris* RN66 strain.

Pain and colleagues (2005) summarized the similarities between *C. hominis* and *C. parvum*. Like *C. parvum*, *C. hominis* seems to lack an apicoplast and has an atypical mitochondrion with a minimal set of mitochondrial proteins. In addition, the two species have a reduced number of metabolic pathways. Energy metabolism is mainly from glycolysis and both aerobic and anaerobic metabolic pathways are available. Several biochemical pathways are absent from the *C. hominis* genome, including the tricarboxylic acid cycle, urea and nitrogen cycle, shikimate pathway, isoprenoid biosynthesis and the type-II fatty-acid biosynthesis pathway. Biosynthetic capabilities are also restricted, *C. hominis* cannot synthesize some simple sugars, amino acids and nucleotides, while starch and amylopectins can be generated from precursors by *C. hominis* enzymes. The apparent lack of biosynthetic capability of *C. hominis* is compensated for by the

presence of an extensive array of amino-acid and other transporters that enable import of essential nutrients from the host.

1.12.3 Genomics and virulence factors

Genome sequences revealed a vast amount of information, contributing to a better knowledge of the microbial biology, pathogenicity, evolution and virulence as described above. The quest for the molecular basis of virulence can exploit the genomic data to search for genes that may ultimately unravel the regulation of virulence at the genetic level (Casadevall and Pirofski, 2001). Particularly, post-genomic analysis is a potential tool to identify genes involved in host-parasite interaction and adaptation. These contingency genes as opposed to conserved housekeeping genes are expected to be highly variable (Moxon *et al.,* 1994). Barry and colleagues (2003) stated that genes responsible for the successful invasion of the host are subject to spontaneous recombination rates higher than the background rate that applies to the other genes in the genome. Contingency genes are common in pathogenic microbes including viral, bacterial, fungal and protozoan pathogens (Henderson *et al.,* 1999, Deitsch *et al.,* 1997). These genes enable, through mutational events, rapid switches in phenotype that are conductive to survival and proliferation in the host and are often associated with telomeres (Barry *et al.,* 2003). Some examples of contingency genes include a variant surface protein (VSG) *in Trypanosoma brucei*, which undergoes antigenic variation to evade the host immune response and allow parasite survival (Yang *et al.,* 2009, Barry *et al.,* 2003) and *var* genes in *Plasmodium falciparum* coding for erythrocyte protein-1, which also undergo antigenic variation to evade the immune system (Kyes *et al.,* 2001). Genetic determinants of host tropism are likely to be among these contingency genes and genomic data would assist in identifying them.

1.13 Aims and scope of the research in this thesis

C. hominis and *C. parvum* are the *Cryptosporidium* species of public health relevance. Genomes representative of these two species have been sequenced and showed only 3-5% sequence divergence. No further investigation of this minimal diversity was performed. The aim of this thesis is to exploit the published genomic data to identify species-specific *Cryptosporidium* genes accountable for this sequence variability. This research objective is novel and might represent a unique opportunity to identify genetic determinants of host tropism and species-specific virulence factors.

The investigation was based on the *in silico* identification of putatively species-specific genes using bioinformatic tools. The predicted specificity was assessed experimentally by PCR in a collection of *Cryptosporidium* clinical isolates provided by the *Cryptosporidium* reference unit and commercially obtained reference strains. The putative specific genes were further characterized at the genomic and molecular level and a preliminary investigation of their role in the host-parasite interaction was undertaken.

A detailed study plan:

- Comparative genomics was performed using Reciprocal BLAST and used to assign a similarity score to every *C. parvum* and *C. hominis* gene
- Initial selection for highly variable genes was undertaken by applying a threshold of 10% sequence identity
- Selected genes were individually tested by BLAST and only genes with no known sequence similarity were retained. In addition, query of genomic databases was performed for ortholog identification. This additional screening allowed elimination of genes showing any sequence similarity to other *Cryptosporidium* species

- Experimental evaluation of the predicted specificity was undertaken by PCR in a collection of *Cryptosporidium* clinical isolates and reference strains
- PCR product sequence analysis was performed to determine genetic polymorphism
- Further investigation and characterization of species-specific genes was planned and involved
    - ❖ Sequence polymorphism in clinical isolates
    - ❖ Evaluation of the diagnostic applications for specific genes
    - ❖ Expression and characterization of the recombinant proteins encoded by the species-specific genes
    - ❖ Production of monoclonal antibodies
    - ❖ Protein localization studies by IFA
    - ❖ Evaluation of the role of specific proteins in host-parasite interaction *in vitro* using blocking assays in Caco2 cell monolayers
    - ❖ Assessment of the immunogenicity of the recombinant proteins by Western Blot screening of sera from infected patients

CHAPTER 2:

Material and methods

2.1 Identification of putative species-specific genes by comparative genomics

2.1.1 Reciprocal BLAST

The genome sequences of *C. parvum* and *C. hominis* have been sequenced as described previously (Abrahamsen *et al*, 2004, Xu *et al*, 2004) and are available online (http://CryptoDB.org). In order to investigate the sequence variability between the genome representatives of these 2 species, estimated to be 3-5%, Reciprocal BLAST was used. Reciprocal BLAST is a variant of BLAST (Basic local alignment search tool), originally described by Altschul and colleagues (1990). Reciprocal BLAST is a common computational tool for predicting putative orthologs (http://www.ncbi.nlm.nih.gov/blast/blast_overview.shtml). Orthologs are homologous sequences found in different species and derived from a common ancestor. Reciprocal BLAST is used to identify orthologous genes between 2 species as follows: Each gene of species 1 is Blasted to a database of gene sequences from the organism of interest (species 2). The highest scoring gene is taken and Blasted to a database of species 1 gene sequences. If the BLAST hit returns the gene originally used for reciprocal BLAST as the highest BLAST scorer, then the 2 genes are considered putative orthologs.

For this study, Reciprocal BLAST was used to compare the genomes of *C. parvum* and *C. hominis* and identify orthologous genes. It conferred a similarity score to each of the ~ 3900 genes of *C. parvum* and *C. hominis*. Applying a threshold of 10% identity allowed the identification of putative species-specific genes. In addition, within this selection, each gene was individually Blasted using Blastn algorithm (http://blast.ncbi.nlm.nih.gov/Blast.cgi) to confirm specificity and reveal sequence similarity to genes from the other *Cryptosporidium* species. Furthermore, orthology queries were performed using the CryptoDB database. Whenever a gene showed sequence similarity, it was eliminated from the selection.

2.1.2 Gene selection

Amongst the putative species-specific genes selected as described in 2.1.1, 13 genes were chosen with preference to annotated genes The name and annotation of the selected genes are detailed in Table 2.1.

2.2 Screening of putative species-specific genes by PCR

2.2.1 Primer design

For each gene, a pair of primers was designed using OligoPerfect™ Designer software (www.invitrogen.com). The primers were ordered from Operon/ Eurofins MWG (Cologne, Germany). Table 2.1 detailed the primers sequences and the PCR product sizes for each gene tested. In addition, reference primers Cry15 and Cry9 amplifying a 555 bp fragment of the COWP (*Cryptosporidium* oocyst wall protein) gene (Spano *et al*, 1997) were used as a positive control.

2.2.2 PCR conditions

PCR conditions were carried out as described previously (Bouzid *et al.,* 2010). The PCR mix consisted of 1.5 mM $MgCl_2$, 0.2 mM of each dNTP (Bioline, UK), 0.6 µM of Forward and reverse primers and 2.5 U of HotStarTaq ® DNA polymerase (Qiagen Ltd., Crawley, UK) in 50 µl reaction. The cycling conditions were as follows: an initial hot-start at 95°C for 15 min, followed by 40 cycles of 94°C for 50s, 60 (or 57) °C for 30s and 72°C for 50s, and a final extension at 72°C for 10 min. PCR reactions were performed using Techne TC-512 thermal cycler (SLS, Nottingham, UK). The PCR products were run on a 2% agarose gel and visualised under UV.

Primer name	Gene function (CryptoDB)	Sequence	Tm (°C)	Annealing temperature (°C)	Size of amplified fragment
cgd2_80 F	ABC transporter family protein	GGA TTG GGG GTG ATA TGT TG	68	60	266 bp
cgd2_80 R		ACC TCC AAG CTG TGT TCC AG	70		
cgd6_200 F	Oocyst wall protein 8	CGT TCC AAC AAT GGT GTG TC	68	60	447 bp
cgd6_200 R		GCA GCT GGA GTG CAA TCA TA	68		
cgd8_2370 F	Adenosine kinase like ribokinase	CAG GAA TTG CTC ACG GAA AT	66	60	685 bp
cgd8_2370 R		CCT TAA ATG CAT CCC CAC AG	68		
Chro.50317 F	RNA polymerase A/beta'/A'' subunit	GAT TTT GAT GGA GGG TCT CG	68	60	752 bp
Chro.50317 R		CTG GCA GCT TCA ACA CCA TA	68		
Chro.30149 F	Ubiquitin-protein ligase 1	GGG ATT AGA TGC AGG TGG TG	70	60	331 bp
Chro.30149 R		TGG ATG CTC CAG CAT TAC AT	66		
Chro.50457 F	Erythrocyte membrane-associated antigen	CCT TTG GAT TGT CCC GAA TA	66	60	394 bp
Chro.50457 R		CAA TGC CAT ATG ATT TGA GAA AAA	65		
cgd6_5020 F	Protein with WD40 repeats	AAC AGG AGC TGA CGA TTG CT	60.4	57	271 bp
cgd6_5020 R		ACA TTG TGC CAT TCC AAG GT	58.35		
cgd2_2430 F	Ximpact ortholog conserved protein seen in bacteria and eukaryotes	GTA ACG CAT GGC GAA CCT AT	60.4	57	389 bp
cgd2_2430 R		AAG ATC AGC CTT GCA GCA TT	58.35		
cgd2_4380 F	signal peptide, repeats, gene anchored to telomere	AAG GGG TGG ACC TAG ATG CT	62.45	57	665 bp
cgd2_4380 R		GAA GAG GTG GGC GTG ATC TA	62.45		
Chro.20156 F	Hypothetical protein	TTC GCT TGA AGC CGT AAA CT	58.35	57	247 bp
Chro.20156 R		GGC ATT GAT ACC AGG CAA GT	60.4		
Chro.50330 F	Leucyl tRNA synthetase	TCG GTA CAG CAT CAG GTT CA	60.4	57	368 bp
Chro.50330 R		GTT TTT GCT CCC CCA GTT TT	58.35		
Chro.00003 F	Sensor histidine kinase	ACC TGC TCT CCG TTG TCA AT	60.4	57	610 bp
Chro.00003 R		ATG CGA ACC TCC TCA CTG TC	62.45		
Chro.50011 F	Hypothetical protein	AGG AAA CGT GGC ATT TTC TG	63.9	57	287 bp
Chro.5001 R		TTG ACA GCA CTT CCT GAA CG	64.1		
Cry-15	Oocyst wall protein gene (Spano *et al.,* 1997)	GTA GAT AAT GGA AGA GAT TGT G	57.08	60	555 bp
Cry-9		GGA CTG AAA TAC AGG CAT TAT CTT G	61.3		

Table 2.1: List of the genes selected for this study, the corresponding primer sequences and PCR product sizes. Primer name is the gene name followed by F or R (for forward and reverse, respectively).

2.3 *Cryptosporidium* DNA

During this study, a panel of clinical samples from the collection of the UK *Cryptosporidium* Reference Unit (CRU), Swansea was used for the screening of putative species-specific genes. These isolates originated from diarrhoea patients with confirmed cryptosporidiosis. DNA was isolated from semi-purified oocyst suspensions prepared from stool samples by saturated-salt solution centrifugation and extracted using QIAamp DNA mini kit spin columns (Qiagen Ltd, Crawley, UK) as previously described by Elwin and colleagues (2001). For each isolate, speciation was performed by PCR-RFLP of the *Cryptosporidium* oocyst wall protein (COWP) gene as previously described by Spano and colleagues (1997) and by real-time PCR using simplex Lib 13 primers for *C. parvum* and *C. hominis* as described by Tanriverdi and colleagues (2003). In addition, genotyping and subtyping of these isolates was performed by sequencing of the SSU rRNA and GP60, as previously described (Xiao *et al.,* 2001 and Mallon *et al.,* 2003 b).

DNA from three *C. hominis* isolates (Ch2, Ch3 and Ch4), 3 *C. parvum* isolates (Cp2, Cp3, and Cp4) and four *C. parvum* anthroponotic subtype isolates (W65, W66, W67 and W70) were tested. The anthroponotic *C. parvum* group was identified using Gp60 subtyping; IIc subtype was only reported to infect humans, which led to this sub-categorisation (Xiao and Fayer, 2008). Table 2.2 details the origin, epidemiological and genotyping data of the isolates tested. In addition, DNA of four rabbit genotype isolates, one isolate from the Northamptonshire outbreak (Chalmers *et al.*, 2009) and three sporadic cases isolates, were also analysed. Furthermore, genomic DNA of three reference strains *C. parvum* Iowa (ATCC/ LGC Promochem, Teddington, UK), *C. parvum* Moredun (Moredun Research Institute, Midlothian, UK) and *C. hominis* TU502 (BEI Resources, Manassas, USA) were tested.

.

Isolate	Host	Origin	COWP- RFLP	18s sequencing (genotyping)	Gp60 sequencing (subtyping)
C. parvum Iowa	Bovine	Iowa, USA	*C parvum*		
C. hominis TU502	Human	Uganda	*C hominis*		
C. parvum Moredun	Cervine	Scotland	*C parvum*		
Ch2	Human	Yorkshire, England	*C hominis*	*C. hominis* **GQ983348**	IbA10G2 **GQ983356**
Ch3	Human	North Wales	*C hominis*	*C. hominis* **GQ983350**	IbA10G2 **GQ983358**
Ch4	Human	Cumbria, England	*C hominis*	*C. hominis* **GQ983352**	IbA10G2 **GQ983360**
Cp2	Human	Devon, England	*C parvum*	*C parvum* **GQ983349**	IIaA18G3R1 **GQ983357**
Cp3	Human	Cumbria, England	*C parvum*	*C parvum* **GQ983351**	IIaA17G1R1 **GQ983359**
Cp4	Human	Grampian, Scotland	*C parvum*	*C. parvum* **GQ983353**	IIaA15G2R1 **GQ983361**
W65	Human	Leicestershire, England	*C parvum*	*C. parvum* **GU971620**	IIcA5G3 **GU971624**
W66	Human	Leicestershire, England	*C parvum*	*C. parvum* **GU971621**	IIcA5G3 **GU971625**
W67	Human	Leicestershire, England	*C parvum*	*C. parvum* **GU971622**	IIcA5G3 **GU971626**
W70	Human	Leicestershire, England	*C parvum*	*C. parvum* **GU971623**	IIcA5G3 **GU971627**
Rabbit 1 (17330)	Human	Northampton-shire, England	*C hominis*	Rabbit genotype **FJ262726**	VaA18 **FJ262732**
Rabbit 2 (18455)	Human	Shropshire, England	*C hominis*	Rabbit genotype **GU971628**	VaA23 **GU971631**
Rabbit 3 (17525)	Human	Suffolk, England	*C hominis*	Rabbit genotype **GU971629**	VaA32 **GU971632**
Rabbit 4 (17435)	Human	Essex, England	*C hominis*	Rabbit genotype **GU971630**	VaA22 **GU971633**

Table 2.2: Details of the host, the geographical origin and the genotyping data of *C. parvum* and *C. hominis* isolates and reference strains, whose DNA was tested during this study. Numbers in bold are the Genbank accession number for sequences used to genotype and subtype these clinical isolates.

In addition, DNA from other *Cryptosporidium* species was also obtained from CRU and tested to further validate the predicted specificity of *in silico* identified genes. DNA from *C. andersoni, C. felis*, cervine genotype, *C. meleagridis* and *C. baileyi* was tested. Table 2.3 summarizes the epidemiological and genotyping data of these isolates.

2.4 Sequence analysis of PCR products and data analysis

Positive PCR products were purified using QIAquick® PCR purification Kit (Qiagen Ltd., Crawley, UK). Purified PCR products were sequenced in both directions using PCR primers in duplicate. In order to confirm sequences, two independent sequencing facilities were used: the genome lab, John Innes Centre (www.jicgenomelab.co.uk) and the sequencing service at the University of Dundee (www.dnaseq.co.uk), both using Dye-terminator chemistry technology and Applied Biosystems automated capillary DNA sequencer (3770 and 3730 model, respectively). Sequences were assembled using CAP3 software (http://pbil.univ-lyon1.fr/cap3.php) (Huang and Madan, 1999) and aligned using AlignX® application of Vector NTI Advance™ 10 software (www.Invitrogen.com). Phylogenetic analysis was performed using MEGA (Molecular Evolutionary Genetic Analysis) software (www.megasoftware.net) (Tamura *et al.,* 2007).

When required, the sequences were analyzed for the presence of endonuclease restriction sites using NEBcutter V2.0 online tool (http://tools.neb.com/NEBcutter2/) (Vincze *et al.,* 2003).

Isolate	Species	Host	COWP- RFLP	18s sequencing	Other loci
W13086	*C. andersoni*	Bovine	Not detected		HSP70 100% *C. andersoni*
W14508	*C. felis*	Human	Not detected	100% *C. felis*	
W15916	cervine genotype	Human	cervine genotype	100% cervine gt	
W10509	*C. meleagridis*	Human	*C. meleagridis*	100% *C. meleagridis*	
W14184	*C. baileyi*	Chicken	Not detected	100% *C. baileyi*	

Table 2.3: Details of the host, the geographical origin and the genotyping data of other *Cryptosporidium* species DNA used for this study. All species and genotypes were identified using 18S rRNA sequencing, except for *C. andersoni*, which was genotyped using HSP70 locus.

2.5 Whole Genome Amplification (WGA)

2.5.1 WGA kits

Three commercial WGA kits were tested during this study: illustra GenomiPhi V2 DNA amplification Kit (GE healthcare, Buckinghamshire, UK), REPLI-g Ultra fast Mini Kit (Qiagen, Crawley, UK) and GenomePlex® Complete WGA Kit (Sigma, Dorset, UK). Illustra GenomiPhi and REPLI-g kits are MDA-based WGA, while the GenomePlex kit is based on the Omniplex technique. All 3 kits were used according to the manufacturer's instructions.

2.5.2 *Cryptosporidium* DNA

Ten clinical samples were selected from the collection of the CRU and used for the WGA kit trial. They originated from diarrhoea patients with confirmed cryptosporidiosis from different geographical locations in the UK (Table 2.4). DNA was isolated as described in 2.3. For each isolate, the speciation was performed by PCR-RFLP of the COWP gene and by real-time PCR using simplex Lib 13 as described in 2.3. In addition, DNA of the *C. parvum* Iowa reference strain (ATCC/LGC Promochem, Teddington, UK) was also tested. All DNA samples were quantified by spectrometry using Nanodrop ND-1000 (Thermo Fisher scientific, Leicestershire, UK).

2.5.3 Quantification of genomic DNA after WGA

After WGA, the amplified products were analyzed by agarose gel electrophoresis as a semi-quantitative approach to evaluate the level of amplification of the genomic DNA. The electrophoresis profiles were also used for densitometry-based DNA quantification using GeneTools software (Syngene, Cambridge, UK).

Sample reference	Age of case	Gender of case	Origin of case	RFLP of the COWP gene (speciation)
W15504	43	F	Scotland	*C hominis*
W15507	9	F	Wales	*C hominis*
W15508	31	Not stated	England	*C hominis*
W15519	68	F	England	*C hominis*
W15521	14	F	Scotland	*C hominis*
W15509	12	Not stated	England	*C parvum*
W15511	19	F	Scotland	*C parvum*
W15516	21	F	England	*C parvum*
W15517	12	M	England	*C parvum*
W15518	2	M	England	*C parvum*

Table 2.4: Details of the origin and epidemiological data of clinical isolates of *C. hominis* and *C. parvum*, whose DNA was used for WGA kits trial.

In addition, 3 methods were used to quantify the amplified genomic DNA: Nanodrop, Hoechst and PicoGreen. Nanodrop is an absorbance-based method and was used to estimate the amount of DNA available in each sample before and after WGA. However, because this method is influenced by the presence of free nucleotides in the reaction mix, alternative quantitation methods were used.

Hoechst 33258 dye was used for quantification of the amplified DNA. It exhibits enhanced fluorescence when bound to dsDNA under high ionic strength conditions (Goumenou and Machera, 2004). For standard curve purposes, serial dilutions of calf thymus DNA stock solution (1 mg/ ml) (Sigma, Dorset, UK) were performed, yielding concentrations ranging from 100 ng/ ml to 2500 ng/ ml. Then, 50 µl of each preparation was placed in duplicates for measurement. All DNA samples were diluted in TNE buffer (100 mM Tris, 1 M NaCl, 10 mM Na_2EDTA, pH 7.5). For each DNA sample, an equal volume of 2x Hoechst Dye solution (200 ng/ ml) (Sigma, Dorset, UK) was added. Fluorescence was read using microplate reader FLUOstar, BMG Labtech (Aylesbury, UK) after 5 min incubation at room temperature. The blank solution was prepared by adding an equal volume of TNE buffer and 2x dye solution. The average value of the blank measurement was subtracted from the duplicate measurements made at each concentration and the results plotted. A linear regression was performed on the standard curve to allow the determination of the DNA concentrations of the samples tested.

PicoGreen was also used as an ultrasensitive fluorescent nucleic acid stain for accurate quantitation of double stranded DNA (dsDNA). Quanti-iT™ Picogreen® dsDNA kit (Molecular Probes, Invitrogen, Paisley, UK) was used according to the manufacturer's instructions.

2.5.4 Integrity and fidelity of amplified genomic DNA

The integrity of the amplified DNA after WGA was evaluated by PCR using Cry15/ Cry 9, Cgd6_5020 and Chro.20156 primers, amplifying 550 bp, 270 bp and 247 bp, respectively. All DNA templates were diluted 1/25 after WGA. PCR conditions were identical for genomic DNA and for post-WGA subsequent amplifications and are as described in 2.2.2.

The fidelity of the amplification was assessed by PCR product sequence analysis before and after WGA of two *C. hominis* samples (W15507, W15519), two *C. parvum* samples (W15511, W15516) and the reference strain *C. parvum* Iowa. PCR products were purified, sequenced and analyzed as described in 2.4.

2.5.5 Comparative analysis of *Cryptosporidium* genomic DNA before and after WGA

The comparative analysis of genomic *Cryptosporidium* DNA and paired WGA DNA was limited to the samples that were prepared using the most promising WGA method. After WGA, the amplified products were compared to the original genomic DNA using a species-specific, semi-quantitative real-time PCR assay. Briefly, real-time PCR using simplex Lib 13 primers (Tanriverdi *et al.,* 2003) was used in triplicate to amplify 2 µl genomic DNA and 2 µl of its paired WGA DNA (both diluted 1 in 25 v/v). Melt curve analysis was performed to identify *Cryptosporidium* species and C_T values were recorded to compare each paired sample, before and after WGA. This assay shows a demonstrable difference of 3 C_T units per 10-fold difference in DNA target copy number (CRU unpublished data). Since the WGA DNA originated from 1 µl of genomic DNA producing 10 µl of product (representing a 10-fold dilution in comparison with the genomic DNA), the final C_T values for the WGA DNA were decreased by 3 C_T units.

2.6 *Cryptosporidium parvum* specific gene (Cops-1): Cgd2_4380

2.6.1 Primers used for Cops-1 amplification

In addition to Cgd2_4380F and R primers described in 2.2.1, Cgd2_4380 FF and FR primers amplifying the full gene length (1434 bp) were designed. The primer sequences are detailed in Table 2.5. Several internal primers were also designed. These internal primers were used to test the presence of the gene in *C. hominis* and to retrieve the gene sequence in a primer walking approach. In addition, a primer amplifying the 3' flanking region to Cgd2_4380 gene was also designed and tested. Primer sequences are detailed in Table 2.5. PCR was carried out as previously described in 2.2.2. The annealing temperature was 55°C for all the primers but was lowered to 52°C or 47°C for some primer combinations to decrease the stringency and allow amplification.

2.6.2 Cops-1 protein encoding gene

The predicted Cops-1 encoded protein was analyzed using online software tools for protein analysis: Protein Calculator v3.3 (http://www.scripps.edu/~cdputnam/protcalc.html), InterProScan (http://www.ebi.ac.uk/Tools/InterProScan/), PSORT II (http://psort.ims.u-tokyo.ac.jp/) and SignalP3.0 (http://www.cbs.dtu.dk/services/SignalP/).

Primer name	Sequence	Gene position	Tm (°C)
Cgd2_4380FF	ATGGGTAATAGTTTAAATGTTTTT	1-24	52.61
Cgd2_4380FR	TTATTTTCGGCATAACGG	1417-1434	53.07
Cgd2_4380_2F	GGGGTGGACCTAGATGCTC	245-263	64.48
Cgd2_4380_2R	GAGGTGGGCGTGATCTAGTAAA	883-904	62.67
Cgd2_4380_flanking_5'end	AAAAGCGCAAGTAATCTGGA	Upstream sequence	56.3
Cgd2_4380_flanking_3R	CCTAAACCTAACCCCCTAAACCT	Downstream sequence	62.77

Table 2.5: Primers used to amplify Cops-1 gene in a primer walking approach. The primers sequence and Tm are detailed. In addition, the position on the Cops-1 gene is detailed for each primer.

2.6.3 Monoclonal peptide antibody to Cops-1

In order to further characterize the Cops-1 protein, a monoclonal peptide antibody was produced by Cambridge Research Biochemicals (www.crbdiscovery.com). Predicted protein sequence as retrieved from CryptoDB database (477 aa), was submitted for peptide design. Two sequences were selected corresponding to high immunogenicity: antigen 1 (Tyle-1) CDHEGFKPPRRTTT-amide (position 404-417) and antigen 2 (Tyle-2) [C]-RSRPPLPTRKPYSGS-amide (position 297-311). Briefly, 5 mg of each antigen was prepared at >95% purity (as determined by HPLC analysis), 3 mg were conjugated to keyhole limpet hemocyanin (KLH) carrier protein and used to immunize Balb/c female mice. Mice bleeds were tested by ELISA and the animal having the highest titre (for each antigen) was used for spleen extraction and fusion with mouse myeloma cells. The cells were cultured in selective medium. After 10 days incubation, hybridoma supernatants were tested for the presence of specific IgG anti-Tyle 1 and Tyle 2 antibodies by ELISA. Immunopositive hybridoma cultures were expanded to large scale culture to allow production of antibodies.

2.7 Cloning and expression of Cops-1 recombinant protein

2.7.1 Cloning of Cops-1

Two approaches were used to clone the Cops-1 gene. Cloning was performed into pGEM®-T Easy Vector System (Promega, Southampton, UK) according to the manufacturer's instructions. Briefly, Iowa strain DNA (supplied as previously described in 2.3) was amplified by PCR using the Expand High Fidelity PCR system (Roche, Hertfordshire, UK) to reduce PCR errors. The full length of the cgd2_4380 gene was amplified using Cgd2_4380_FF and FR primers in a 50 µl reaction. The PCR product was visualized by agarose gel electrophoresis and then purified using the QIAquick® PCR purification Kit as described in 2.4. The DNA was quantified using Nanodrop ND-1000. 50 ng of pGEM®-T Easy Vector, 80 ng of PCR

product and 3U of T4 DNA ligase were incubated overnight at 4°C as instructed by the manufacturer. The ligation mix was used to transform JM109 High efficiency competent cells (genotype: e14-(McrA-) *recA1 endA1 gyrA96 thi-1 hsdR17* (r_{K-} m_{K+}) *supE44 relA1* Δ(*lac-proAB*) [F′ *traD36 proAB* $lacI^{q}ZΔM15$]). The positive clones were identified through the blue/white selection on LB/Ampicillin/IPTG/X-Gal agar plates. At least 3 positive colonies were used to isolate the recombinant plasmid using a Qiagen plasmid purification Midi kit according to the manufacturer's instructions (Qiagen, Ltd., Crawley, UK). Briefly, one colony was used to inoculate 5 ml of LB (Luria-Bertani) medium (1% bacto-tryptone, 0.5% yeast extract, 1% NaCl, pH 7.0) containing 50 µg/ ml Ampicillin. After 8h incubation at 37°C, the culture was diluted 1/500 into 100 ml LB selective medium. After overnight incubation, the cells were harvested by centrifugation at 6000 x *g* for 15 min at 4°C. The pellet containing the plasmid was then purified using the spin columns provided. Alternatively, the PCR product was cloned using forward primer: Cgd2_4380FF_*XhoI* CTCGAGATGGGTAATAGTTTAAATGTTTTT (Tm= 60.53°C) and a reverse primer: Cgd2_4380FR_*BamHI* GGATCCTTATTTTCGGCATAACGG (Tm= 62.86°C). The underlined sequences correspond to the restriction sites *XhoI* and *BamHI*, respectively. The PCR product was visualized by agarose gel electrophoresis before digestion using *XhoI* and *BamHI* enzymes to allow cloning into the expression vector.

In addition, Champion™ pET100 Directional TOPO® Expression Kit (Invitrogen, Paisley, UK) was used according to the manufacturer's instructions. Briefly, Cgd2_4380FF_TOPO primer (CACCATGGGTAATAGTTTAAATGTTTTT) was designed (Tm= 57.8°C). The underlined sequence will allow directional cloning of the PCR product by the GTGG overhang in the cloning vector pET100/D-TOPO® and to the action of Topoisomerase I. Amplification of the full length of Cgd2_4380 gene from Iowa DNA was performed using Cgd2_4380FF_TOPO and Cgd2_4380FR primers. PCR conditions were as described above. The PCR product was integrated into the pET100/D-TOPO® vector by incubation for 30 min at room temperature and used to transform One shot® TOP10 competent cells

(genotype: *end*A1, *rec*A1, *gyr*A96, *thi, hsd*R17 (r_k^-, m_k^+), *rel*A1, *sup*E44, Δ(*lac-pro*AB), [F′ *tra*D36, *pro*AB, *laq*I^qZΔM15]). Positive transformants were identified by PCR using T7 and Cgd2_4380_FR primers. The recombinant plasmid was isolated as described above.

The isolated recombinant plasmids from pGEM®-T Easy Vector and pET100/D-TOPO® vector were sequenced in both directions to confirm gene sequence and orientation.

2.7.2 Expression vectors

Two cloning vectors were used as described in 2.7.1. For pGEM®-T Easy Vector containing the Cgd2_4380 gene, 1 µg of plasmid preparation was used for double digestion using *BamHI* and *XhoI* enzymes (10U each) in 1x SuRE/Cut Buffer B (Roche, Hertfordshire, UK) in a final volume of 20 µl. The Expression vector pTrcHisA (commonly used in our laboratory and originally from Invitrogen) was digested using the same restriction enzymes. The linearized pTrcHisA vector and the Cgd2_4380 gene (released from the pGEM®-T Easy Vector by digestion) were purified using QIAquick Gel Extraction Kit (Qiagen, Ltd., Crawley, UK) and quantified by Nanodrop ND-1000. The molar ratio insert to vector was determined according the following formula:

$$\frac{\text{ng of vector} \times \text{kb size of insert}}{\text{kb size of vector}} \times \text{molar ratio of } \frac{\text{insert}}{\text{vector}} = \text{ng of insert}$$

Different molar ratios were tested to determine the optimal ligation conditions: 1:1, 3:1 and 10:1 and ligation was performed overnight at 4°C using T4 DNA ligase (Promega, Southampton, UK). The pET100/D-TOPO® vector is used as a cloning and expression vector, so no further subcloning was required.

2.7.3 Bacterial strains used for expression

The recombinant pTrcHisA vector was used to transform DH5α competent cells (genotype: F– Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17 (rK–, mK+) *pho*A *sup*E44 λ– *thi*-1 *gyr*A96 *rel*A1). 10 ng of vector was used to transform a 200 µl cell aliquot. Cells were then heat shocked at 42°C for 45s and incubated in 400 µl SOC medium for 1h. 100 µl of each transformation was then plated on selective LB plate containing 50 µg/ ml Ampicillin.

The pET100/D-TOPO® vector containing the insert was used to transform BL21 Star™(DE3) competent cells (genotype: F– *ompT hsd*SB(rB–, mB–) *gal dcm rne*131 (DE3)) according to the manufacturer's instructions. Briefly, 10 ng of plasmid DNA was used to transform 50 µl of BL21 Star™ (DE3) One Shot® cells, which were then heat shocked at 42°C for 30s and incubated in 250 µl SOC medium for 1h. The entire transformation reaction was then added to 10 ml of LB containing 50 µg/ ml Ampicillin.

2.7.4 Specialized bacterial expression strains

In addition to standard bacterial expression strains (described in 2.7.3), specialized bacterial strains were used in order to optimize the expression of the recombinant protein. These strains were kindly donated by Dr Matt Hutchings (School of Biological Sciences, UEA). Two of these strains were engineered to contain extra copies of genes that encode rare tRNAs, which allows high-level expression: BL21-CodonPlus-RP (genotype: F– *ompT hsdS*(rB– mB–) *dcm*+ Tetr *gal endA* Hte [*argU proL* Camr]) and Rosetta™2(DE3) (genotype: F^- *ompT* $hsdS_B(r_B^- m_B^-)$ *gal dcm* (DE3) pRARE2 (Cam^R)). The former supplies tRNAs for seven rare codons (AUA, AGG, AGA, CUA, CCC, CGG, GGA) and the latter supplies tRNAs for three rare codons (AGA, AGG, CCC). In addition, two bacterial strains engineered to allow expression of membrane proteins (Wagner *et al.,* 2008) were used: C41(DE3) (genotype: F –

ompT hsdSB (rB- mB-) gal dcm (DE3)) and C43 (DE3) (genotype: F – *ompT hsdSB (rB- mB-) gal dcm* (DE3)).

These bacterial strains were made chemically competent following this protocol: 100 µl of liquid culture were plated on LB-agar plate without antibody selection. After overnight culture at 37°C, a single colony was used to inoculate 15 ml LB and incubated overnight. 500 µl of this culture was used to inoculate 100 ml LB and the bacterial growth was monitored by OD_{600}. When the mid-log phase was reached (OD_{600} range 0.4-0.6), the cells were centrifuged (5000 xg for 5 min at 4°C). Bacterial pellet was re-suspended in 10 ml of 0.1M $CaCl_2$ and incubated on ice for 20 min. Cells were then centrifuged as previously and the pellet was re-suspended in 10 ml of 0.1M $CaCl_2$ containing 10% glycerol. Cells were then aliquoted and frozen in a dry ice/ ethanol bath and stored at -80°C.

These specialized expression strains were transformed as described in 2.7.3. BL21-CodonPlus-RP and Rosetta™2(DE3) strains have pACYC and pRARE plasmids, respectively, conferring Chloramphenicol resistance, so Chloramphenicol (34 µg/ml) was added to the LB medium for host strain selection.

2.7.5 Expression of Cops-1 recombinant protein

The expression of the recombinant protein was induced at mid-log phase by addition of 1mM IPTG (isopropyl-beta-D-thiogalactopyranoside). A pilot expression was performed from 5 ml cultures and the expression was monitored every hour for a time course (1-6 hours). A non-induced culture served as a negative control. At each time point, 500 µl aliquot was removed and centrifuged at 6000 xg for 5 min and cell pellets were stored at -20°C until use. Each pellet was then re-suspended in SDS-PAGE Sample buffer, boiled for 5 min and used for protein analysis by Coomasie staining and Western blot.

In addition, pET100/D/*lac*Z, a control expression vector provided with the Champion™ pET100 Directional TOPO® Expression Kit, was used. It encodes β-galactosidase and the fusion protein has an expected size of 121 kDa. The control expression plasmid was used to transform BL21 Star™ (DE3) One Shot® cells as described previously.

The pilot expression allowed the determination of the optimal expression conditions, which were adopted for the production of recombinant protein on a larger scale. Briefly, one positive colony was used to inoculate 5 ml of liquid LB media containing 50 µg/ ml Ampicillin and 34 µg/ml of Chloramphenicol (when appropriate). After overnight incubation, the whole culture was added to 100 ml liquid LB media containing the same antibiotics as detailed above. The bacterial growth was monitored by spectrophotometry until mid-log phase (OD_{600}= 0.5). The culture was then split into 2 flasks of 50 ml each. 1 mM IPTG was added to one flask, while the other was kept as uninduced culture. Both flasks were cultured for 4 h at 37°C.

2.7.6 Preparation of cell lysate

After expression, the cells (both induced and uninduced) were harvested by centrifugation at 6,000 x g for 15 min at 4°C. The cell pellets were stored at -80°C until use. Preparation of protein samples was carried out according to the Champion™ pET100 vector manual. Briefly, cell lysates were thawed on ice and then re-suspended in 500 µl of lysis buffer (400 mM NaCl, 100 mM KCl, 10% glycerol, 0.5% glycerol, 0.5% Triton X-100, 10 mM imidazole, 500 mM potassium phosphate, pH 7.8). The samples were subjected to 3 cycles of freeze/ thaw and then centrifuged at 15,000 rpm for 1 min. The supernatants were transferred to clean tubes. In order to prevent protein degradation, 20 µl of complete EDTA-free protease inhibitor cocktail (Roche, Hertfordshire, UK) and 10 µl of phosphatase inhibitor cocktail set II (Calbiochem, Merck Chemicals Ltd., Nottingham, UK) were added. The protein samples were kept at -20°C until use.

The pellet was re-suspended in lysis buffer and used for the detection of insoluble proteins.

2.8 Western Blot

2.8.1 Protein samples preparation

For Western Blot analysis, recombinant proteins prepared as described in 2.7.6 were used. Alternatively, native antigen preparations were prepared as follows: $2x10^8$ oocysts were pelleted by centrifugation at 2500 rpm for 10 min at 4°C. The pellet was washed in PBS and the suspension was centrifuged at 10000 rpm for 10 min at 4°C and this step was repeated 3 times. The pellet was then resuspended in PBS, then 5 µl of complete EDTA-free protease inhibitor cocktail and 5 µl of phosphatase inhibitor cocktail were added. The reaction was kept on ice and sonicated for 15 min. The parasite lysate was subjected to 3 cycles of freeze/ thaw and then centrifuged at 15000 rpm for 30 min at 4°C. The supernatant was collected and aliquoted and the antigen preparations were kept -20°C until use.

2.8.2 Protein quantification

Protein quantification was carried out using BCA™ protein assay kit (Pierce, Thermo Fisher Scientific, Leicestershire, UK) according to the manufacturer's instructions. Briefly, a working reagent solution was prepared by mixing 50 parts of BCA Reagent A with 1 part of BCA Reagent B. 200 µl of the working solution was added to 25 µl of protein samples (tested in duplicates) in a microplate format. In addition, 25 µl of bovine serum albumin (BSA) Standards ranging from 20–2000 µg/ml was used. The samples were incubated for 30 min at 37°C and the absorbance was read using a DYNAtech MRX plate reader at 550 nm. A standard curve was prepared by plotting the blank-corrected absorbance for

each BSA standard versus its concentration and was used to determine the protein concentration of each protein sample.

2.8.3 SDS-PAGE

The protein preparations were analyzed by sodium dodecyl sulphate-polyacrylamide gel (SDS-PAGE) electrophoresis (Laemmli *et al.,* 1970). The separation and stacking gels had a concentration of 10% and 4%, respectively. Gels were made with 30% acrylamide/bis solution, 10% sodium dodecyl sulphate (SDS), 10% ammonium persulphate and TEMED (all the reagents are from BioRad, Hertfordshire, UK). Protein samples were mixed with SDS-PAGE sample buffer (0.25 M Tris-HCl pH 6.8, 8% SDS, 30% Glycerol, 0.02% Bromophenol Blue, 10% β-mercaptoethanol) and heated at 100°C for 5 min. The protein samples were loaded on the SDS-PAGE gel alongside Precision plus Protein standards dual colour marker (BioRad, Hertfordshire, UK). The gels were run at 100 V for 2h at room temperature in a 1x running buffer made from 10x stock (0.25 M tris, 1.92 M Glycine, 10% SDS). The gels were then stained with Coomassie blue solution (0.2% Coomassie Blue, 7.5% Acetic Acid, 50% Ethanol) for one hour and then de-stained in the Coomassie destaining solution (20% methanol, 5% glacial acetic acid) for protein visualization.

2.8.4 Protein transfer

Proteins were transferred from the acrylamide gel to a Polyvinylidene Fluoride (PVDF) membrane (Thermo Fisher scientific, Leicestershire, UK) using a Trans-Blot SD Semi-dry Transfer cell (Biorad, Hertfordshire, UK) at 25 V for 30 min. The gel and membrane were soaked in Transfer buffer (25 mM Tris pH8.5, 0.2 M Glycine, 20% Methanol) prior to transfer.

2.8.5 Immunolabelling

After transfer, the membrane was blocked in 1x TBST (Tris-Buffered Saline Tween-20) (25 mM Tris, 150 mM NaCl, 0.05% Tween 20) containing 5% skimmed dried milk for 1 h. The membrane was then washed 3 times for 5 min with TBST and incubated with the primary antibody 9E1 (1/100 dilution) to label the membrane overnight at 4°C. The membrane was then washed 4 times for 15 min with TBST to remove the excess of unbound antibody. The membrane was then incubated with the secondary antibody Goat anti-mouse conjugated to horseradish peroxidase (HRP) (Jackson Immuno Research, Stratech Scientific Ltd, Suffolk, UK) diluted 1:20,000 in TBST containing 5% skimmed dried milk for 1 hour at room temperature. After incubation, the membrane was washed 4 times for 15 min in TBST and finally in PBS (phosphate buffered saline). Alternatively, for the detection of the recombinant protein, a monoclonal anti- polyhistidine peroxidase conjugate antibody (Sigma, Dorset, UK) was used. The membrane was incubated with the antibody diluted 1:10,000 in TBST containing 5% skimmed dried milk for 2 hours at room temperature and then washed 4 times for 15 min in TBST and rinsed in PBS.

The detection of the HRP signal was performed by adding 1 ml per membrane of Supersignal® West Pico Chemiluminescent Substrate (Pierce, Thermo Fisher Scientific, Leicestershire, UK) for 5 min. The reaction was revealed on an X-ray film developed using SRX-101A X-ray processor (Konica Minolta, Banbury, UK).

2.8.6 Reprobing of the membrane

When reprobing of the labelled membrane was desirable, the membrane was washed in TBST for 5 min and then in deionised water for 5 min before being incubated in a 0.2 M sodium hydroxide solution for 5 min. The membrane was then washed in deionised water for 5 min and blocked in TBST solution containing 5% skimmed dried milk. The immunolabelling was then performed as described in 2.8.5 using a different antibody.

2.9 Immunofluorescence assay (IFA)

2.9.1 Oocysts suspensions

Purified *Cryptosporidium* oocyst suspensions were used for IFA testing. *C. parvum* Moredun oocysts (Moredun Research Institute, Midlothian, UK) and *C. parvum* Iowa (Bunch Grass Farm, Deary, ID) were purchased at $1x10^9$ titre. In addition, purified clinical oocyst suspensions were provided by CRU and tested by IFA. They correspond to the clinical isolates, whose DNA was tested by PCR as described in 2.3. The parasite concentration in clinical samples was determined using an improved Neubauer hemacytometer counting chamber, with 10 µl of each diluted oocyst suspension being added to the chamber. For each of the four counting grids, the number of cells was counted and an average was calculated. The grid surface is 0.1 mm below the coverslip. The volume of each grid (of 1 mm^2 area) is therefore 0.1 cubic mm. The number of cells per cubic mm is the average calculated number multiplied by 10^4. The calculated parasite titres for each suspension are summarized in table 2.6.

2.9.2 Antibodies and labelling reagents

Cryptosporidium oocyst suspensions, as detailed in 2.9.1, were stained using the monoclonal antibody (9E1) as primary antibody. Different secondary antibodies were used to detect the reaction. A Fluorescein (FITC)-conjugated Goat Anti-Mouse IgG (Jackson ImmunoResearch, Stratech Scientific Ltd, Suffolk, UK) was used at 1:200 dilution. Alternatively, a Cy™5- conjugated Donkey Anti-Mouse IgG (Jackson ImmunoResearch, Stratech Scientific Ltd, Suffolk, UK) or Alexa Fluor®546 Goat Anti-Mouse IgG (Molecular Probes, Invitrogen, Paisley, UK) were used at 1:800 and 1:200 dilutions, respectively.

***Cryptosporidium* isolate**	**Concentration (oocysts/ml)**
Ch2	$22x10^5$
Ch3	$1x10^5$
Ch4	$43x10^5$
Cp2	$2x10^7$
Cp3	$1x10^7$
Cp4	$0.5x10^5$

Table 2.6: Determination of the titre of clinical *Cryptosporidium* oocyst suspensions.

In addition, a staining protocol with two antibodies was used for improved binding specificity. Biotinylated Rat Anti-Mouse IgG (1:1000) was used as a secondary antibody and was subsequently detected using streptavidin substrate conjugated to Alexa Fluor®568 (1:100) (both reagents were purchased from Invitrogen, Paisley, UK). Furthermore, the effect of a further blocking step using Streptavidin/Biotin Blocking kit (Vector laboratories, Peterborough, UK) was also assessed.

A direct conjugation kit was utilized to prepare fluorescent conjugates of the monoclonal antibody in order to eliminate possible cross-reaction of conjugated secondary antibody with the sample. The subtype of the monoclonal antibody (9E1) was determined using Iso-Gold™ Rapid Mouse-Monoclonal Isotyping Kit (Amsbio, AMS Biotechnology Ltd., Abingdon, UK) according to the manufacturer's instructions. Protein quantitation of 9E1 was determined by absorbance at 280 nm using Nanodrop ND-1000. Zenon® Alexa Fluor®594 Mouse IgG1 Labelling kit (Molecular Probes, Invitrogen, Paisley, UK) was used according to the manufacturer's instructions and the freshly labelled antibody was utilized in IFA. Alexa Fluor®594 Monoclonal Antibody Labelling kit (Molecular Probes, Invitrogen, Paisley, UK) was also used for long term stable labelling of the monoclonal antibody.

A Fluorescein isothiocyanate (FITC)-labelled genus-specific anti-*Cryptosporidium* species monoclonal antibody: Crypto-cell (TCS BioSciences Ltd., Buckingham, UK) staining the oocyst wall of *Cryptosporidium* spp. was used as a counter-stain according to the manufacturer's instructions.

2.9.3 Immunolocalization of Cops-1

Approximately 10^3 oocysts were deposited on a single welled-microscope slide (Hendley-Essex, Essex, UK), which was cleaned with methanol-soaked tissue prior to use. The parasite suspension was allowed to air-dry at room temperature and fixed by applying 20 µl of methanol. Then, 25 µl of the monoclonal peptide

antibody (9E1) was used to stain the oocysts in a humidified staining chamber for 1h at room temperature. The antibody excess was washed twice in PBS. The secondary antibody and/or the appropriate substrate was used at the recommended dilution as detailed in 2.9.2 and incubated for 30 min at room temperature in a humidified staining chamber. The secondary antibody was then washed twice in PBS. For directly conjugated 9E1 antibody, this step was omitted. Crypto-cell antibody (30 µl) was added to the slide (when applicable) and incubated at 37°C for 15 min. The antibody was then carefully removed and the slide was stained using 50 µl of a 1:2000 solution of 4',6-diamidino-2-phenylindole (DAPI) for 2 min. The slide was then washed in deionized water and allowed to air-dry. It was then mounted in Fluoromount-G™ (SouthernBiotech, Cambridge BioScience Ltd., Cambridge, UK) and the edges of the coverslip were sealed using clear nail polish. Zeiss AxioPlan 2 microscope (Zeiss, Hertfordshire, UK) was used for the microscopic observation of the stained slides.

2.10 Gliding assay

Approximately, 10^6 oocysts were excysted in a solution of 0.8% sodium deoxycholate at 37°C for 30 min. The mixture was then filter-sterilized (0.2 µm) to separate sporozoites from empty shells and unexcysted oocysts. 5 µl of purified sporozoite suspension was deposited on a SuperFrost microscope slide (VWR International Ltd., West Sussex, UK). Excysted sporozoites undertake gliding motility and sporozoite proteins involved in host cell attachment and invasion are shed in these trails. The proteins present in trails are able to adhere to the slide when left to dry at room temperature. The slides were then stained using 9E1 monoclonal antibody. As a positive control, slides were stained using 4C1 monoclonal antibody, kindly provided by Dr. A. Sheoran (Tufts Cummings School of Veterinary Medicine, Massachusetts, USA), raised against a sporozoite protein deposited in trails (Feng *et al.,* 2006). A FITC-conjugated Goat Anti-Mouse IgG secondary antibody was used as described in 2.9.2. The slides were also stained with DAPI as detailed in 2.9.3. The formation of gliding trails was

revealed by microscopic observation of stained slides using Zeiss AxioPlan 2 microscope and FITC, DAPI and DIC filters.

2.11 Co-culture of *Cryptosporidium* and Caco-2 cells

2.11.1 Parasite preparation

Purified *C. parvum* oocysts of the Iowa isolate (10^9) were purchased from Bunch Grass Farm (Idaho, USA) and stored in PBS pH 7.2 containing 1000 U Penicillin and 1000 µg Streptomycin at 4°C. A 10 ml aliquot of this suspension was subjected to surface sterilization by suspension in 10% (vol/vol) commercial bleach solution (sodium hypochlorite) in PBS as previously described (Choudhry *et al.,* 2008). The parasites were washed three times in PBS and pelleted by centrifugation at 13000 rpm for 6 min before enumeration using an improved Neubauer hemacytometer counting chamber as described in 2.9.1 to evaluate the number of parasite in the purified suspension. The purified oocyst suspension was used for co-culture.

2.11.2 Excystation of the oocysts

The bile salt sodium deoxycholate (Sigma, Dorset, UK) was suspended in 0.1% wt/vol in DMEM (Dulbecco's Modified Eagle Medium) and filter-sterilized. Then, 250 µl of this solution was equilibrated to 37°C for 10 min, 10^6 oocysts were added and the suspension was incubated at 37°C in an aerobic incubator as previously described (Choudhry *et al.,* 2008). The excystation was monitored microscopically; when the majority of the oocysts (~ 80%) were excysted, the reaction was stopped by taking the reaction back to room temperature.

2.11.3 Cell culture

The Caco-2 (Human colorectal adenocarcinoma) epithelial cell line was kindly donated by Dr. N. Belshaw (Institute of Food Research, Norwich). The cells were routinely grown in T-75 Tissue culture Nunclon™ coated flasks (Nunc, Thermo Fisher Scientific, Leicestershire, UK) and maintained in DMEM supplemented with 10% heat-inactivated foetal calf serum, 10 mM L-glutamine, 100 U of penicillin/ml and 100 µg of streptomycin/ml (all reagents are from Gibco, Invitrogen, Paisley, UK). Cells were grown in an incubator at 37°C with 5% CO_2 and 95% air until confluence. The culture medium was removed and the cells were washed twice in PBS. Adherent cells were detached by adding 1 ml trypsin (Sigma, Dorset, UK) and incubation for 5-7 min at 37°C (until all the cells were detached). The trypsin action was neutralized by adding 10 ml of DMEM. The cells were homogenised by pipetting up and down. Then, 5 ml of the cell suspension was used to seed a new culture flask containing 10 ml of culture media. When a cell count was required, it was performed as described in 2.9.1. All cell culture manipulations were carried out under sterile conditions in an advanced Bio safety class II cabinet (Microflow, BioQuell, Hampshire, UK).

2.11.4 Cell monolayers growth

$2x10^5$ cells from the cell suspension as described in 2.11.3 were used to seed sterile 13 mm diameter coverslips placed in 24-well plastic tissue culture plates (Nunc, Thermo Fisher Scientific, Leicestershire, UK) in 1 ml of supplemented DMEM medium. The cells were grown to confluence over 3-5 days.

2.11.5 Infection of cell monolayers

When Caco-2 cell monolayers reached confluence, the culture medium was removed and the coverslips were washed twice with PBS before the addition of $1x10^6$ oocysts in 500 µl of serum-free DMEM. As a control, coverslips were

incubated in medium without oocysts. In some experiments, the cell monolayers were infected with excysted oocysts as described in 2.11.2. For invasion assays, the oocysts were incubated with the cells for either 3 or 24 hours at 37°C to allow host-cell attachment and invasion. After incubation, the culture medium was removed and the cells were washed twice in PBS before adding 1 ml of supplemented DMEM as described in 2.11.3. The cell monolayers were then cultured for a further 24 hours at 37°C.

2.11.6 Fixation and staining of cell monolayers

The culture medium was removed from the infected cell monolayers and the wells were washed twice with PBS to remove any remaining traces of medium. Each coverslip was incubated with 100 µl of 1% paraformaldehyde at room temperature for 15 min and then washed three times with PBS to remove any traces of paraformaldehyde. After fixation, the cells were stained with Giemsa (Giemsa stock solution diluted in 0.1 M phosphate buffer pH 6.8) for 2h at room temperature. The slides were then washed and allowed to air-dry, before being mounted. The intracellular multiplication stages of the parasite were identified by microscopic observation.

2.11.7 Inhibition of *Cryptosporidium* adhesion-invasion of the host cells *in vitro*

The role of Cops-1 protein in host-cell attachment and invasion was evaluated in a blocking assay by pre-incubation of the parasite suspension with 9E1 before coculture in Caco-2 cell monolayers. The monoclonal antibody 9E1 was incubated with $1x10^6$ oocysts from a purified suspension as described in 2.11.1. The 9E1 antibody was used at two concentrations (1mg/ml and 0.1 mg/ml). As a negative control, a mouse IgG1 isotype (AbD Serotec, MorphoSys UK Ltd, Oxford, UK) was used at 0.1 mg/ml. As a positive control, $1x10^6$ oocysts were incubated in PBS (no blocking). Antibody and parasite suspensions were incubated at 37°C for 90 min as previously described (Nishikawa *et al.,* 2000) and the mixture was used to infect confluent Caco-2 cell monolayers in duplicate as described in

2.11.5. After 3h incubation, parasite suspensions were removed and the monolayers were cultured for a further 24 h as described in 2.11.5. Coverslips were washed, fixed and stained as described in 2.11.6. The parasites infecting host cells were observed by bright field microscopy. The effect of the pre-incubation with the antibody was assessed by parasite count for each condition from 20 randomly chosen fields.

2.12 Screening of *Cryptosporidium* antigens by Western Blot using sera from cryptosporidiosis patients

2.12.1 *Cryptosporidium* antigens

The recombinant Cops-1 protein produced as detailed in 2.7.5 and 2.7.6 was screened by Western Blot using patient's sera. In addition, native antigen preparations from *C. hominis* and *C. parvum* oocysts were also screened. *C. hominis* oocysts were kindly donated by Prof. G. Widmer (Tufts Cummings School of Veterinary Medicine, Massachusetts, USA) and *C. parvum* Iowa oocysts were purchased as detailed in 2.9.1. Approximately, $2x10^6$ oocysts were concentrated by centrifugation 13000 rpm for 10 min at 4°C. The pelleted parasites were re-suspended in a small volume of PBS. The oocysts were frozen in a dry ice/ ethanol bath and immediately thawed in a water bath at 42°C. This cycle was repeated 4 times to ensure the breakage of the oocyst wall. Then, 2 µl of protease inhibitors and 2 µl of phosphatase inhibitors were added as previously described in 2.7.6 and Western Blot analysis was performed as described in 2.8.

2.12.2 Cryptosporidiosis patient sera

Native or recombinant antigen preparation proteins were tested by immunoblot using sera from *Cryptosporidium*-infected patients to check the immunogenicity of Cops-1 protein. The sera are from laboratory-confirmed cryptosporidiosis

cases kindly donated by Dr. K. Elwin (*Cryptosporidium* Reference Unit, Singleton Hospital, Swansea, UK). *Cryptosporidium* isolates were genotyped, so, it was possible to differentiate between *C. parvum* and *C. hominis* serological responses. Initially, one *C. hominis* serum and *one C. parvum* serum were tested at 1:50 dilution. These sera were previously tested at CRU and showed high level of expression of the 15/17 and 27KDa *Cryptosporidium* immunodominant proteins. In addition, a negative serum (no detected cryptosporidiosis) was tested as a negative control. The reaction was detected using a Goat Anti-Human IgG HRP-conjugated secondary antibody (Abcam, Cambridge, UK) diluted 1:20,000.

CHAPTER 3:

Identification of species-specific markers of anthroponotic *Cryptosporidium* species using comparative genomics

3.1 Introduction

At least seven *Cryptosporidium* species infect humans (Xiao and Fayer, 2008); however, only two species are of most significance for public health causing the majority of human cases both as sporadic and outbreak related cases: *C. hominis* and *C. parvum* (Cacciò and Pozio, 2006, Cacciò, 2005, Xiao and Ryan, 2004, Morgan *et al.,* 1999). *C. parvum* is zoonotic and mostly associated with human and ruminant infections, while *C. hominis* is generally restricted to humans (Robertson and Gjerde, 2007). These two *Cryptosporidium* species differ mainly in host range, but differences in geographical and temporal distribution, pathogenicity and risk factors have also been identified (Cacciò, 2005, Hunter and Thompson, 2005). Formerly, these species were considered one single species, with the anthroponotic subtype called *C. parvum* genotype 1 and the zoonotic subtype called *C. parvum* genotype 2. Subsequently, based on molecular and phylogenetic analysis, in combination with biological differences, Morgan-Ryan and colleagues (2002) suggested the creation of new species: *C. hominis* for *C. parvum* genotype 1. In addition, some *Cryptosporidium* species and genotypes, initially not associated with human infection, were shown to cause cryptosporidiosis outbreaks. These species and genotypes are therefore considered emergent. This is for example the case of the rabbit genotype, identified as the aetiological agent in a medium-sized outbreak of waterborne human cryptosporidiosis in the Northamptonshire area (Robinson and Chalmers, 2009, Chalmers *et al.,* 2009).

The public health relevance of *C. parvum* and *C. hominis* has driven a bias in *Cryptosporidium* research towards these two human infective species. In fact, genomes representatives of *C. parvum* and *C. hominis* species have been sequenced (Abrahamsen *et al.,* 2004, Xu *et al.,* 2004). In addition, the genome sequencing of *C. muris*, a less relevant *Cryptosporidium* species from a public health perspective, is underway (Widmer *et al.,* 2007). The genomic data for all three genome representatives are available online (http://CryptoDB.org).

The genome sizes for *C. parvum* and *C. hominis* are 9.11 and 9.16 Mb, respectively (Table 3.1). The GC content is ~ 30% and the coding region is ~ 6 Mb (Xu *et al.,* 2004). The number of published genes is slightly higher in *C. hominis* than in *C. parvum*: 3,994 genes versus 3,952 genes. The average gene length is comparable between the two species: 1.57 kb and 1.72 kb, for *C. hominis* and *C. parvum*, respectively (Table 3.1).

The genome analysis of *C. hominis* and *C. parvum* showed extremely streamlined metabolic pathways and a reliance on the host for nutrients (Abrahamsen *et al.,* 2004). Biosynthesis capabilities are limited, explaining an extensive array of transporters that enable import of essential nutrients from the host (Pain *et al.,* 2005, Xu *et al.,* 2004). Genome sequences showed that *Cryptosporidium* species have genes associated with apical complex organelles despite the fact that they lack an apicoplast and possess a degenerate mitochondrion that has lost its genome (Abrahamsen *et al*, 2004, Xu *et al*, 2004). The existence of a relict mitochondrion was later confirmed by ultrastructural studies (Keithly *et al.,* 2005).

About 60% of *Cryptosporidium* genes exhibit similarity to known genes. The annotations for *Cryptosporidium*, *Plasmodium* and *Saccharomyces* are remarkably similar, suggesting that their phenotypic differences are a reflection of non-conserved or previously unreported gene families of unknown function rather than to the functional specialization of conserved gene families (Xu *et al.,* 2004).

(a) The genome	*C. hominis*	*C. parvum*
Size (Mb)	9.16	9.11
No. of physical gaps	246	5
No. of contigs	1413*	n.a.
(G+C) content (%)	31.7	30.3
Coding regions†		
Coding size (Mb)	6.29	6.80
Percentage coding	69	74
(G + C) content (%)	32.3	31.9
No. of genes	3,994	3,952
Mean gene length (bp)	1,576	1,720
Gene density (bp per gene)	2,293	2,305
Genes with introns (%)‡	5–20%	5%
Hits nr§	2,331	2,483
Percentage hits nr§	58	63
Intergenic regions		
Non-coding size (Mb)	2.87	2.32
Percentage not coding	31	25
(G+C) content (%)	30.3	25.6
No. of intergenic regions	4,003	3,960
Mean length (bp)	716	585
RNAs		
No. of tRNA genes	45	45
No. of 5S rRNA genes	6	6
No. of 5.8S,18S and 28S	5	5
(b) The proteome		
Total predicted proteins	3,994	3,952
Hypothetical proteins	2,779	2,567
Gene ontology		
Biological process	1,239	n.d.
Cellular component	1,265	n.d.
Molecular function	1,235	n.d.
Structural features		
Transmembrane domain	786	n.d.
Signal peptide	421	n.d.
Signal anchor	221	n.d.

Table 3.1: Genomic and proteomic characteristics of *C. hominis* TU502 and *C. parvum* Iowa strains based on full genome sequences (reproduced from Xu *et al.,* 2004)

Several comparative genomics studies were performed after the completion of genome sequences of apicomplexan parasites of medical and veterinary importance. Templeton and colleagues (2004) showed that *Cryptosporidium* spp and *Plasmodium* spp. share over 150 ancestral "apicomplexan" proteins, mainly involved in interactions with eukaryotic host cells and the biogenesis of the apical complex. Gordon and Sibley (2005) used genome sequences of *Toxoplasma gondii, Plasmodium* spp., *Cryptosporidium* spp. and *Theileria* spp. to show the conservation of actin-like proteins among these parasites relying on actin-based motility for cell invasion. In addition, comparative genomics of *Plasmodium* spp., *Cryptosporidium* spp. and *Toxoplasma gondii* revealed that calcium-regulated proteins (plant-like pathways for calcium release channels and calcium-dependent kinases) were also conserved (Nagamune and Sibley, 2006).

Genome comparison of *C. hominis* and *C. parvum* showed a high level of similarity, which limited the ability of comparative genomics to improve annotation, identify conserved non-coding elements and study gene and protein evolution (Widmer *et al.,* 2007). More importantly, this high sequence similarity hindered better understanding of host specificity and virulence mechanisms of these *Cryptosporidium* species, as was initially anticipated from the genome projects (Widmer *et al.,* 2002). In fact, *C. hominis* and *C. parvum* genomes exhibit only 3–5% sequence divergence, with no large insertions, deletions or rearrangements (Xu *et al.,* 2004). The authors stated that the gene complements of the two species are essentially identical because the few *C. parvum* genes not found in *C. hominis* are proximal to known sequence gaps (there are 246 gaps in the *C. hominis* genome sequence). However, the total number of genes was similar: 3,994 and 3,952 for *C. hominis* and *C. parvum*, respectively (as shown in Table 3.1). The significance of the 42 missing genes is not clear. In fact, uncertainty about the amount of sequence variation between *C. parvum* and *C. hominis* persists due to the incomplete status of the *C. hominis* genome. It has been concluded, however, that the phenotypic differences between *C. hominis* and *C. parvum* are caused by polymorphisms in coding regions and differences in gene regulation (Pain *et al.,* 2005, Xu *et al.,* 2004). Nevertheless, the role of what genetic variability is present in the phenotypic differences between *C.*

hominis and *C. parvum* is now much more accessible for investigation. In fact, this subset of genes displaying the most variation may include hitherto unnoticed genetic determinants of host specificity and virulence.

3.2 Aims

The aim of this study is to identify potential genetic determinants of host tropism for *C. parvum* and *C. hominis* species. This was undertaken using a comparative genomic approach, aiming to investigate the reported 3-5% sequence divergence between the two species. These areas of high sequence variation in the coding regions are putatively species-specific genes. These genes could be responsible for the phenotypic and pathogenic differences between these two important human infective *Cryptosporidium* species.

3.3 Results

3.3.1 Reciprocal BLAST

Based on the published genome sequences for *C. parvum* and *C. hominis* available on CryptoDB database, the Reciprocal BLAST tool was used to identify putative species-specific genes as described in 2.1.1. This comparative genomic tool was used to assign a similarity score to every *C. hominis* gene when tested by BLAST against the genomic database of *C. parvum.* Similarly, each coding sequence from *C. parvum* was tested by BLAST against *C. hominis* genomic sequences and a similarity score was assigned. Putatively species-specific genes correspond to regions of high sequence variation; therefore, a threshold of 10% sequence identity was adopted. Only genes showing a level of identity below 10% were considered. This initial screening allowed the identification of 117 and 272 putative species-specific genes for *C. hominis* and *C. parvum*, respectively (Appendix II). The majority of *C. parvum* putative specific genes were annotated, while *C. hominis* selection corresponded mainly to hypothetical proteins. All the putative species-specific genes were individually tested

using Blastn algorithm as described in 2.1.1. In addition, the genomic database CryptoDB was used to identify orthologs in other *Cryptosporidium* species. This secondary screening was undertaken to increase the prediction stringency. The number of the predicted putative species-specific genes was 93 and 211 genes for *C. hominis* and *C. parvum*, respectively (Appendix III).

3.3.2 PCR testing of putative species-specific genes in *C. hominis* and *C. parvum* strains

The specificity of a subset of putative species-specific genes identified *in silico* was assessed experimentally by PCR in a collection of *C. hominis* and *C. parvum* clinical isolates obtained from the *Cryptosporidium* reference Unit as described in 2.3. In addition, DNA of three reference strains was also tested: *C. hominis* TU502, *C. parvum* Iowa and *C. parvum* Moredun. For each gene, primers were designed and PCR was carried out as described in 2.2. PCR products were purified and sequenced in both directions and sequence analysis was performed as described in 2.4.

Initially, ten putative species-specific genes were tested by PCR. Surprisingly, 90% (9/10) of the genes tested were common to *C. hominis and C. parvum*. Figure 3.1 shows the PCR results for the amplification of cgd2_80 and chro.50330 genes. There was no discernable difference between PCR results of clinical isolates and reference strains by agarose gel electrophoresis. All isolates DNA was amplified by PCR, except for one isolate (Cp4), which did not amplify using Chro.30149 primers.

Further testing of additional predicted species-specific genes confirmed the general trend, therefore, the majority of the genes identified *in silico* seems to be common to both *Cryptosporidium* species.

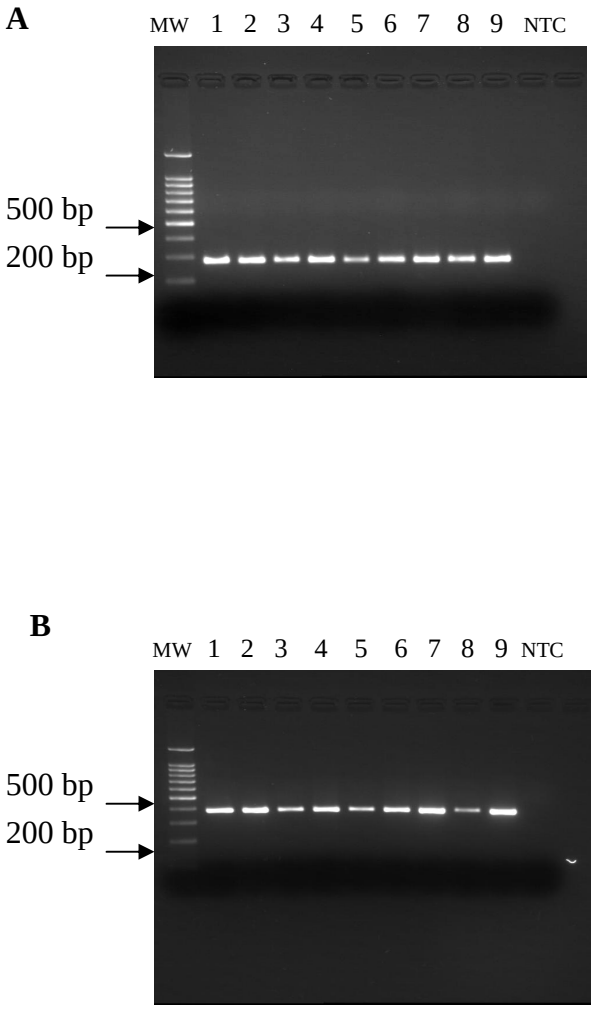


Figure 3.1: Amplification of *Cryptosporidium* DNA from clinical isolates and reference strains. A: amplification of 266 bp of Cgd2_80 gene, B: amplification of 368 bp of Chro.50330 gene. Both *Cryptosporidium* species and all isolates were PCR positive. MW: molecular weight, 1: Cp2, 2: Cp3, 3: Cp4, 4: Ch2, 5:Ch3, 6: Ch4, 7: Iowa, 8: Moredun, 9: TU502, NTC: non template control.

3.3.3 PCR testing of putative species-specific genes in other *Cryptosporidium* species

Whether the observed ubiquity of the predicted specific genes represented the closeness between *C. hominis* and *C. parvum* or whether these genes are orthologous in other *Cryptosporidium* species was assessed by experimentally testing DNA from other *Cryptosporidium* species and genotypes. *C. andersoni, C. felis*, cervine genotype, *C. meleagridis* and *C. baileyi* DNAs were kindly donated by CRU and used for PCR screening as described in 2.3 using primers designed to amplify the ten novel genetic loci.

C. meleagridis DNA was amplified by PCR for 8/10 genes (80%). Only, Cgd2_2430 and Chro.20156 PCR reactions were negative (Table 3.2). Interestingly, for Cgd2_2430 gene, only *C. andersoni* DNA was amplified by PCR. For Cgd6_5020, only *C. felis* DNA was PCR positive and for Chro.30149 primers, cervine genotype DNA was amplified. *C. andersoni*, cervine genotype and *C. felis* DNA was amplified by 10% (1/10) of primers tested. *C. baileyi* DNA was not amplified by any of the primers tested.

3.3.4 Sequence analysis of PCR products

All positive PCR products were purified and sequenced using forward and reverse primers in two independent sequencing facilities as described in 2.4. The alignments of PCR product sequences for each gene are shown in Appendix IV. PCR product sequences were submitted to GenBank and are available online (GU904212-GU904405). One PCR product of *C. meleagridis* DNA using Chro.50330 primers did not give good sequence and was therefore excluded from the analysis. PCR products for *C. andersoni*, *C. felis* and cervine genotype, as mentioned in 3.3.3, did not yield good quality sequences and they were not included in the analysis.

	C. andersoni	*C. felis*	Cervine genotype	*C. meleagridis*	*C. baileyi*
Cgd2_80	-	-	-	+	-
Cgd2_2430	+	-	-	-	-
Cgd6_200	-	-	-	+	-
Cgd6_5020	-	+	-	+	-
Cgd8_2370	-	-	-	+	-
Chro.20156	-	-	-	-	-
Chro.50317	-	-	-	+	-
Chro.50330	-	-	-	+	-
Chro.30149	-	-	+	+	-
Chro.50457	-	-	-	+	-

Table 3.2: PCR results from other *Cryptosporidium* species using newly designed primers amplifying fragments from the ten novel genetic loci. PCR product sizes are shown in table 2.1.

The majority of the genes tested were common to both *Cryptosporidium* species. Not surprisingly, the sequences in the different strains were highly conserved, but some polymorphism was seen mainly as Single nucleotide polymorphisms (SNP)s. A total of 78 SNPs were detected from 4150 nucleotides, this corresponds to an average of 1 SNP every 53 bp. The details of the position and type of SNPs for each gene are shown in Appendix V. The SNP result is summarized in Table 3.3. The number of SNPs was variable for each gene, ranging from 1 SNP every 30 bp for Cgd2_2430 to less than one SNP per 330 bp for Chro.30149 (Table 3.3).

Interestingly, 78.3% (61/78) of the SNPs were species-specific, thus defining an interesting feature of this subset of genes identified by comparative genomics. The proportion of species-specific SNPs ranged from 66.7% for Cgd8_2370 and Chro.50317 genes to 100% for Chro.50330 and Chro.50457. The majority of the SNPs detected (64.2%) were synonymous, thus maintaining the protein sequence. The non synonymous SNPs were not evenly distributed between the loci. In general, the proportion of the non synonymous SNPs was low for the majority of the genes ranging from 0% to 25% for Chro.50330 and Cgd6_200, respectively (Table 3.3). On the contrary, for Chro.50317 and Chro.20156 genes, 66.7% and 83.4% of the SNPs were non synonymous. In addition to 61 species-specific SNPs allowing discrimination between *C. hominis* and *C. parvum*, the sequence analysis revealed five SNPs specific for rabbit genotype isolates and three SNPs specific for anthroponotic *C. parvum* subtype (Appendix V).

Some of the newly identified species specific SNPs were predicted to form restriction sites, which was confirmed experimentally by PCR-RFLP. The sequences alignments were used to identify differential restriction sites between the main species tested. For example, the sequence alignment of Cgd6_200 PCR products showed a SNP at position 202 (C→ T) in *C. hominis* isolates, which creates a *ClaI* restriction site (ATCGAT). *C. hominis* and *C. parvum* PCR products were digested with *ClaI* and the digestion profile is presented in Figure 3.2. For *C. hominis*, the 447 bp PCR product was digested into two fragments of 210 bp and 240 bp, as predicted, while *C. parvum* PCR product lacks the restriction site and remained undigested.

Gene name	Gene annotation	PCR product size	Number of SNPs detected	Average number of nucleotides per SNP	Number of Species specific SNPs (%)	Number of non synonymous SNPs (%)
Cgd2_80	ABC transporter family protein	266 bp	7	38	6 (85.5%)	1 (14.3%)
Cgd2_2430	Ximpact ortholog conserved protein seen in bacteria and eukaryotes	389 bp	13	30	9 (69.3%)	3 (23.1%)
Cgd6_200	Oocyst wall protein 8	447 bp	8	56	6 (75%)	2 (25%)
Cgd6_5020	Protein with WD40 repeats	271 bp	2	136	2 (100%)	1 (50%)
Cgd8_2370	Adenosine kinase like ribokinase	685 bp	12	58	8 (66.7%)	1 (8.4%)
Chro.20156	Hypothetical protein	247 bp	6	42	5 (83.4%)	5 (83.4%)
Chro.50317	RNA polymerase A/beta'/A'' subunit	752 bp	15	51	10 (66.7%)	10 (66.7%)
Chro.50330	Leucyl tRNA synthetase	368 bp	3	123	3 (100%)	0 (0%)
Chro.30149	Ubiquitin-protein ligase 1	331 bp	0	331		
Chro.50457	Erythrocyte membrane-associated antigen	394 bp	12	33	12 (100%)	5 (41.7%)

Table 3.3: Summary of genetic polymorphism detected by PCR product sequence analysis of 10 genetic loci from *C. parvum* and *C. hominis* isolates. 78 SNPs were detected, 78.3% of which were species specific. For each gene, the number and proportion of species-specific SNPs were provided. The effect of the genetic polymorphism on amino acid composition was also indicated.

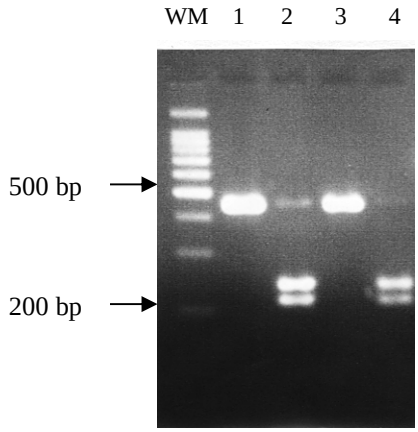


Figure 3.2: PCR-RFLP using *ClaI* restriction enzyme of Cgd6_200 gene PCR products. The *C. hominis* PCR product is digested into 2 fragments of 210 bp and 240 bp; however, *C. parvum* PCR product is not digested. MW: molecular weight, 1: Cp2, 2: Ch2, 3: Cp3, 4: Ch3,

3.3.5 Multi-locus analysis (MLA)

For MLA analysis, 11 gene sequences were included. Ten genes were derived as in 3.3.3, in addition to sequences of the COWP gene. The retrieved sequences allowed comparison of a total of 4469 bp. A Neighbor-Joining Tree was generated using MEGA software. The tree showed clear discrimination between *C. parvum* and *C. hominis* isolates (Figure 3.3 (A)). Within each group, there were two clusters corresponding to isolate subtypes. *C. parvum* anthroponotic subtype isolates (W65, W66, W67, W70) were separated from the other *C. parvum* isolates and the rabbit genotype isolates were segregated from the *C. hominis* isolates. All groups and clusters were supported by high bootstrap values. Unweighted Pair Group Method with Arithmetic Mean (UPGMA) phylogenetic method was also tested to construct a phylogenetic tree, which gave identical topology with similar bootstrap values.

Within each main group, there was no discrimination between the different isolates, despite distinct GP60 subtypes. However, TU502 strain showed some sequence divergence and was grouped separately within the *C. hominis* cluster. This is due to the presence of a unique SNP at position 132 on Cgd8_2370 gene (Appendix V). This SNP was confirmed by three independent rounds of sequencing reactions.

C. meleagridis DNA did amplify for 8/10 loci tested as detailed in 3.3.3. *C. meleagridis* sequences were included in MLA and used as an out group for the construction of a Neighbor-Joining phylogenetic tree as described above. Sequences from Cgd8_2370 and Chro.50330 genes did not generate high quality sequences. Therefore, the differences between this strain and the other isolates were based only on 2853 bp comparisons for seven genetic loci. The phylogenetic tree having *C. meleagridis* as an out group showed an identical clustering of the isolates tested with similar bootstrap values (Figure 3.3 (B)).

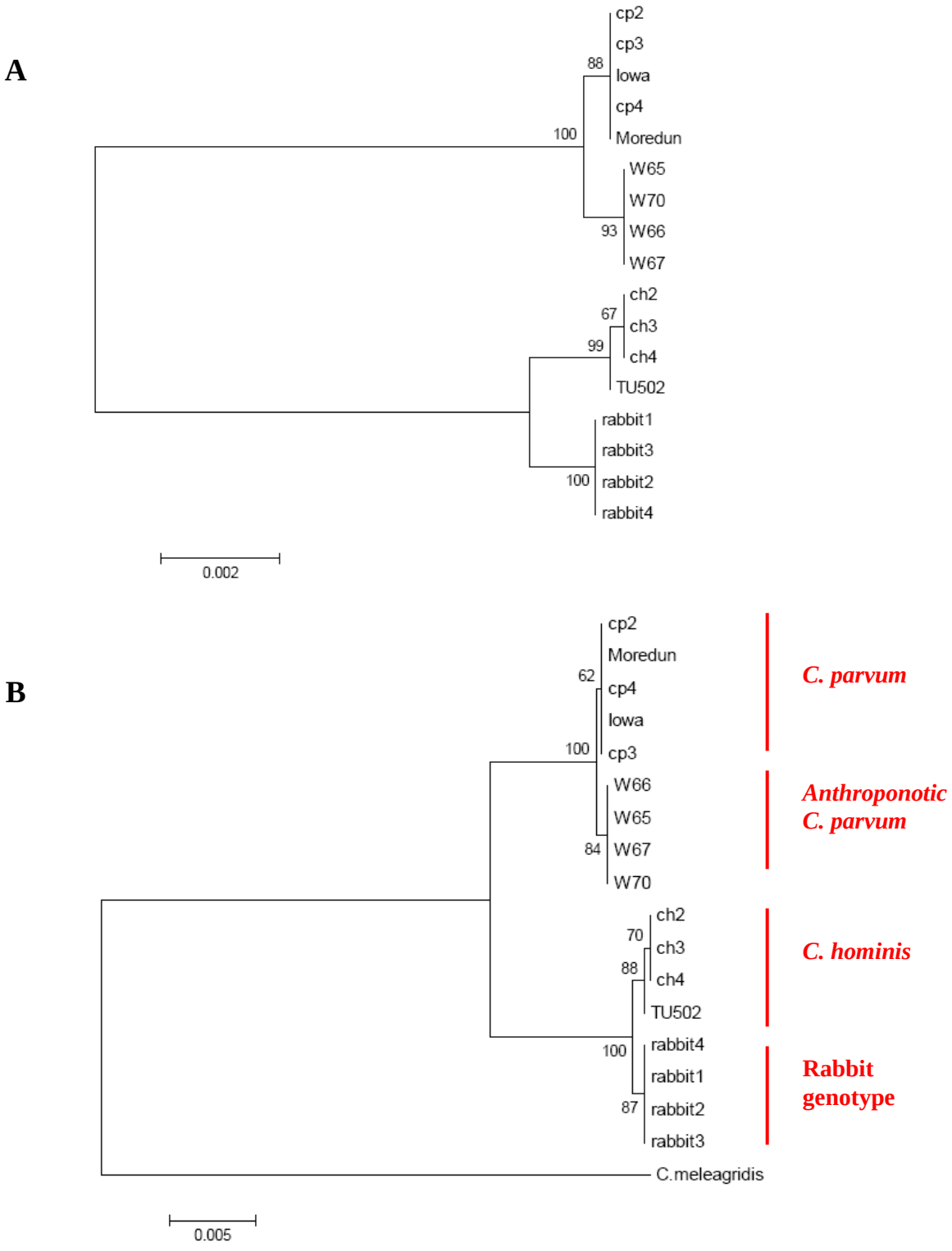


Figure 3.3: Phylogenetic Tree based on the gene sequences of 10 new loci and the COWP gene sequence. The trees were constructed using Neighbor- Joining algorithm of MEGA software. A: Phylogenetic tree constructed with *C. meleagridis* sequences. B: Phylogenetic tree with *C. meleagridis* as an out-group. *Cryptosporidium* species and subtypes for each cluster are shown in red.

3.3.6 SNP analysis

The SNP analysis was performed using the logical function "IF" of the Microsoft Excel software to discriminate between variables. Isolate groups and subtypes were compared in a pair-wise manner. When the SNPs are identical between the two groups, the value "0" is attributed, while if the two SNPs are different, the value "1" is assigned. Then, the values are summed for each group to represent the genetic variability between the main isolate groups. The calculated values are presented in Table 3.4. The newly identified SNPs showed a clear genetic difference pattern between species and subtypes of *Cryptosporidium* and the number of base pair differences between the groups is shown in Table 3.4.

It is noticeable that the genetic differences of *C. hominis* and *C. parvum* to *C. meleagridis* were comparable (5.50 and 5.05%, respectively). This is supported by minimal genetic variability between *C. hominis* and *C. parvum* (1.72%). Interestingly, the genetic difference between *C. parvum* and *C. parvum* anthroponotic subtype was 0.13%, while a slightly higher genetic difference was observed between *C. hominis* and rabbit genotype isolates (0.27%).

	C. hominis	*C. parvum*	Anthroponotic *C. parvum*	Rabbit genotype	*C. meleagridis*
C. hominis	0				
C. parvum	77 (1.72%)	0			
Anthroponotic *C. parvum*	78 (1.75%)	5 (0.12%)	0		
Rabbit genotype	12 (0.27%)	75 (1.68%)	76 (1.70%)	0	
C. meleagridis	157 (5.50%)	144 (5.05%)	144 (5.05%)	155 (5.50%)	0

Table 3.4: Genetic differences (number and percentage of base pair polymorphisms) between the main groups and subtypes of *Cryptosporidium* tested.

3.3.7 *C. parvum* specific gene (Cops-1)

Among the subset of genes tested for specificity by PCR, only one gene was *C. parvum* specific as predicted by comparative genomics. This gene Cgd2_4380, gave a correct size PCR product (665 bp) from *C. parvum* DNA but not from *C. hominis* DNA (Figure 3.4). Different PCR conditions were tried (different primer concentration and lower annealing temperature) to allow amplification of *C. hominis* DNA with no success. The PCR product was negative from rabbit genotype DNA. However, when other *Cryptosporidium* spp. DNA was tested, only, *C. meleagridis* gave amplification.

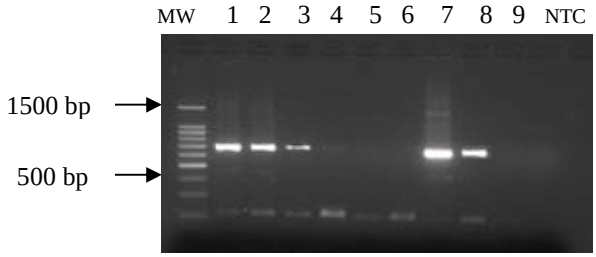


Figure 3.4: Amplification of 665 bp of Cgd2_4380 gene from *C. parvum* isolates and reference strains DNA. *C. hominis* DNA did not amplify. MW: molecular weight, 1: Cp2, 2: Cp3, 3: Cp4, 4: Ch2, 5: Ch3, 6: Ch4, 7: Iowa, 8: Moredun, 9: TU502, NTC: non template control.

3.3.8 *C. hominis* specific gene (Chos-1)

Among the initial genes tested by PCR, none was *C. hominis* specific. Subsequent testing of additional putatively specific *C. hominis* genes allowed the identification of one *C. hominis* specific gene: Chro.50011. The primers enabled amplification of the right product size (287 bp) from *C. hominis* isolates but not from *C. parvum* DNA (Figure 3.5). Different conditions were tried to optimize the amplification of *C. parvum* DNA (as described in 3.7.4), but were unsuccessful. The rabbit genotype DNA gave the same profile as *C. hominis* DNA (not shown). Testing of other *Cryptosporidium* spp. DNA (*C. andersoni*, *C. felis*, cervine genotype, *C. meleagridis* and *C. bailey)* did not allow amplification even from *C. meleagridis* DNA.

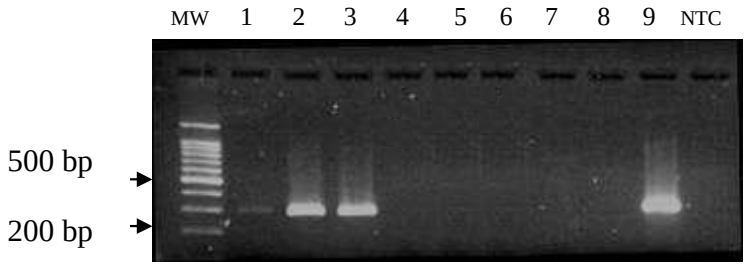


Figure 3.5: Amplification of 287 bp of Chro.50011 gene from *C. hominis* clinical isolates and reference strains DNA. No amplification of *C. parvum* DNA was possible.
MW: molecular weight, 1: Ch2, 2: Ch3, 3: Ch4, 4: Cp2, 5: Cp3, 6: Cp4, 7: Iowa, 8: Moredun, 9: TU502, NTC: non template control.

3.4 Discussion

In this study, comparative genomic tools were used to identify putative species-specific genes for *C. hominis* and *C. parvum*. The bioinformatic primary and secondary screening allowed the identification of 93 and 211 genes for *C. hominis* and *C. parvum*, respectively. Initially, ten of these genes were tested by PCR in a collection of *Cryptosporidium* clinical isolates and reference strains. The PCR screening of the putative species-specific genes showed that the majority of the genes were not as predicted. In fact, 90% of the genes tested were common to *C. hominis* and *C. parvum* isolates and some were present in *C. meleagridis*. This result reflects the quality and the status of the published genome sequences. *C. hominis* TU502 genome is neither completed nor fully assembled, which clearly hindered the accuracy of the comparative genomic analysis performed. The majority of the predicted genes are not specific and may correspond to sequence gaps reported (Xu *et al.,* 2004).

Further testing of an additional ten genes confirmed the general trend of orthology. An improved comparative genomic analysis has been made possible by the fast progress made towards the completion of *C. muris* genome. In fact, 8.9 Mb from the *C. muris* genome has recently been made available for download from CryptoDB database, of which 7.2 Mb corresponds to coding sequences. Based on these newly added genomic sequences, 58.4% of the selected putative species-specific genes appear to have orthologs in *C. muris*. This information, if known previously, would have decreased dramatically the number of putative species-specific genes predicted by comparative genomics, which would assist the successful discovery of genetic determinants of host tropism.

In order to verify whether these common genes reflected the closeness between *C. hominis* and *C. parvum* or whether they are orthologous genes common to other *Cryptosporidium* species. *C. andersoni*, *C. felis*, cervine genotype, *C. meleagridis* and *C. baileyi* DNA were also tested. *C. meleagridis* DNA amplified using 80% of the primers tested, *C. andersoni*, cervine genotype and *C. felis* DNA amplified with only 10% of primers. This result is in accordance with the taxonomy and evolution of *Cryptosporidium* species (Xiao *et al.,* 2004). In fact, *C. meleagridis* is the closest species to the cluster formed by *C. hominis*, *C. parvum* and rabbit genotype based on partial SSU rRNA gene (Xiao *et al.,* 2004). *C. meleagridis* DNA did not amplify using primers for Cgd2_2430 and Chro.20156 genes. The lack of amplification could be explained by either nucleotide mismatch in the primer region or the absence of these genes.

PCR product sequence analysis of these novel genetic markers showed interesting genetic variation as Single Nucleotide Polymorphisms (SNPs). 78 SNPs were detected, 78.3% (61) of which were species-specific. The presence of species-specific SNPs was reported previously for several genetic markers and was exploited for *Cryptosporidium* genotyping and subtyping (Sulaiman *et al.,* 1999). PCR-RFLP of the 18S rRNA (Xiao *et al.,* 1999), the COWP gene (Spano *et al.,* 1997), the DHFR gene (Gibbons *et al.,* 1998), TRAP-C1 (Spano *et al.,* 1998) and TRAP-C2 gene (Sulaiman et el., 1998), Polythreonine (Poly-T) repeats (Carraway *et al.,* 1997) and HSP70 (Gobet and Toze, 2001) allowed discrimination between *Cryptosporidium* species from various sources.

The majority of the SNPs detected (64.2%) were synonymous. It has long been assumed that synonymous SNPs are inconsequential as the primary sequence of the protein is preserved. However, it has been demonstrated that synonymous mutations can alter the structure, function and expression level of the protein by

affecting messenger RNA splicing, stability, protein folding and structure (Hunt *et al.,* 2009).

C. meleagridis PCR product sequence analysis allowed data enrichment. In fact, *C. meleagridis* species have 108 additional SNPs, 20 of which are in the Chro.30149 gene (Appendix V). For this gene, *C. meleagridis* has in average 1 SNP every 15 nucleotides. Interestingly, no SNPs were detected in this gene from *C. hominis* and *C. parvum* comparison. Chro.30149 has a predicted function as Ubiquitin ligase. Surprisingly, all the detected SNPs are synonymous. This gene, being a housekeeping gene, shows a low level of sequence divergence between species and isolates when compared to contingency genes consistently under environmental pressure and characterized by high spontaneous recombination rates (Barry *et al.,* 2003), such as the hypervariable GP60 locus used for *Cryptosporidium* typing (Widmer, 2009).

The new SNPs uncovered at ten loci allowed the construction of a robust and novel multi-locus analysis (MLA). The Neighbor-Joining phylogenetic tree constructed clearly grouped and discriminated with high bootstrap values the previously described lineage of anthroponotic *C. parvum* from *C. parvum* and *C. hominis* from the rabbit genotype. These novel genetic loci can potentially be used for genotyping purposes and to expand the genetic markers already used for epidemiological analysis.

In addition, the newly identified SNPs were used to determine genetic differences between the main *Cryptosporidium* species and subtypes tested. This analysis showed that the genetic difference between *C. hominis* and *C. parvum* was only 1.72%. When compared to *C. meleagridis*, the genetic differences were estimated to be 5% for *C. hominis* and *C. parvum*. Within the *C. parvum* group, the anthroponotic subtype isolates showed only a 0.12% difference from the main zoonotic *C. parvum* isolates. Similarly, rabbit genotype isolates exhibited only 0.27% genetic differences

to *C. hominis* isolates. This closeness between *C. hominis* and rabbit genotype was also observed in the commonly used genotyping loci (SSU rRNA, 70 kDa HSP, Actin and COWP), which showed an extremely low sequence variability (> 0.51%).

In this study, ten novel genetic markers were studied for the first time. PCR screening and sequencing of genes experimentally found to be common to both species provided *de novo* sequence information at incomplete regions of the *Cryptosporidium* genome projects and was used to examine polymorphism in these regions. The sequence analysis revealed several interesting species-specific SNPs that can be exploited for genotyping and subtyping purposes. In addition, these results gave insights about the variability and the population structure of the human infective *Cryptosporidium* species circulating in the UK.

It is anticipated that more genomic data from a multitude of *Cryptosporidium* isolates from different species and genotypes will be available shortly. This is feasible due to the major biotechnological advances, mainly next generation sequencing. This technology would enable extensive comparative genome sequence surveys, which in conjunction with *in vitro* and *in vivo* studies, would improve the understanding of this important protozoan parasite and would assist in the search for novel intervention strategies against human cryptosporidiosis (Jex and Gasser, 2010, Striepen and Kissinger, 2004).

The aim of this chapter was to identify putative species-specific genes for *C. hominis* and *C. parvum*. However, the majority of the predicted species-specific genes were found to be common to *C. parvum* and *C. hominis*. Despite this limitation, Evidence was found that one *C. parvum* gene (Cops-1) and one *C. hominis* gene (Chos-1) generated the appropriate PCR product from only one species. This result suggests that within the predicted genes, a small proportion may be species specific. No PCR amplification was possible from the other species even after relaxing PCR

conditions. A second pair of primers targeting a different region within the coding sequence for each gene was designed and tested and the PCR results were comparable between the two set of primers.

Cops-1 and Chos-1 are the only genes confirmed experimentally to be species-specific and were identified after testing 6/211 (2.85%) of the predicted *C. parvum* specific genes and 16/93 (17.2%) of the putatively *C. hominis* specific genes. This result questions the accuracy of the *in silico* prediction, which is likely attributable to incomplete status of the *C. hominis* genome. As discussed above, comparative genomic analysis with the three *Cryptosporidium* species: *C. hominis*, *C. parvum* and *C. muris* would improve future analysis and offer realistic chances for *in silico* identification of genetic determinant of host tropism.

The putative species-specific genes identified in this study could be exploited for diagnosis purposes to discriminate between *C. hominis* and *C. parvum* in a duplex PCR assay. This technique could discriminate between these two relevant *Cryptosporidium* species in a single reaction. However, this technique would need to be tested for sensitivity and specificity before realistically considering its potential as a diagnostic tool. In addition, the sequence of these species-specific loci should be determined in several clinical and environmental isolates to evaluate the genetic polymorphism.

If these putative species-specific genes were proven to be determinants of host tropism, they are likely to be contingency genes under selection pressure reflecting the characteristics of host-parasite interaction in the preferred niche. Characterization of the newly discovered species-specific genes warrants further investigation and would reveal the relevance of these findings. Additionally, the determination of the biological role of the encoded proteins, focussing on the involvement in host-parasite interaction should be assessed. Finally, screening of all

predicted species-specific genes may uncover additional undiscovered genetic determinants potentially involved in host tropism and virulence.

3.5 Summary

In this study, comparative genomic tools were used to identify putatively species-specific genes for human infective *Cryptosporidium* species. A subset of these genes identified *in silico* was tested experimentally by PCR using a collection of clinical isolates and reference strains. The majority of these genes were common to both species. This was attributable to the incomplete and partially assembled state of *C. hominis* genome, which had impaired the accuracy of the *in silico* prediction. The fast progress towards the finishing of the gastric *C. muris* genome would allow better comparative genomics and would offer insights about the evolution, host tropism and pathogenicity of *Cryptosporidium* species.

PCR product sequences revealed interesting SNPs, the majority of which were species specific. Sequence analysis of these novel genetic loci allowed the construction of a robust and novel multi-locus analysis (MLA). The Neighbor-Joining phylogenetic tree clearly grouped and discriminated with high bootstrap values the *Cryptosporidium* species and subtypes tested.

Despite this limitation, evidence was found for one *C. parvum* (Cops-1) and one *C. hominis* (Chos-1) putative species-specific gene. These genes could be used as diagnostic targets. The evaluation of the potential of these genes as species determinant and virulence factors should be carried out to validate the relevance of these findings.

CHAPTER 4:

Whole Genome Amplification

4.1 Introduction

The availability of adequate amounts of high quality genomic DNA is essential for several molecular diagnostic techniques and typing methods, particularly PCR. Genotyping and subtyping of clinical and environmental isolates is desirable as it allows source tracking and improved understanding of molecular epidemiology and population structure (Smith *et al.,* 2007, Burgos *et al.,* 2004, Mallon *et al.,* 2003, Anderson *et al.,* 2000, Han *et al.,* 2000). Unfortunately, such isolates are often unculturable, yielding a very limited amount of DNA for study. This is the case of *Cryptosporidium*, which is particularly difficult to propagate. Although *Cryptosporidium* oocysts are excreted in high numbers in the faeces during acute clinical episodes (Goodgame *et al.,* 1993), the purification methods used to obtain clean DNA, suitable for downstream molecular methods, usually result in losses. Thus, *Cryptosporidium* DNA, especially from sub-clinical infections and environmental samples is considered a precious resource. These issues have limited biological studies to *C. parvum* calf-propagated strains, particularly Iowa reference strain (Cama *et al.* 2006).

Several subtyping techniques have been applied to *Cryptosporidium* species using different markers: glycoprotein GP60 (Strong *et al.,* 2000, Leav *et al.,* 2002), double stranded RNA element (Leoni *et al.,* 2003) and mini and microsatellite repeats (Cacciò *et al.,* 2000, Mallon *et al.,* 2003). Ideally, each new isolate should be tested using a panel of markers. However, this is usually limited by the amount of DNA available. Whole genome amplification (WGA) can be used to increase the amount of nucleic acid available from clinical and environmental samples of waterborne pathogens (reviewed in Bouzid *et al.,* 2008) and application of this technique to *Cryptosporidium* isolates should address the perceived need for multi-locus typing (Smith *et al.,* 2006).

The first described WGA methods were degenerate oligonucleotide primed PCR (DOP-PCR) (Telenius *et al.,* 1992, Cheung and Nelson, 1996) and primer extension preamplification (PEP) (Zhang *et al.,* 1992). However, these PCR-based techniques produced short products (< 3 kb) and were limited by substantial amplification bias and incomplete coverage of genetic markers (Park *et al.,* 2005, Dean *et al.,* 2002, Hawkins *et al.,* 2002, Paunio *et al.,* 1996).

New strategies for WGA have been developed including multiple displacement amplification (MDA) and OmniPlex WGA (Park *et al.,* 2005). MDA is an isothermal amplification using degenerate hexamers and the bacteriophage phi-29 DNA polymerase, which possesses high processivity, strand-displacement abilities and a proofreading activity resulting in error rates 100 times lower than the Taq polymerase (Dean *et al*, 2002, Nelson *et al.,* 2002, Hawkins *et al.,* 2002, Eckert and Kunkel, 1991, Esteban *et al.,* 1993). MDA was firstly described by Blanco *et al.* (1989) and then used for WGA of different targets such as lymphoma and leukemia clinical specimens (Luthra and Medeiros, 2004), complex mixtures of DNA (Shoaib *et al.,* 2008), whole blood and tissue culture cells (Dean *et al.,* 2002), human blastomeres (Snabes *et al.* 1994), plasmid constructs and whole bacterial genomes (Detter *et al.,* 2002). The OmniPlex WGA technique uses libraries of 200–2,000 bp fragments created by random chemical cleavage of genomic DNA, followed by ligation of adaptor sequences to both ends and PCR amplification. (Bergen *et al.,* 2005, Barker *et al.,* 2004). This fragmentation/ligation/PCR-based method amplifies the entire genome several thousand-fold, and could be even re-amplified to achieve a final amplification of over million-fold without introducing inaccuracies (Langmore, 2002). Currently, several commercial kits for MDA and Omniplex based WGA are available and have been tested for several microorganisms. No trial of WGA kits for the amplification of *Cryptosporidium* DNA has been published to date.

In this study, a short supply of *Cryptosporidium* DNA and a reliance on CRU resources were the main limitations faced when testing numerous genetic loci (Chapter 3) and a solution was sought in WGA.

4.2 Aims

The aim of this study is to evaluate the suitability of WGA commercial kits for the accurate expansion of genomic DNA from *Cryptosporidium* clinical isolates DNA. This includes the ability to amplify the genomic DNA (success rate). In addition, the integrity and fidelity of the amplification is also assessed. The assessment of the WGA kits performance would determine the usefulness of the amplified DNA for downstream genotyping and subtyping applications.

4.3 Results

4.3.1 WGA kits and DNA templates concentration

Three WGA commercial kits were trialled during this study: illustra GenomiPhi V2 DNA amplification Kit, REPLI-g Ultra fast Mini Kit and GenomePlex® Complete WGA Kit as described in 2.5.1.

The performance of WGA kits was tested by evaluation of the amplification of *Cryptosporidium* genomic DNA from clinical isolates and reference strains as described in 2.5.2. Prior to WGA, each DNA sample was quantified by spectrophotometry as described in 2.7.1. DNA concentrations are shown in Table 4.1. They ranged from 3.5 to 13.2 ng/ µl (with a mean of 7.7 ng/ µl). One µl of each sample was used for WGA. The manufacturer's recommended starting genomic DNA concentration was 1 ng for the illustra kit and 10 ng for both REPLI-g and GenomePlex kits. WGA techniques are well established in the literature to give

amplification levels from 10-fold less than the lowest template concentration (0.3 ng) to 20-fold higher than the highest concentration (300 ng) (Dean *et al.,* 2002), which raises the prospect for them to be used widely for clinical samples. Thus, in this study all of the samples lie well within the range that would expect to give good amplification by the kits under test and the range recommended by the manufacturers.

4.3.2 Success rate of WGA kits

The ability of WGA kits to amplify genomic DNA from *Cryptosporidium* clinical samples was assessed by agarose gel electrophoresis, which also served as a semi-quantitative approach to assess the level of amplification. A successful amplification was considered when genomic DNA was visible on the gel. One sample (W15508) did not amplify with any of the three kits tested.

Illustra GenomiPhi and REPLI-g kits genomic amplified DNA was of high molecular weight (~10 kb) for all the tested samples. The success rate was estimated by agarose gel electrophoresis, when the genomic DNA was visible, the WGA amplification was considered successful. The success rate was 90.9% (10/11 samples) for illustra GenomiPhi kit and 45.4% (5/11) for REPLI-g kit (Figure 4.1). For the GenomePlex amplified samples, the generated DNA had a smeary aspect and was of smaller size ranging between 200 and 1,000 bp (mean size ~ 400 bp) (Figure 4.1). The success rate of the GenomePlex kit was 72.7 % (8/11 samples). For illustra GenomiPhi and REPLI-g kits, samples W15521, W15516 and Iowa showed the strongest bands, suggesting better amplification. However, when the GenomePlex kit was used, W15504, W15521, W15516 and W15517 showed a high level of amplification, but, surprisingly, commercially obtained Iowa DNA did not.

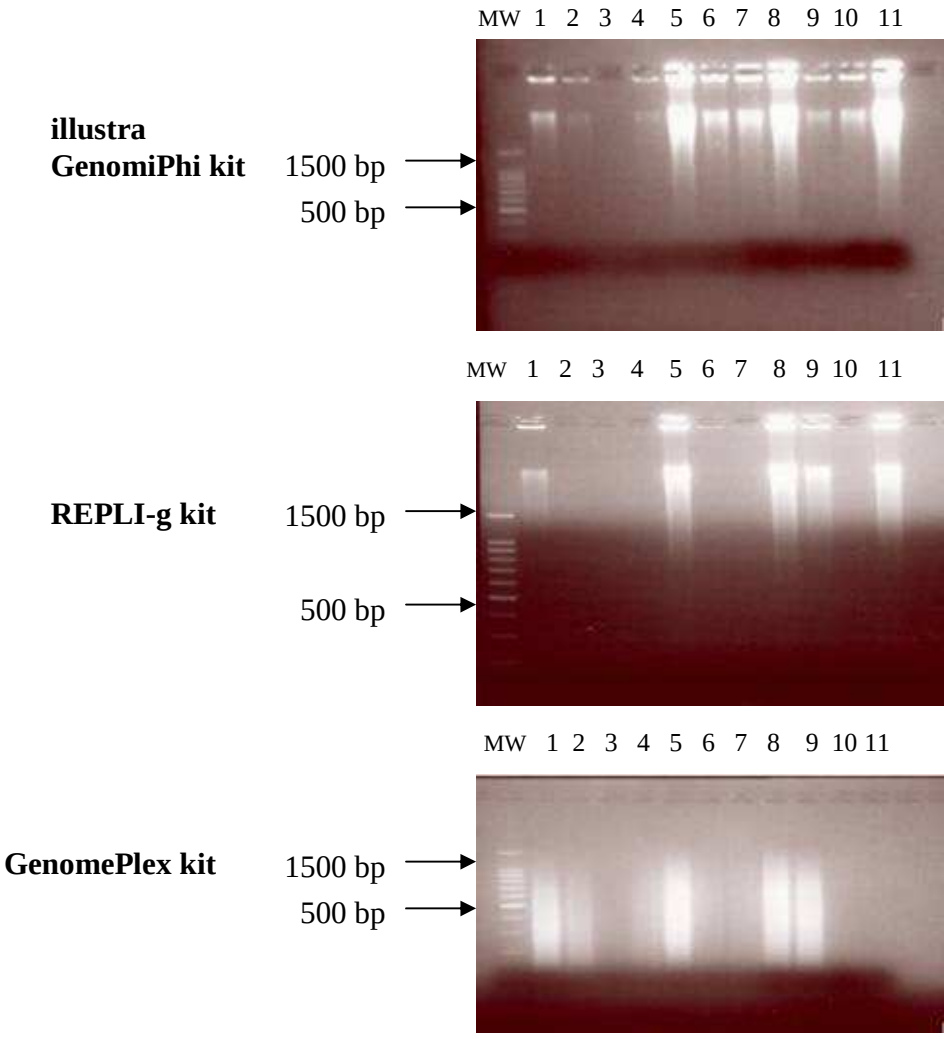


Figure 4.1: Agarose gel electrophoresis of amplified genomic DNA. MW: molecular weight marker, 1: W15504, 2: W15507, 3: W15508, 4: W15519, 5: W15521 (1-5: *C. hominis* samples), 6: W15509, 7: W15511, 8: W15516, 9: W15517, 10: W15518 (6-10: *C. parvum*), 11: *C. parvum* Iowa.

4.3.3 Yield of WGA kits

DNA quantification after WGA was initially assessed by four methods: nanodrop spectrophotometry, Hoechst and PicoGreen fluorimetry and agarose gel-based ethidium bromide fluorescence densitometry as described in 2.5.3. For the first two methods, the calculated DNA concentrations were similar in all the samples (Table 4.1) and not in accordance with the electrophoresis patterns. This is presumably due to the presence of residual random hexamers in the reaction mix. Thus, these methods may require additional purification before DNA quantification. PicoGreen quantification was more accurate as previously reported (Singer *et al.,* 1997, Ahn *et al.,* 1996), because it is not affected by the residual hexamers. This technique gave good quantitative data for two of the three kits tested (illustra GenomiPhi and GenomePlex kits). PicoGreen quantification was in accordance with the electrophoresis pattern, for example W15521 and W15516 had the highest concentration, while W15508 has a concentration of 2.5 ng/ µl (corresponding to the detection threshold of the technique). For REPLI-g amplified samples, only gel-based densitometry was able to provide DNA concentrations without an additional purification step.

For each kit, yield was determined by calculating the amount of DNA in the final reaction volume (10 µl for illustra GenomiPhi kit, 20 µl for REPLI-g kit and 75 µl for GenomePlex kit). Yield range was calculated based on the highest and lowest sample concentrations. The level of amplification was determined as a ratio of concentrations between template and WGA DNA for each sample. The typical yield of illustra GenomiPhi kit was in the 0.7-7 µg range, corresponding to 6-60 fold-amplification (Table 4.2). The highest yield was 10 µg achieved from Iowa DNA, giving over 180 fold-amplification. The typical yield of GenomePlex kit was 4.5- 46 µg range and the level of amplification was in the 10-70-fold. For the REPLI-g kit, the yield was in the 0.6-2 µg range, corresponding to 50-160 fold amplification.

Sample reference	Nanodrop concentration (ng/µl)	Post-WGA DNA quantification								
		Nanodrop (ng/µl)			Hoechst 33258 (ng/µl)			PicoGreen (ng/µl)		
		illustra GenomiPhi	REPLI-g	Genome Plex	illustra GenomiPhi	REPLI-g	Genome Plex	illustra GenomiPhi	REPLI-g	Genome Plex
W15504	12	891.5	1120.9	1072.6	588.9	751.9	740.9	77.3	2.5	302.5
W15507	4.3	887.4	1221.5	999.2	594.4	876.8	672.9	2.5	2.5	178.9
W15508	3.5	967.2	1144.5	917.2	573.6	657.8	646.7	2.5	2.5	2.5
W15519	5.5	890.7	1233.3	968.3	613.6	764.9	995.3	2.5	2.5	118.9
W15521	11.5	805.2	1053.7	889.9	580.5	619.4	822.6	704.1	2.5	369.9
W15509	6.7	865.6	1152.7	912.9	574.2	606.1	570.6	78.5	2.5	7.5
W15511	6.2	819.9	1150.3	941.7	578.7	592.8	596.9	197.2	2.5	60.1
W15516	8.9	815.3	1105.8	938.1	699.6	568.7	766	604.9	2.5	609.8
W15517	6.4	875.8	1109.5	979.3	598.4	573.8	726.8	2.5	2.5	410.3
W15518	13.2	890.0	1179.7	947.0	585.3	663.6	692.6	68.7	2.5	2.5
Iowa	5.8	791.0	1221.7	903.3	631.1	586.4	649	1086.1	2.5	2.5

Table 4.1: Quantification of *Cryptosporidium* DNA before and after WGA using the 3 trialled kits by densitometry, Nanodrop, Hoechst and PicoGreen methods. Nanodrop was used to determine the DNA concentration in clinical samples prior to WGA (2^{nd} column). After WGA, the DNA concentration was determined using Nanodrop and Hoechst, but the concentrations were similar for all the samples. Therefore, PicoGreen was used to determine the DNA concentration post-WGA. Picogreen DNA quantification was more accurate and in accordance with the electrophoresis patterns.

4.3.4 Integrity of the amplified DNA

The integrity of the amplified genomic DNA was assessed by comparing PCR results using three primer sets before and after WGA as described in 2.5.4. For Cry 15/9 primers, amplifying 550 bp fragment of the COWP gene, all 11 samples were positive before WGA. After WGA, 10/11 illustra GenomiPhi amplified samples (91%), 6/11 REPLI-g amplified samples (54.6%) and 9/11 GenomePlex amplified samples (81.9%) were PCR positive (Figure 4.2). For Cgd6_5020 primers, amplifying 270 bp, all 11 samples were also positive before WGA. After WGA, 10/11 illustra GenomiPhi amplified samples (91%), 8/11 REPLI-g amplified samples (72.8%) and 11/11 GenomePlex amplified samples (100%) were PCR positive (Figure 4.3). For Chro.20156 primers, amplifying 247 bp, 10/11 samples were positive before WGA, with only the W15519 sample negative. After WGA, 9/11 illustra GenomiPhi amplified samples (81.9%), 7/11 REPLI-g amplified samples (63.7%) and 11/11 GenomePlex amplified samples (100%) were PCR positive (Figure 4.4).

The overall post-WGA PCR success rates from the three kits were 87.8%, 63.6% and 93.9% for illustra GenomiPhi, REPLI-g and GenomePlex kits, respectively (Table 4.2).

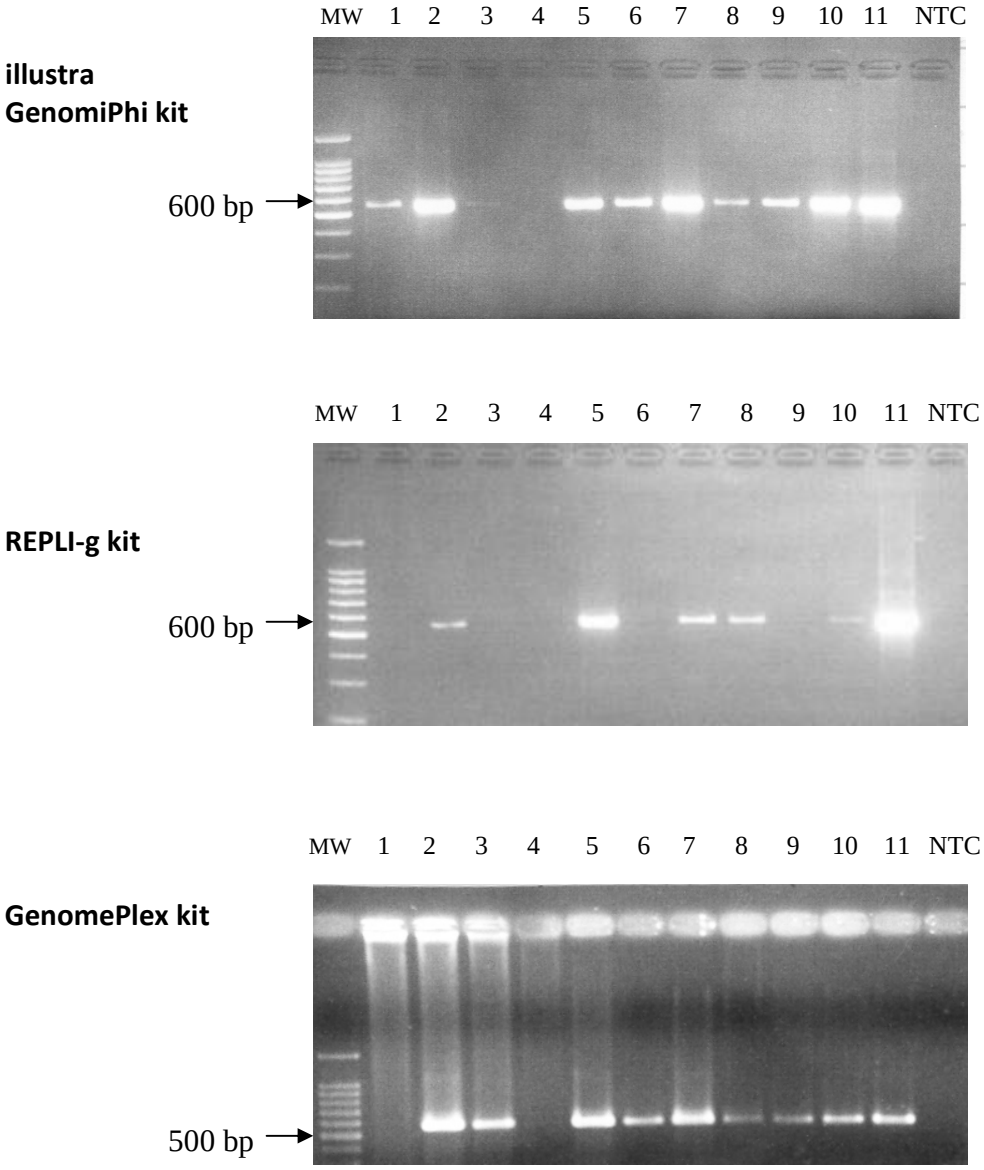


Figure 4.2: PCR products of WGA-amplified *Cryptosporidium* DNA isolates using Cry15/9 primers. MW: molecular weight marker, 1: W15504, 2: W15507, 3: W15508, 4: W15519, 5: W15521, 6: W15509, 7: W15511, 8: W15516, 9: W15517, 10: W15518, 11: *C. parvum* Iowa, NTC: non template control.

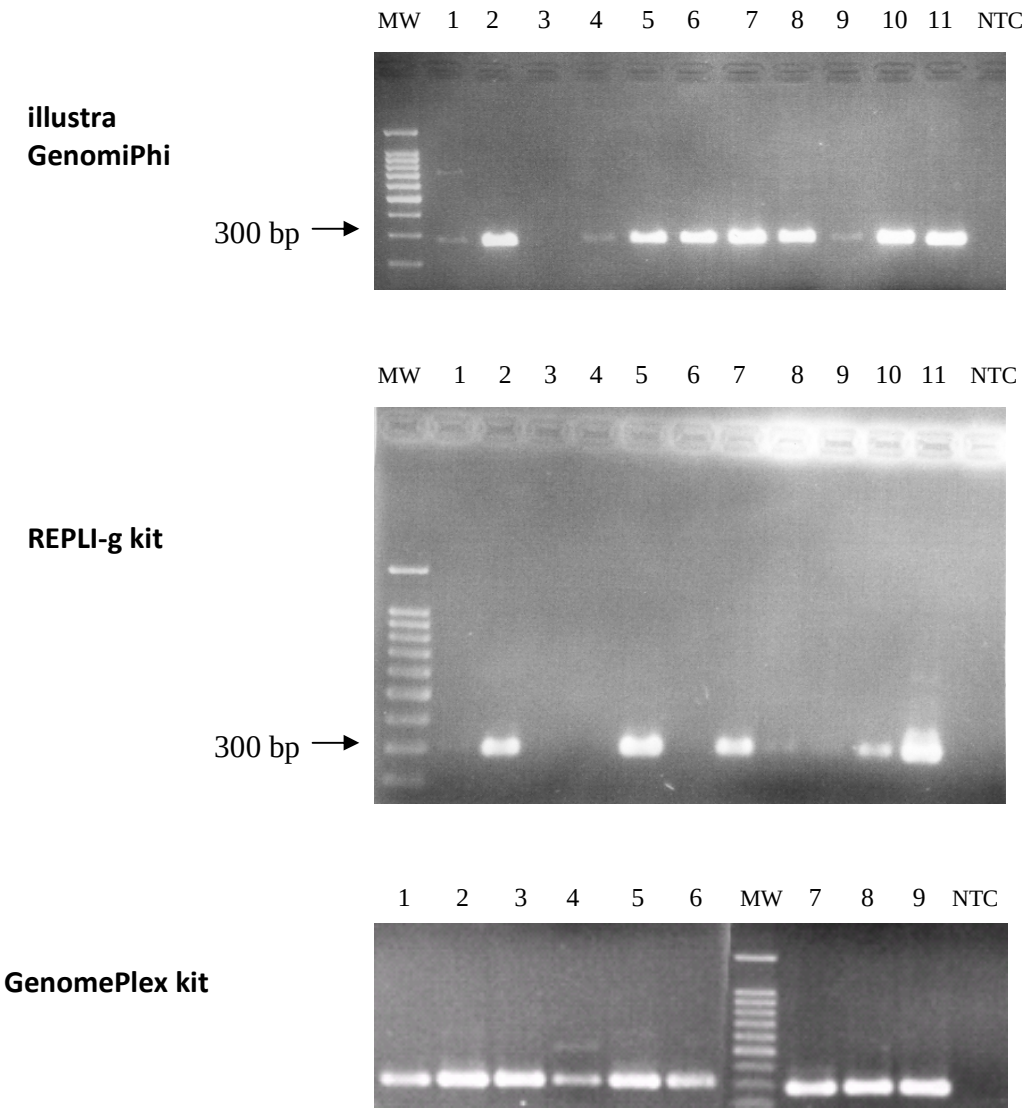


Figure 4.3: PCR products of WGA-amplified *Cryptosporidium* DNA isolates using cgd6_5020 primers. MW: molecular weight marker, 1: W15504, 2: W15507, 3: W15508, 4: W15519, 5: W15521, 6: W15509, 7: W15511, 8: W15516, 9: W15517, 10: W15518, 11: *C. parvum* Iowa, NTC: non template control.

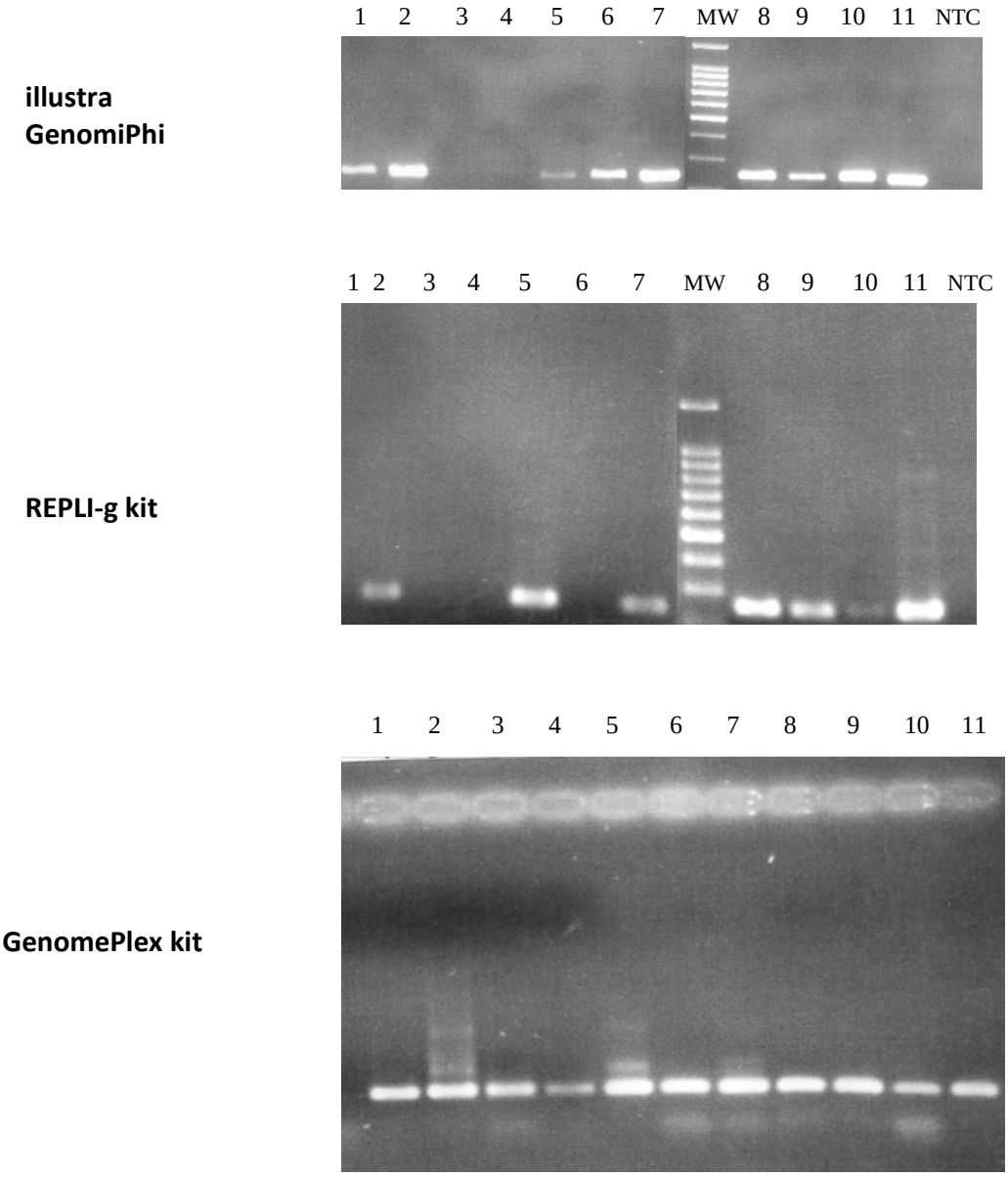


Figure 4.4: PCR products of WGA-amplified *Cryptosporidium* DNA isolates using chro.20156 primers. MW: molecular weight marker, 1: W15504, 2: W15507, 3: W15508, 4: W15519, 5: W15521, 6: W15509, 7: W15511, 8: W15516, 9: W15517, 10: W15518, 11: *C. parvum* Iowa, NTC: non template control.

	Success rate (electro-phoresis)	Size of amplified product	PicoGreen quantification	Yield	Level of amplification	Secondary amplification (post WGA PCR)	Fidelity (nucleotide sequence match)	Specific amplification of *Cryptosporidium* DNA (Real-Time PCR)
Illustra GenomiPhi	90.9%	> 10 kb	Yes	0.7-10 µg	6-180 fold	87.8%	100%	Yes (11/11 samples)
REPLI-g	45.4%	> 10 kb	No	* 0.6-2 µg	* 50-160 fold	63.6%	99.62%	Not tested
GenomePlex	72.7%	< 1 kb	Yes	4.5-46 µg	10-70 fold	93.9%	99.87%	Not tested

Table 4.2: Comparison of the performances of the 3 trialled WGA commercial kits. The performance was evaluated by comparing the success rate, the size of the amplified genomic DNA and the yield for each kit. In addition, the integrity, fidelity and the specific amplification of *Cryptosporidium* DNA were evaluated.

Success rate was the number of samples with positive amplification as determined by agarose gel electrophoresis. The size of amplified product was determined by comparison to a molecular weight marker. Picogreen quantification column showed the successful quantification without prior sample purification. Yield and level of amplification were determined based on PicoGreen results for illustra GenomiPhi and GenomePlex kits. * for REPLI-g kit, densitometry was used to determine the yield and the level of amplification. Secondary amplification was the success rate of post-WGA PCR reactions. Fidelity was determined by the number of identical nucleotides before and after WGA.

4.3.5 Fidelity of WGA kits

The fidelity of amplification was assessed by comparing PCR product sequences, generated with and without a WGA intermediate step. For Cry 15/9 primers, PCR product sequences were identical using all three kits for the W15507, W15511 and Iowa isolates (Appendix VI). For the W15516 isolate, WGA using REPLI-g kit produced three nucleotide errors and the use of GenomePlex kit produced one error, corresponding to 99.34% and 99.78% sequence concordance, respectively. For Cgd6_5020 and Chro.20156 primers, PCR product sequences were identical before and after WGA for all the samples. The overall fidelity of the three kits was 100%, 99.62% and 99.87% for illustra GenomiPhi, REPLI-g and GenomePlex kits, respectively (Table 4.2).

Interestingly, one of the samples tested (W15519) failed to amplify using Chro.20156 primers without WGA or after WGA using illustra GenomiPhi and REPLI-g kits, but did amplify when GenomePlex amplified DNA was used. The identity of the PCR product was confirmed by sequencing.

4.3.6 Comparative analysis of *Cryptosporidium* genomic DNA before and after WGA

The evaluation of the specific amplification of *Cryptosporidium* DNA after WGA was performed by comparing the amplified products and the original genomic DNA using a species-specific, semi quantitative Real-time PCR assay as described in 2.5.5. Melt curve analysis of genomic DNA and WGA DNA amplified using the Illustra GenomiPhi kit confirmed the specific amplification of *Cryptosporidium* DNA after WGA in each of the samples tested. The unadjusted C_T values showed that 8/11 of the samples had a lower C_T after WGA than before, indicating that in these samples the WGA did amplify *Cryptosporidium* DNA. After adjustment to allow for the 10-fold dilution

applied through the WGA process, all of the samples had a lower C_T value after WGA, confirming that *Cryptosporidium* DNA was present in higher copy numbers in the samples post WGA than before (Table 4.3).

The highest difference of C_T before and after WGA was 10.13 for sample W15516, which corresponds to over 30-fold increase in DNA target copy number. This is in accordance with the electrophoresis result. For the other samples, the difference in C_T values before and after WGA ranged between 2.63 (for sample W15504) and 8.11 (for sample W15511), corresponding to an 8 to 27 fold increase in copy number. For sample W15508, the difference in C_T value was 0.73 supporting a poor amplification.

Sample	Genomic DNA / WGA DNA	Mean C_T (adjusted)	C_T value difference	Fold increase in copy numbers
W15504	Genomic DNA	31.63 ±0.14	2.63 ±0.45	8.8 ±0.45
	WGA DNA	29.00 ±0.43		
W15507	Genomic DNA	22.62 ±0.19	4.53 ±0.21	15 ±0.21
	WGA DNA	18.09 ±0.11		
W15508	Genomic DNA	29.06 ±0.16	0.73 ±0.80	2.5 ±0.80
	WGA DNA	28.33 ±0.79		
W15519	Genomic DNA	32.56 ±0.71	2.54	8.5
	WGA DNA	30.02		
W15521	Genomic DNA	23.72 ±0.11	4.41 ±0.82	14.8 ±0.82
	WGA DNA	19.31 ±0.82		
W15509	Genomic DNA	28.82 ±0.25	4.09 ±0.41	13.7 ±0.41
	WGA DNA	24.73 ±0.33		
W15511	Genomic DNA	23.57 ±0.22	8.11 ±0.22	27 ±0.22
	WGA DNA	15.46 ±0.06		
W15516	Genomic DNA	29.57 ±0.41	10.13	33.8
	WGA DNA	19.44		
W15517	Genomic DNA	28.87 ±0.04	4.08 ±0.54	13.6 ±0.54
	WGA DNA	24.79 ±0.53		
W15518	Genomic DNA	25.99 ±1.11	7.81 ±1.13	26 ±1.13
	WGA DNA	18.18 ±0.25		
Iowa	Genomic DNA	19.15 ±0.03	7.96 ±0.06	26.6 ±0.06
	WGA DNA	11.46 ±0.61		

Table 4.3: Real-time PCR analysis of *Cryptosporidium* DNA before and after WGA and estimation of the increase in copy numbers after WGA using illustra GenomiPhi kit.

4.4 Discussion

WGA commercial kits were successfully used for the amplification of *Cryptosporidium* genomic DNA from clinical isolates. Illustra GenomiPhi and GenomePlex kits successfully amplified the majority of the tested isolates (90.9% and 72.7%, respectively). The REPLI-g Kit, however, amplified less than half of the samples. One sample, W15508, did not amplify with any of the three kits tested. While the template concentration was well within the recommended range for the kits under test, it did correlate with the sample having the lowest concentration of template DNA tested and thus may reflect some degree of degradation of the DNA in that sample or relative enrichment of inhibitors carried through the DNA purification process.

For MDA-based kits (illustra GenomiPhi and REPLI-g), it was noticeable that most of the DNA remains in the well when run out on an agarose gel, this is due to the formation of very high molecular weight DNA, independent of the genome size, as previously reported by Detter *et al.* (2002). OmniPlex-amplified DNA was of smaller size, because this method is based on fragmentation of the genomic DNA followed by linker ligation to enable amplification (Thorstenson *et al.* 1998; Fiegler *et al.* 2007).

PicoGreen DNA quantification of the amplified DNA proved effective for Illustra GenomiPhi and GenomePlex kits. The yield range was 0.7-10 µg and 4.5-46 µg, respectively, which is in accordance with the manufacturer's claims. The apparent higher yield of GenomePlex kit is due to a higher reaction volume (10 µl versus 75 µl). The PicoGreen quantification of REPLI-g kit amplified samples was not effective without prior purification of the amplified DNA. This could be in part explained by the fact that high proportion of the amplified DNA may be present as single-stranded product rather than dsDNA.

The integrity of the amplified DNA was assessed by the ability to generate PCR products for three genes. The overall post-WGA PCR success rate was 87.8%, 63.6% and 93.9%, for illustra GenomiPhi, REPLI-g and GenomePlex kits, respectively. For one sample, the W15519 isolate, no PCR amplification using Chro.20156 primers was observed before WGA or after WGA using illustra GenomiPhi and REPLI-g kits. However, GenomePlex amplified DNA of the same sample was PCR positive and the identity of the PCR product was confirmed by sequencing. These results suggest that the use of WGA amplified DNA as PCR template could increase PCR sensitivity from clinical samples. Similar findings were reported for the detection of *Trypanosoma* species from blood samples (Pinchbeck *et al.,* 2008). Further work to more thoroughly test the integrity of the amplified DNA and assessing amplification biases should focus on amplification of longer sequences and broader genomic coverage utilizing sequences from each of the eight nuclear chromosomes, from the telomeres, centromeres, ribosomal DNA, mitochondrial DNA, and repetitive regions.

The fidelity of the amplification was assessed by sequence analysis of the PCR products, before and after WGA. For this sample set, the overall error rate observed was 0% for illustra GenomiPhi kit, 0.38% for REPLI-g kit and 0.13% for GenomePlex kit. Interestingly, all the errors arose from the same clinical sample using the same set of primers. This could be explained by a variety of factors, such as the presence of impurities affecting the enzyme proofreading activity, the secondary structure of the DNA, or by a low concentration of the starting material, which can decrease the amplification fidelity as previously reported (Bergen *et al.*, 2005).

Sequence analysis of WGA amplified PCR products using Cry 15/9 primers showed the preservation of six species-specific Single Nucleotide Polymorphisms (SNP)s, one of which at position 66 is of particular interest as it corresponds to an *RsaI* restriction site used for *Cryptosporidium* genotyping as previously described (Spano

et al., 1997). In addition, sequence analysis showed the preservation of one species-specific SNP for Cgd6_5020 gene and five species-specific SNPs for Chro.20156.

Comparative analysis of *Cryptosporidium* genomic DNA and paired WGA DNA using a real-time PCR assay confirmed that *Cryptosporidium* DNA was specifically amplified using illustra GenomiPhi kit, resulting in higher copy numbers post WGA than before for all the samples tested.

For this collection of *Cryptosporidium* clinical isolates, the illustra GenomiPhi WGA kit had the best performance, with 90.9% success rate, generating high concentration of high molecular weight DNA with 100% fidelity. The additional cost of WGA is not prohibitive for clinical usage when added to the PCR cost for routine detection of *Cryptosporidium* DNA. These are preliminary results, highlighting the usefulness of MDA based WGA for the accurate amplification of *Cryptosporidium* genomic DNA for the purposes of creating a bank of clinical isolates, thus enabling extensive genetic testing.

These results were obtained from semi-purified oocyst suspensions, themselves requiring reasonable numbers of oocysts in the original sample. They carry significant contamination of bacterial and fungal DNA, which can also be amplified by the WGA process. Real-time PCR for faecal marker DNAs such as genes from *Bacteroides*, *Clostridium* and *E. coli* might be useful to determine whether there is any predisposition to amplifying contaminating DNAs rather than the *Cryptosporidium* DNA in the samples. Additional investigation of the effect of other oocyst purification methods (immunomagnetic separation) on the performance of WGA should be carried out. Importantly, since many investigators use DNA extracted from raw stool without prior oocyst purification, independent validation of the suitability of this material for WGA and downstream analysis should also be undertaken.

This study investigated the suitability of *Cryptosporidium* DNA after WGA for genotyping purposes. The COWP marker was tested as it is routinely used for *Cryptosporidium* speciation by RFLP, together with 2 novel markers. These results showed efficient and specific amplification of *Cryptosporidium* DNA. Further validation of these WGA techniques for the preservation of routine subtyping targets of *Cryptosporidium* (GP60 and mini and micro-satellite repeats) would be necessary and desirable before adopting WGA for routine characterization of clinical and environmental isolates of *Cryptosporidium* species.

4.5 Summary

During this study, three WGA commercial kits were trialled to evaluate their efficiency to resolve the issue of the limited DNA amounts available from clinical samples especially with the perceived need for extensive genetic testing. This is particularly relevant in the case of fastidious in culture pathogens like *Cryptosporidium*. The performance of these kits was assessed on the base of the success rate, the coverage and the fidelity of amplification of genomic DNA from five *C. hominis* and five *C. parvum* clinical isolates and *C. parvum* Iowa reference strain.

The higher success rates were obtained using Illustra GenomiPhi and GenomePlex kits: 90.9% and 72.7%, respectively, REPLI-g Kit, however, amplified less than half of the samples. The yield was comparable between the three kits and in accordance with the manufacturer's claims. Using these DNA samples, REPLI-g Kit had the lowest performance with a low success rate, poor coverage, reduced fidelity and inadequacy for direct PicoGreen DNA quantification. The main limitation of the GenomePlex kit was the production of small-sized DNA due to the fragmentation and linker ligation steps required for Omniplex based technique for WGA, which might limit downstream applications.

One kit, Illustra GenomiPhi, amplified the majority of *Cryptosporidium* samples tested generating high amounts of genomic DNA with high fidelity. These results highlight the usefulness of MDA based WGA for the accurate amplification of *Cryptosporidium* genomic DNA and for the creation of genomic banks of clinical isolates, which would enable extensive genetic testing.

CHAPTER 5:

Molecular and genetic characterization of Cops-1

5.1 Introduction

During the genomic era, advances in sequencing technologies have generated ever larger and more complex genomic data sets that have poured into public databases (Collins *et al.,* 2003), which have started to improve the understanding of the biology of health and disease at an unprecedented level of molecular detail. Thus, genomics has come to be considered a central and cohesive discipline in biomedical research (Guttmacher *et al.,* 2003, Collins *et al.,* 2003). The first completed bacterial genome project was *Haemophilus influenzae* (Fleischmann *et al.,* 1995); subsequently several genome sequencing projects were conducted including many important human pathogens. Initially, the simplest known free-living model organisms *Escherichia coli* and *Bacillus subtilis* and the first eukaryote *Saccharomyces cerevisiae* were sequenced as reviewed by Fraser and colleagues (2000). Since then, the genomes of many more microbes, mammals, vertebrates have been published and culminating in publication of the human genome project as reviewed by Collins and colleagues (2003).

In the post-genomic era, great progress in sequencing technologies and platforms has been achieved and the amount of novel microbial genomic information being generated is so vast that only a multidisciplinary approach integrating bioinformatics, statistics and mathematics is able to assess it effectively (Medini *et al.,* 2008). Apicomplexan parasites are a successful group of eukaryotic parasites. The diseases they cause in humans and animals pose major threats to world health and global economy (Tomley, 2009). Protozoan parasites have not been left behind in the rush to genomics and proteomics (Wastling *et al.,* 2009). Genomes of 32 apicomplexans are now finished and are deposited in comprehensive and publicly accessible genomics and proteomics databases (Zarlenga and Gasbarre, 2009). These databases give investigators the ability to analyze biodiversity, evolution and population genetics, functional genomics, host-parasite relationships and epigenetics, transmission dynamics,

pathogenicity determinants, metabolic pathways, gene expression and could drive the discovery of novel diagnostic, drug and vaccine targets (Rider and Zhu, 2010, Zarlenga and Gasbarre, 2009, Tomley, 2009, Medini *et al.,* 2008). While the biology of each parasite differs, all the apicomplexan parasites share many functions and pathways related to their intracellular niche (Tomley, 2009). Approximately, 150 ancestral apical complex apicomplexan proteins are conserved between *Cryptosporidium* spp and *Plasmodium* spp (Templeton *et al.,* 2004). In particular, actin-like proteins implicated in cell invasion are characteristically conserved among *Toxoplasma gondii, Plasmodium* spp., *Cryptosporidium* spp. and *Theileria* spp (Gordon and Sibley, 2005).

Genome sequences of *C. parvum* and *C. hominis* were published in the same year (Abrahamsen *et al.,* 2004, Xu *et al.,* 2004). Genome sizes were comparable and quite small compared to *P. falciparum* and *T. gondii* genomes. The reduced size is partly due to a lack of genes for plastid maintenance (apicoplast) and variant surface antigen gene families, the reduced number of introns and smaller intergenic regions (Rider and Zhu, 2010, Xu *et al.,* 2004). The complete genome sequences of these two *Cryptosporidium* species has stimulated recent progress in identifying conserved vital *Cryptosporidium* molecules through genome comparisons and data mining, including DNA replication machinery, DNA repair, transcription factors, gene silencing and regulation, histone modification and methylation as reviewed by Rider and Zhu (2010). Tomley (2009) stated that the priority for apicomplexan research in the post-genomic era should be the functional characterization of unique parasite molecules, which have no counterparts in other organisms and which are implicated in parasite-specific metabolic pathways or host-parasite interactions at the molecular and structural levels.

C. hominis and *C. parvum* comparative genomics showed that the two genome sequences are very similar and exhibit only 3–5% sequence divergence, with no large insertions, deletions or rearrangements (Xu *et al.,* 2004). The authors concluded that the phenotypic differences between *C. hominis* and *C. parvum*

are caused by polymorphisms in coding regions and differences in gene regulation (Pain *et al.,* 2005, Xu *et al.,* 2004). The role of what genetic variability is present in the phenotypic differences between *C. hominis* and *C. parvum* is now much more accessible for investigation. In fact, the subset of genes displaying the most variation may include valuable epidemiological markers and unnoticed genetic determinants of host specificity and virulence. This hypothesis was investigated in Chapter 3 and although the majority of the predicted genes were common to *C. parvum* and *C. hominis* and some to *C. meleagridis*, evidence was found of one *C. parvum* putative specific gene (Cops-1) and one *C. hominis* putative specific gene (Chos-1).

5.2 Aims

The aim of this chapter is to characterize Cops-1 gene and its gene products. The investigation includes determination of the full length gene sequence to assess genetic polymorphism in *C. parvum* isolates, characterization of the protein using a monoclonal peptide antibody and evaluation of the antigenicity of Cops-1 protein by screening of patient sera using the recombinant protein. Finally, *in vitro* assessment of the role of Cops-1 protein in host-cell attachment and invasion using a blocking assay with the monoclonal antibody in coculture with Caco-2 cell line.

5.3 Results

5.3.1 Cops-1 predicted features

The initial annotation of the Cops-1 gene provided by the *C. parvum* genome project shows no apparent orthologs in other genomes, including *C. hominis*. The limited annotation provided describes a serine-rich protein containing repeated motifs, with an N-terminal secretory peptide, situated proximal to the telomeric repeats of chromosome 2 (1.08Mb). The Cops-1 genomic position as retrieved

from CryptoDB is 983,586- 985,019. The telomeric location might have hindered the detection of Cops-1 in other *Cryptosporidium* species. In fact, the telomeric regions are highly repetitive and renowned to be tricky to assemble, especially in genome sequencing projects. In *C. hominis*, there is no contig covering the Cops-1 genomic region. Similarly, Cops-1 was not detected in the ongoing genome sequencing of *C. muris* so far. Based on *C. parvum* genome sequence, Cops-1 is 1434 bp, the encoded protein is predicted to be 50.164 kDa with an isoelectric point of 11.

Using InterPro software, Cops-1 protein is predicted to have a signal peptide (position 1-24) and a transmembrane domain (position 5-27). SignalP 3.0 software predicts that Cops-1 is likely to be a secreted protein and predicts that the protein would be cleaved between position 24 and 25 due to the recognition of the motif "LQT-FF", generating a mature polypeptide of 47.477 kDa. In addition, PSORT software analysis showed that the protein has an N-myristoylation motif "MGNSLNV" and is more likely to have nuclear location. However, this motif would be cleaved off in the mature protein. The gene has putative N-glycosylated site and multiple potential sites for phosphorylation, suggesting that under appropriate conditions, Cops-1 protein can be heavily phosphorylated. In fact, the protein sequence shows 47 potential serine phosphorylation sites, 18 threonine sites and 5 tyrosine phosphorylation sites. The Cops-1 protein is predicted to have a hydrophobic N terminus, but one which would be cleaved off in the mature protein.

Using BLAST, it was noted that from positions 74 to 264 in the Cops-1 of *C. parvum,* there was low level similarity (32% identity over 477 amino acids) to a region of the product of the *Drosophila melanogaster* gene "*shot*" (short stop), which is a calcium-binding protein involved in cross-linking microtubules to microfilaments. This fly protein is very large (5155 amino acids) and no other part of the Short polypeptide has detectable similarity to Cops-1, so the biological significance of this homology is questionable.

5.3.2 Cops-1 as diagnostic target

The Cops-1 gene was the only putative *C. parvum* specific gene confirmed experimentally, amplifying 665 bp from *C. parvum* but not *C. hominis* isolates as shown in Chapter 3. The full length Cops-1 gene was amplified from *C. parvum* DNA only (Figure 5.1). Agarose gel electrophoresis of PCR products using Cgd2_4380 F and R primers showed that, at increased UV exposure, these primers amplified three bands from *C. parvum* Iowa DNA of 655, 450 bp and 200 bp (Figure 5.1). The 200 bp band was also present in PCR products from *C. hominis*. This band could serve as a useful internal control for the presence of human infective *Cryptosporidium*. When DNA from other *Cryptosporidium* species was tested, only *C. meleagridis* was PCR positive, generating a 665 bp product, no 200 bp band was visible. This assay as it stands confirms and discriminates *C. hominis* from *C. parvum*.

5.3.3 Cops-1 has an ortholog in *C. hominis*

The identity of the 200 bp band amplified from *C. hominis* DNA was confirmed by sequencing, since its intensity was strong enough to allow purification and sequencing as described in 2.4. Interestingly, the retrieved sequence showed 99% sequence identity to Cgd2_4380 (Figure 5.2). The 200 bp is a fraction of the 650 bp. Therefore, Cops-1 seems to have an ortholog in *C. hominis*. CpCops-1 and ChCops-1 are used to name the gene in *C. parvum* and *C. hominis*, respectively.

5.3.4 Cops-1 PCR products sequence analysis

Cops-1 PCR products were used for sequence analysis. Sequences from Cp2, Cp3, Cp4 and Moredun isolates were identical to the published Cgd2_4380 gene sequence. The anthroponotic subgroup isolates showed four SNPs specific to this subgroup. The alignment of Cops-1 PCR product sequences are shown in Appendix VII. *C. hominis* PCR products, despite being shorter, showed high sequence similarity to PCR products from *C. parvum* isolates.

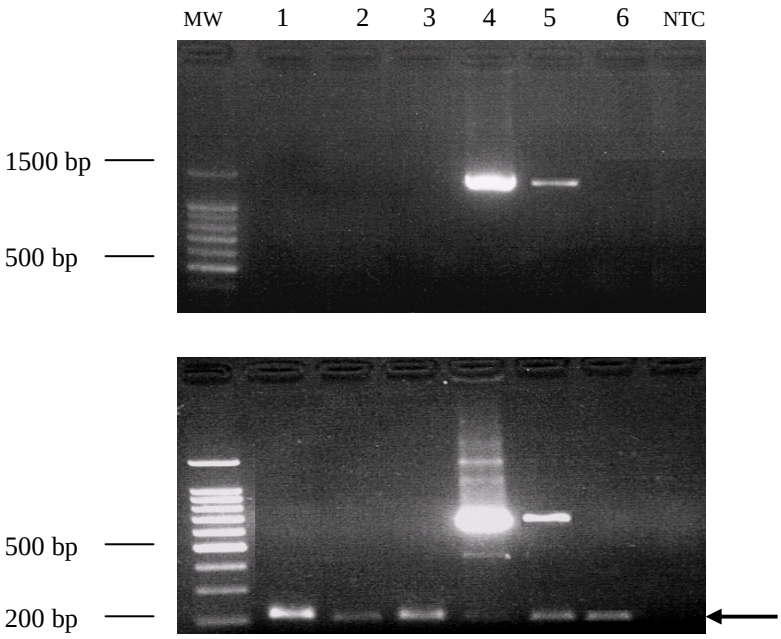
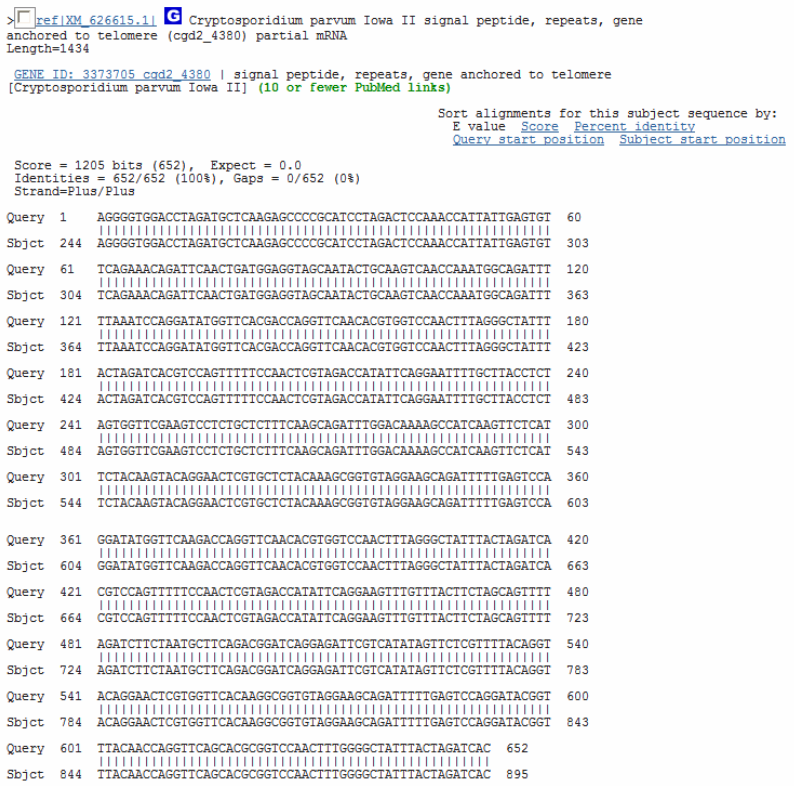


Figure 5.1: Amplification of full and partial Cops-1 from *C. parvum* and *C. hominis* isolates. **A**: amplification of the full gene length using Cgd2_4380 FF and FR primers. PCR was positive from *C. parvum* DNA only. **B**: Amplification of a fragment of Cops-1 using Cgd2_4380 F and R primers. From *C. parvum* DNA, the primers amplified 3 bands of 650 bp, 450 bp and 200 bp. The 200 bp band was also visible in *C. hominis* PCR products (arrow). MW: molecular weight, 1: Ch2, 2: Ch3, 3: Ch4, 4: Iowa, 5: Moredun, 6: TU502, NTC: non template control.

A



>ref|XM_626615.1| Cryptosporidium parvum Iowa II signal peptide, repeats, gene anchored to telomere (cgd2_4380) partial mRNA
Length=1434

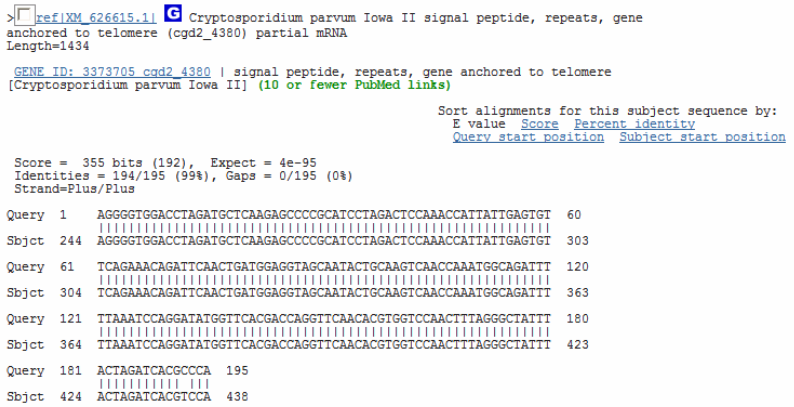
GENE ID: 3373705 cgd2_4380 | signal peptide, repeats, gene anchored to telomere [Cryptosporidium parvum Iowa II] (10 or fewer PubMed links)

Sort alignments for this subject sequence by:
E value Score Percent identity
Query start position Subject start position

Score = 1205 bits (652), Expect = 0.0
Identities = 652/652 (100%), Gaps = 0/652 (0%)
Strand=Plus/Plus

```
Query  1    AGGGGTGGACCTAGATGCTCAAGAGCCCCGCATCCTAGACTCCAAACCATTATTGAGTGT  60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  244  AGGGGTGGACCTAGATGCTCAAGAGCCCCGCATCCTAGACTCCAAACCATTATTGAGTGT  303

Query  61   TCAGAAACAGATTCAACTGATGGAGGTAGCAATACTGCAAGTCAACCAAATGGCAGATTT  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  304  TCAGAAACAGATTCAACTGATGGAGGTAGCAATACTGCAAGTCAACCAAATGGCAGATTT  363

Query  121  TTAAATCCAGGATATGGTTCACGACCAGGTTCAACACGTGGTCCAACTTTAGGGCTATTT  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  364  TTAAATCCAGGATATGGTTCACGACCAGGTTCAACACGTGGTCCAACTTTAGGGCTATTT  423

Query  181  ACTAGATCACGTCCAGTTTTTCCAACTCGTAGACCATATTCAGGAATTTTGCTTACCTCT  240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  424  ACTAGATCACGTCCAGTTTTTCCAACTCGTAGACCATATTCAGGAATTTTGCTTACCTCT  483

Query  241  AGTGGTTCGAAGTCCTCTGCTCTTTCAAGCAGATTTGGACAAAAGCCATCAAGTTCTCAT  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  484  AGTGGTTCGAAGTCCTCTGCTCTTTCAAGCAGATTTGGACAAAAGCCATCAAGTTCTCAT  543

Query  301  TCTACAAGTACAGGAACTCGTGCTCTACAAAGCGGTGTAGGAAGCAGATTTTTGAGTCCA  360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  544  TCTACAAGTACAGGAACTCGTGCTCTACAAAGCGGTGTAGGAAGCAGATTTTTGAGTCCA  603

Query  361  GGATATGGTTCAAGACCAGGTTCAACACGTGGTCCAACTTTAGGGCTATTTACTAGATCA  420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  604  GGATATGGTTCAAGACCAGGTTCAACACGTGGTCCAACTTTAGGGCTATTTACTAGATCA  663

Query  421  CGTCCAGTTTTTCCAACTCGTAGACCATATTCAGGAAGTTTGTTTACTTCTAGCAGTTTT  480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  664  CGTCCAGTTTTTCCAACTCGTAGACCATATTCAGGAAGTTTGTTTACTTCTAGCAGTTTT  723

Query  481  AGATCTTCTAATGCTTCAGACGGATCAGGAGATTCGTCATATAGTTCTCGTTTTACAGGT  540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  724  AGATCTTCTAATGCTTCAGACGGATCAGGAGATTCGTCATATAGTTCTCGTTTTACAGGT  783

Query  541  ACAGGAACTCGTGGTTCACAAGGCGGTGTAGGAAGCAGATTTTTGAGTCCAGGATACGGT  600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  784  ACAGGAACTCGTGGTTCACAAGGCGGTGTAGGAAGCAGATTTTTGAGTCCAGGATACGGT  843

Query  601  TTACAACCAGGTTCAGCACGCGGTCCAACTTTGGGGCTATTTACTAGATCAC  652
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  844  TTACAACCAGGTTCAGCACGCGGTCCAACTTTGGGGCTATTTACTAGATCAC  895
```

B



>ref|XM_626615.1| Cryptosporidium parvum Iowa II signal peptide, repeats, gene anchored to telomere (cgd2_4380) partial mRNA
Length=1434

GENE ID: 3373705 cgd2_4380 | signal peptide, repeats, gene anchored to telomere [Cryptosporidium parvum Iowa II] (10 or fewer PubMed links)

Sort alignments for this subject sequence by:
E value Score Percent identity
Query start position Subject start position

Score = 355 bits (192), Expect = 4e-95
Identities = 194/195 (99%), Gaps = 0/195 (0%)
Strand=Plus/Plus

```
Query  1    AGGGGTGGACCTAGATGCTCAAGAGCCCCGCATCCTAGACTCCAAACCATTATTGAGTGT  60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  244  AGGGGTGGACCTAGATGCTCAAGAGCCCCGCATCCTAGACTCCAAACCATTATTGAGTGT  303

Query  61   TCAGAAACAGATTCAACTGATGGAGGTAGCAATACTGCAAGTCAACCAAATGGCAGATTT  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  304  TCAGAAACAGATTCAACTGATGGAGGTAGCAATACTGCAAGTCAACCAAATGGCAGATTT  363

Query  121  TTAAATCCAGGATATGGTTCACGACCAGGTTCAACACGTGGTCCAACTTTAGGGCTATTT  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  364  TTAAATCCAGGATATGGTTCACGACCAGGTTCAACACGTGGTCCAACTTTAGGGCTATTT  423

Query  181  ACTAGATCACGCCCA  195
            ||||||||||| |||
Sbjct  424  ACTAGATCACGTCCA  438
```

Figure 5.2: BLAST result of the PCR product using Cgd2_4380 F and R primers amplifying 655 bp from *C. parvum* Iowa DNA (A) and 200 bp from *C. hominis* TU502 DNA (B). **A**: The sequence and location of *C. parvum* PCR product was as expected (662 bp) corresponding to hybridization sites for the Forward primer (243-262) and Reverse primer (888-907) and showed 100% sequence identity to the published Cgd2_4380 gene sequence. **B**: The 200 bp fragment amplified from *C. hominis* DNA showed 99% sequence identity to Cgd2_4380 gene and corresponded to a fragment of the *C. parvum* PCR product.

Nevertheless, the alignment revealed four *C. hominis*-specific SNPs. Rabbit genotype PCR products were faint and did not generate good quality sequences and therefore were excluded from the analysis. Each of the anthroponotic *C. parvum* SNPs were non-synonymous, resulting in a change of the protein sequence. Of the *C. hominis* SNPs, two were synonymous and two non-synonymous. Overall, 6/8 (75%) of the SNPs detected were non synonymous. Sequence analysis of *C. meleagridis* PCR product showed only one SNP difference between this fragment and the *C. parvum* (non-anthroponotic) sequences and that this change was non-synonymous.

The 650 bp PCR product sequences from *C. parvum* were used to build a Neighbor-Joining Tree (NJ), which showed a good discrimination of *C. parvum* and *C. parvum* anthroponotic subtype (Figure 5.3 A). The sequences of the 200 bp PCR product, detected in all isolates except rabbit genotype, were also used to construct a NJ Tree (Figure 5.3 B). This tree showed that the variability associated with this short fragment has a good discriminatory power and allowed discrimination of *Cryptosporidium* genotypes and subtypes, which was consistent with the polymorphism seen in the other genetic loci used for MLA as shown in Chapter 3.

5.3.5 Determination of the full gene length of ChCops-1

These initial findings suggested that Cops-1 gene, first thought to be *C. parvum* specific, has an ortholog in *C. hominis*. The primers (Cgd2_4380 FF and FR) that had initially been used to amplify the full gene length in *C. parvum* did not generate PCR product from *C. hominis* DNA. In addition, a set of primers targeting the 5' upstream region and 3' downstream region (Cgd2_4380_flanking_5'end and Cgd2_4380_flanking_3R) were also tested on *C. hominis* genomic DNA. However, these, too, did not generate any PCR product. Therefore, a primer walking approach was used in an attempt to determine the full sequence of the ChCops-1 gene. Several primers were designed as shown in Figure 5.4 and tested.

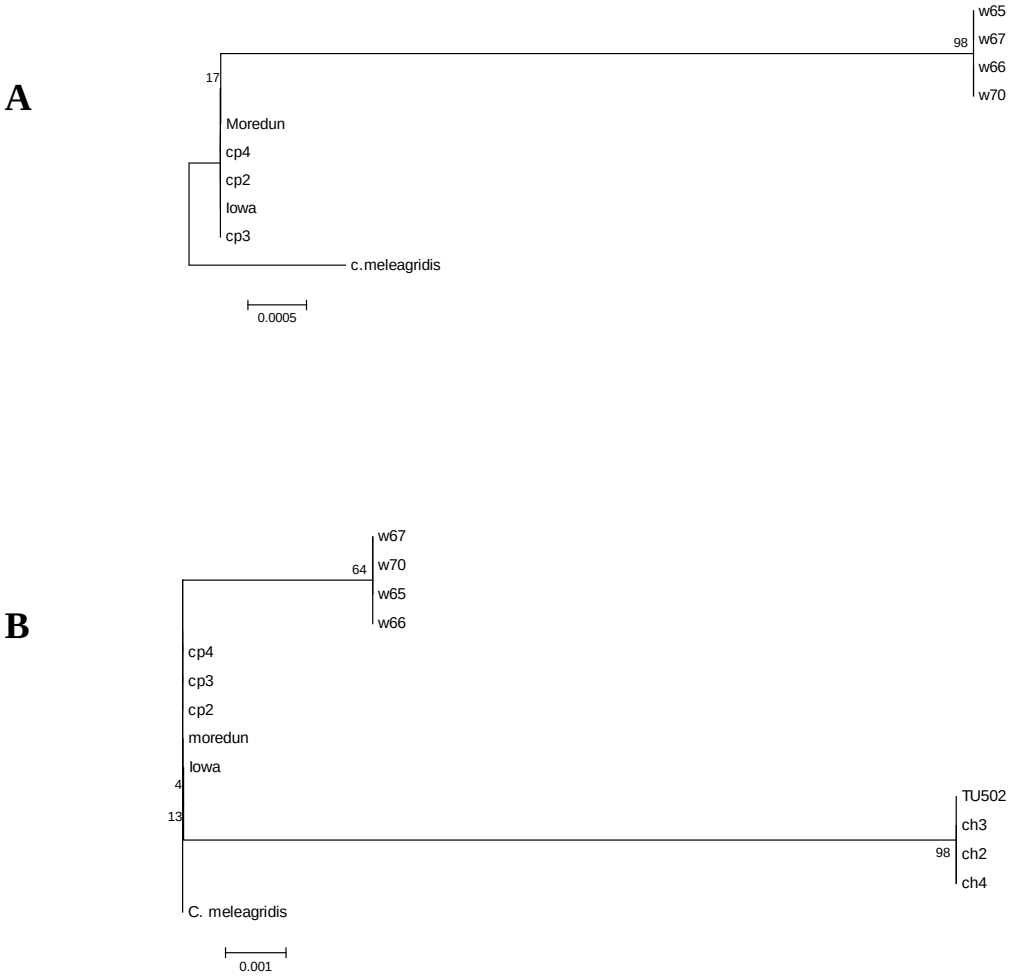
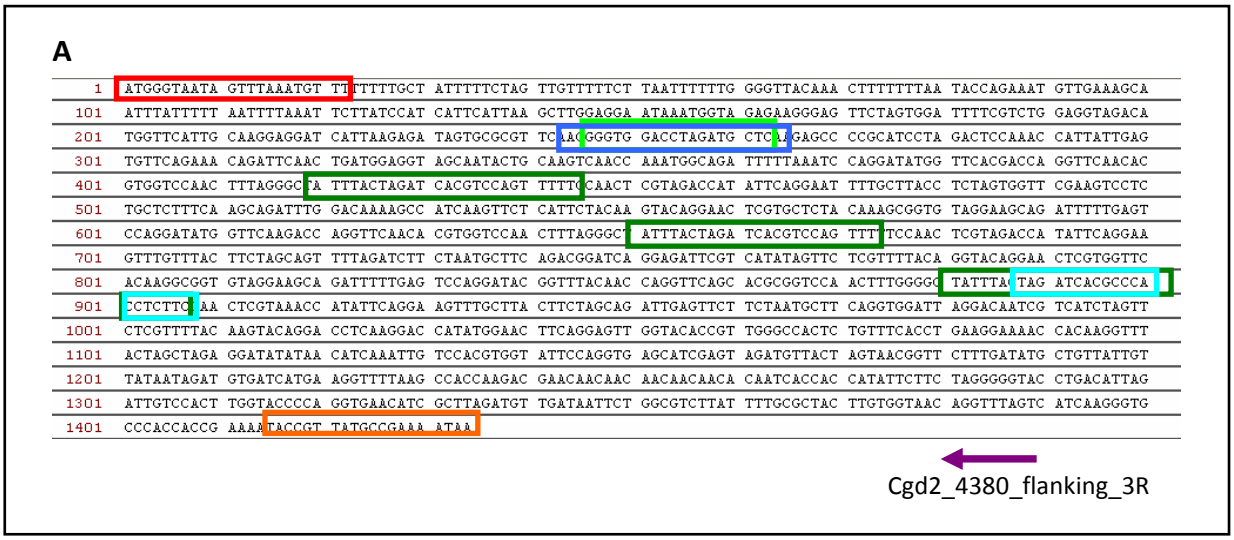
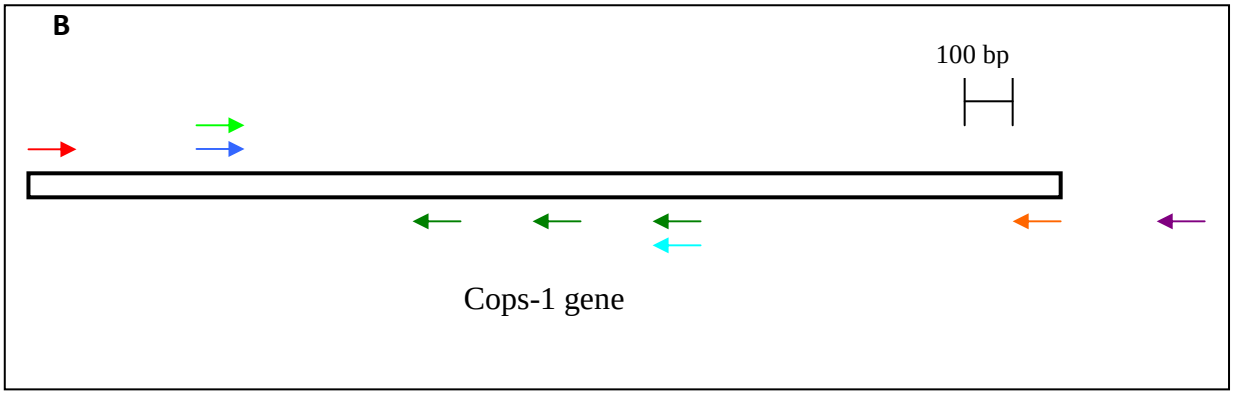


Figure 5.3: Neighbour-joining trees of Cops-1 sequences from *C. hominis, C. parvum* and *C. meleagridis*, based on: **A**: sequences of 650 bp PCR product retrieved from *C. parvum* and *C. meleagridis*. **B**: sequences of 200 bp PCR product retrieved from *C. parvum, C. meleagridis* and *C. hominis*.





C

	PCR products	
Primer combination	*C. parvum*	*C. hominis*
Cgd2_4380 F/R	650 bp 450 bp 200 bp	200 bp
Cgd2_4380 FF/FR	1434 bp	No
Cgd2_4380 2F/2R	650 bp 450 bp 200 bp	450 bp 200 bp
Cgd2_4380 FF/2R	900 bp 650 bp 450 bp	650 bp 450 bp
Cgd2_4380 FF/flanking_3R	1500 bp	1300 bp

Figure 5.4: Details of the different primers used for primer walking approach aiming to uncover the full gene sequence of ChCops-1. **A**: primer sequences and position on the Cops-1 gene, **B**: schematic representation of the primer locations, **C**: PCR product results from *C. parvum* and *C. hominis* DNA.

Cgd2_4380_2F and 2R were designed based on the 200 bp band sequences (as described in 5.3.3). They overlap with Cgd2_4380 F and R primers and they target the repeat region. PCR products using Cgd2_4380_2F and 2R primers amplified three bands (650, 450, 200 bp) from *C. parvum* Iowa and only two bands 450, 200 bp) from *C. hominis* DNA (Figure 5.5). No PCR product was amplified from TU502 DNA. Since repeats are prone to duplication and loss, the difference observed in Figure 5.5 could be the result of a reduced repeat number. Each of the three bands amplified from *C. parvum* DNA represents a repeat, while *C. hominis* PCR profile suggests that Cops-1 could be truncated at the C terminus in part from loss of a repeat.

In order to evaluate the magnitude of truncation of ChCops-1, several primer combinations were tested. The following primer combinations allowed amplification of *C. parvum* DNA only: Cgd2_4380 F/FR, Cgd2_4380_2F/FR and Cgd2_4380 FF/R, therefore, did not assist in uncovering the ChCops-1 sequence. Cgd2_4380 FF and 2R primers generated three bands of 900, 650 and 450 bp from *C. parvum* DNA and two bands of 650 and 450 bp from *C. hominis* DNA, thus enabling the retrieval of extra length of ChCops-1. Interestingly, no amplification was achieved from the TU502 isolate. The use of Cgd2_4380 FF and Cgd2_4380_flanking_3R primers allowed amplification of a faint band > 1500 bp from Ch3 (not shown). PCR products from Ch2, Ch4 and TU502 were negative.

All the positive PCR products were sequenced and assembled to retrieve ChCops-1 gene sequence. The 5' end was conserved between ChCops-1 and CpCops-1. PCR products from *C. hominis* clinical isolates (Ch2, Ch3 and Ch4) were identical. Sequencing of Ch3 PCR product (using Cgd2_4380 FF and Cgd2_4380_flanking_3R primers) enabled to retrieve the 3' end of ChCops-1 gene and the sequences were validated at two independent sequencing facilities as described in 2.4. The sequencing allowed the retrieval of 1737 bp from *C. parvum* Iowa and 1559 bp from *C. hominis* Ch3 isolate (Appendix VIII).

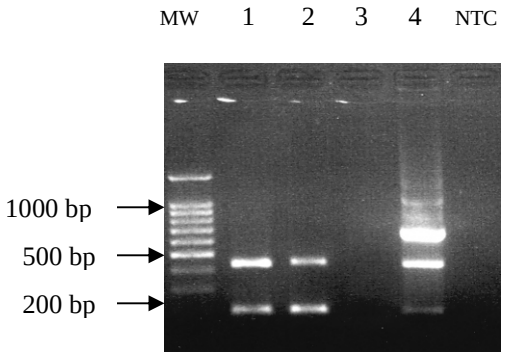


Figure 5.5: PCR results of Cgd2_4380_2F and 2R primers using *C. parvum* and *C. hominis* DNA. From *C. hominis* DNA, 2 bands of 450 and 200 bp were amplified. With *C. parvum* DNA, an additional band of 650 bp was also visible. No amplification using these primers was observed from TU502 DNA. MW: molecular weight, 1: Ch3, 2: Ch4, 3: TU502, 4: Iowa, NTC: non template control.

Using these primers (Cgd2_4380 FF and Cgd2_4380_flanking_3R), the PCR product also included the downstream flanking region, therefore, the 3'end of the ChCops-1 gene was determined by the presence of a stop codon TGA at position 1263. Therefore, ChCops-1 seems to be truncated to 1263 bp, while CpCops-1 is 1434 bp (Appendix VIII). Based on the ChCops-1 sequence, a reverse primer (Cops-1_*C. hominis*_R) targeting the 3' end of the gene was designed, however, PCR product using Cgd2_4380_FF and Cops-1_*C. hominis*_R primers failed to amplify DNA from either *C. hominis* or *C. parvum* isolates.

5.3.6 Comparison of CpCops-1 and ChCops-1

Cops-1 gene was first thought to be *C. parvum* specific. Subsequent analyses showed that the gene is also present in *C. meleagridis*. In addition, a truncated ortholog was detected in *C. hominis*, which was determined using a primer walking approach to be 1263 bp. Therefore, Cops-1 is considered an orthologous gene and CpCops-1 and ChCops-1 were used to name the gene in *C. parvum* and *C. hominis*, respectively. The CpCops-1 sequence in *C. parvum* clinical isolates is identical to the published Cops-1 gene sequence (Cgd2_4380). ChCops-1 has eluded *C. hominis* genome project. In *C. hominis* clinical isolates (ch2, ch3, ch4) ChCops-1 sequence was identical on only a portion of the gene (650 bp). The full gene length was amplified and sequenced from only Ch3 isolate. The full length CpCops-1 and ChCops-1 are 1434 and 1263 bp, respectively. The two sequences exhibit 78.8% sequence identity when tested by GENESTREAM software (Pearson *et al.*, 1997). The Alignment of CpCops-1 and ChCops-1 is shown in Appendix VIII. The difference in size between CpCops-1 and ChCops-1 corresponds to a truncation of ~ 170 bp in *C. hominis*. Alignment of CpCops-1 and ChCops-1 revealed 101 SNPs. This corresponds to an average of one SNP every 13 nucleotides for *C. hominis* and one SNP every 15 nucleotides for *C. parvum*.

CpCops-1 and ChCops-1 encoded proteins are predicted to be 477 aa and 420 aa, respectively. The protein sequences are 70% identical. The protein sequence alignment are shown in Figure 5.6. The N terminus of the CpCops-1 and ChCops-1

encoded protein is conserved, thus maintaining the gene's features (signal peptide, transmembrane domain and myristoylation motif). In addition, ChCops-1 has also a cleavage site between position 24 and 25 "FQT-FF". The biological significance of these findings needs to be assessed, especially that the predicted features, at the N terminus, support a role of cops-1 in host adaptation and virulence. Based on the protein sequences alignment, 79 aminoacid substitutions were detected, therefore 78.3% of the SNPs are non synonymous.

5.3.7 Cloning of CpCops-1

CpCops-1 gene was cloned in order to characterize the encoded protein. The full length gene was amplified from Iowa DNA. Two approaches were used to clone the Cops-1 gene. Cloning was performed into pGEM®-T Easy cloning vector and pTrcHisA expression vector. Alternatively, the full gene length was cloned using Champion™ pET100 Directional TOPO® Expression Kit. The construct orientation and sequence was checked by PCR using Cgd2_4380_FF and T7 terminator primers. Colonies which had the correct insert were used for expression of the recombinant protein. A map of the recombinant pET100/D-TOPO® vector is shown in Appendix IX.

5.3.8 Expression of Cops-1 recombinant protein

Both the recombinant pTrcHisA vector and pET100/D-TOPO® vector were trialled for the expression of the recombinant Cops-1 protein as described in 2.7.5. A pilot expression study was performed to determine the optimal expression time.

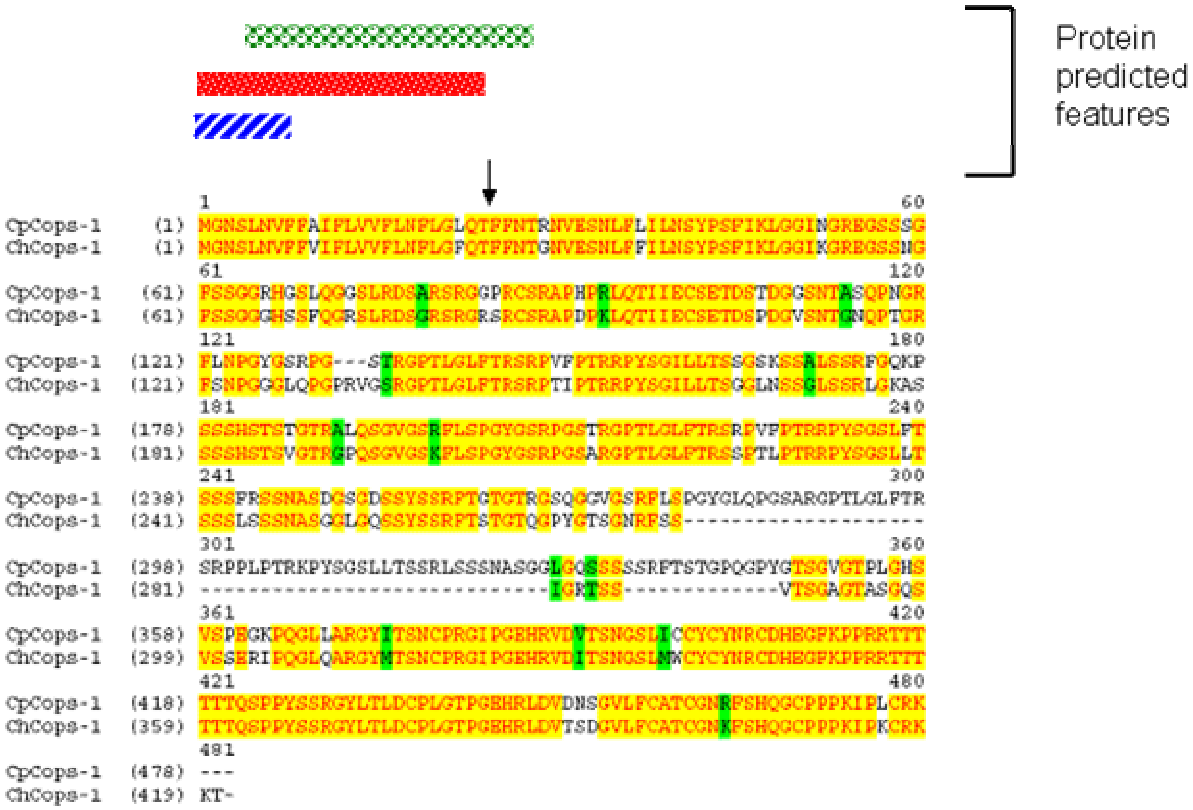


Figure 5.6: Protein sequence and characteristics of CpCops-1 and ChCops-1. The two proteins are 70% identical. The N terminus is conserved between *C. hominis* and *C. parvum*, thus maintaining the gene predicted features: signal peptide, transmembrane domain and myristoylation motif.

For DH5α transformed using the pTrcHisA recombinant vector, both supernatant and pellet samples failed to demonstrate detectable expression of the recombinant protein (~ 50 kDa) after Coomasie staining. Because DH5α is not an expression bacterial strain, the recombinant pTrcHisA vector was transformed into BL21 bacterial strain adapted for expression of recombinant proteins and the level of expression was compared between the two strains after 4h induction (Figure 5.7). No detectable expression of the recombinant protein was achieved by BL21 either.

In addition, the expression of the recombinant His-tagged proteins was tested by Western Blot revealed with anti-His peroxidase conjugated secondary antibody. Western Blot analysis supported SDS-PAGE results as no His-tagged proteins were detected at different time points and with the two bacterial strains. Similarly, the use of the high efficiency quick cloning and expression kit Champion™ pET100 Directional TOPO® allowed successful cloning of the Cgd2_4380 gene but no expression of the recombinant protein was detected, even though high levels of His-tagged recombinant β-galactosidase, used as a control, was detected.

The absence of expression of the recombinant protein despite the stability of the construct in the bacterial host suggests some limitation in protein translation. A Cops-1 protein sequence analysis, performed using ProtParam software (http://www.expasy.ch/tools/protparam.html), showed that the following aminoacids are abundant in the Cops-1 gene sequence: Serine: 17.4%, Glycine: 14.7%, Arginine: 9.4%). Therefore, a likely explanation for the lack of expression of the recombinant protein might be linked to differences in codon usage between the eukaryotic machinery of *Cryptosporidium* and the prokaryotic system of *E. coli*, especially for these predominant codons.

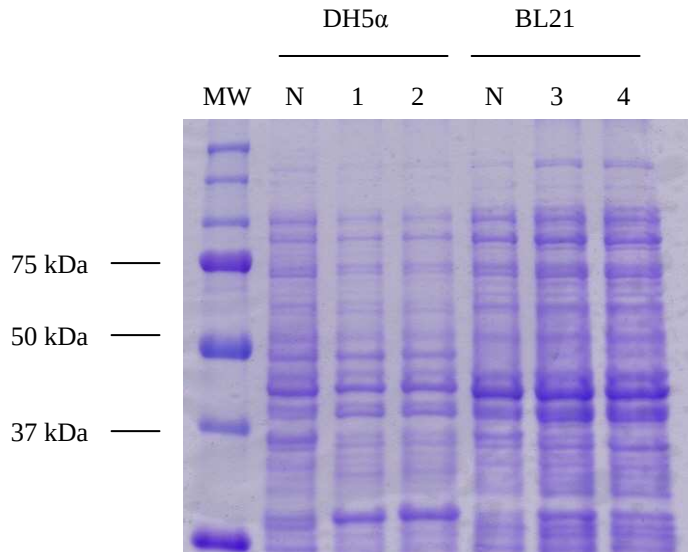


Figure 5.7: Comparison of the level of protein expression of two bacterial strains DH5α and BL21 Star™ (DE3) after 4h induction. There is no over-expression of a 50 kDa recombinant protein. MW: molecular weight marker, N: not induced, 1 and 2: recombinant DH5α clones harbouring the recombinant pTrcHisA vector, 3 and 4: recombinant BL21 Star™ (DE3) clones harbouring the recombinant pTrcHisA vector.

To confirm this hypothesis, the codon frequency of the Cops-1 gene was calculated using the codon usage bioinformatics software (www.ebioinfogen.com/biotools/codon-usage.htm). The codon frequency analysis showed that 17 codons occur at a frequency ≥ 2% (Table 5.1).

The average usage for each coding trinucleotide by *C. parvum* and *E. coli* was published by Fayer (1997). The predominant codons in the Cops-1 gene were assessed for efficient translation by the *E. coli* machinery. 11/17 (64.7%) codons had at least 20% difference in the average codon usage between *C. parvum* and *E. coli* (Table 5.1). For example, the codons GGA, TTA, AAT, CCA, CAA and AGA are extensively used by *Cryptosporidium* in comparison to *E. coli*. These results could explain the inability of *E. coli* bacterial strains to efficiently express the recombinant Cops-1 protein.

5.3.9 Use of specialized bacterial strains for the expression of recombinant protein

In order to resolve the codon bias observed, specialized bacterial strains were assessed for the expression of Cops-1 recombinant protein. This includes strains engineered for the expression of membrane proteins (C41 and C43) and strains having tRNAs for rare codons (Rosetta™2 and BL21-CodonPlus-RP) as detailed in 2.7.4. The recombinant pET100/D-TOPO® vector was used to transform these bacterial strains. Although several C41 colonies were Ampicillin resistant, screening for Cops-1 by PCR was negative. However, C43, Rosetta™2 and BL21-CodonPlus-RP strains successfully maintained the recombinant plasmid. After induction, the bacterial lysates were used to visualize protein expression. SDS-PAGE stained with Coomasie blue did not show clear expression of the recombinant protein at 50 kDa. However, immunoblot showed distinctive expression of the recombinant protein from 1h post-induction (Figure 5.8). The recombinant His-tagged protein has an approximate size of 53 kDa to account for the N terminal tag. No his-tagged protein was observed in the non-induced lysate.

Calculated codon frequency (codon usage bio tool)			Percent codon used by (Fayer, 1997)	
Amino acid	**Codon**	**Number (%) within Cgd2_4380 gene sequence**	***C. parvum* codon usage**	***E. coli* codon usage**
Phe	TTT	25 (5.3%)	63	50
Gly	GGA	29 (6%)	37	9
Gly	GGT	30 (6.3%)	50	37
Leu	TTG	11 (2.3%)	23	12
Leu	TTA	16 (3.4%)	35	10
Asn	AAT	16 (3.4%)	68	40
Pro	CCA	33 (7%)	62	19
Gln	CAA	13 (2.8%)	80	30
Arg	AGA	22 (4.7%)	70	4
Arg	CGT	13 (2.8%)	17	43
Ser	AGT	19 (4%)	17	13
Ser	TCA	30 (6.3%)	33	11
Ser	TCT	21 (4.4%)	34	18
Thr	ACA	18 (3.8%)	42	12
Thr	ACT	19 (4%)	47	19
Val	GTT	10 (2%)	54	28
Tyr	TAT	12 (2.6%)	73	53

Table 5.1: Codons present at a frequency ≥ 2% in the Cgd2_4380 gene as determined by the codon usage software. The percentage was calculated based on the total number of 478 trinucleotides (including stop codon). The average of codon usage in percentage was adopted from Fayer (1997). The highlighted codons showed over 20% difference between *C. parvum* and *E. coli* codon usage.

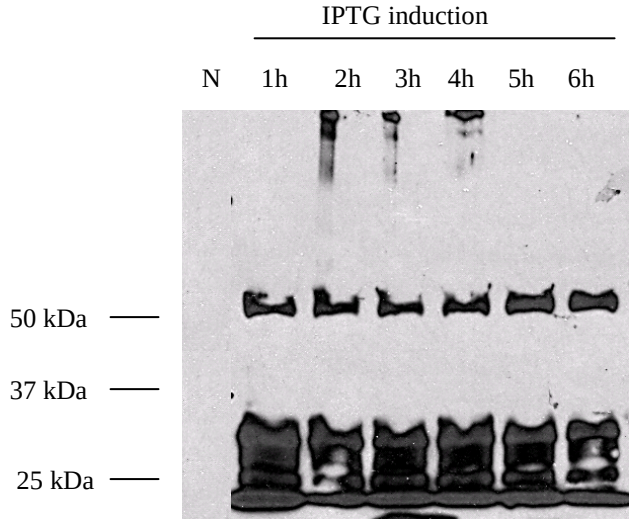


Figure 5.8: Immunoblot using anti-polyhistidine secondary antibody from BL21-CodonPlus-RP lysates after IPTG induction. The bacterial lysates showed an expression of the His-tagged recombinant protein of ~ 50 kDa from 1h up to 6h after induction. N: not induced, 1-6: cell lysate were taken at one hour interval after IPTG induction.

The level of expression of the recombinant protein from the different bacterial strains was assessed 5h post induction. Immunoblot showed similar levels of expression of the recombinant protein between Rosetta™2 and BL21-CodonPlus-RP strains (Figure 5.9). No expression of the His-tagged protein was visible before induction. C43 strain, however, did not allow expression of the full length recombinant protein although some short fragments were detected (Figure 5.9).

Both immunoblots (time course expression and comparison of expression levels between bacterial strains) showed that there is either degradation of the newly produced recombinant protein as early as 1h post induction or production of truncated product. The same profile occurred also after increasing the amount of protease and phosphatase inhibitors and adopting more stringent handling procedures.

5.3.10 Monoclonal peptide antibody anti-Cops-1

In order to investigate the biological function of Cops-1 encoded protein, a monoclonal peptide antibody was produced as described in 2.6.3. Initially, two peptide regions were chosen due to high immunogenicity and are shown in Figure 5.10. The peptide sequences were synthesized and used to immunize mice. Mice bleeds were screened by ELISA and the spleens of the two highly reactive mice were used for cell fusion. For Tyle-1, screening of fusion cells showed only weak and unstable IgG production. In addition, no specific IgG or IgM secretion was detected after ten days culture. Therefore, no anti-Tyle-1 antibody was produced. For Tyle-2 peptide, 13/900 clones were ELISA positive (OD450> 0.5). All 13 clones were expanded to assay viability, growth and stability and tested by specific anti-Tyle-2 IgG ELISA to check for stable IgG production. Only four hybridoma cultures expressed high specific antibody in their supernatant. All supernatants were further screened by IFA using purified Moredun oocyst suspension as described in 2.9. There was good concordance between ELISA and IFA results. The 9E1 hybridoma supernatant was found to have the strongest reactivity and was used for further IFA staining.

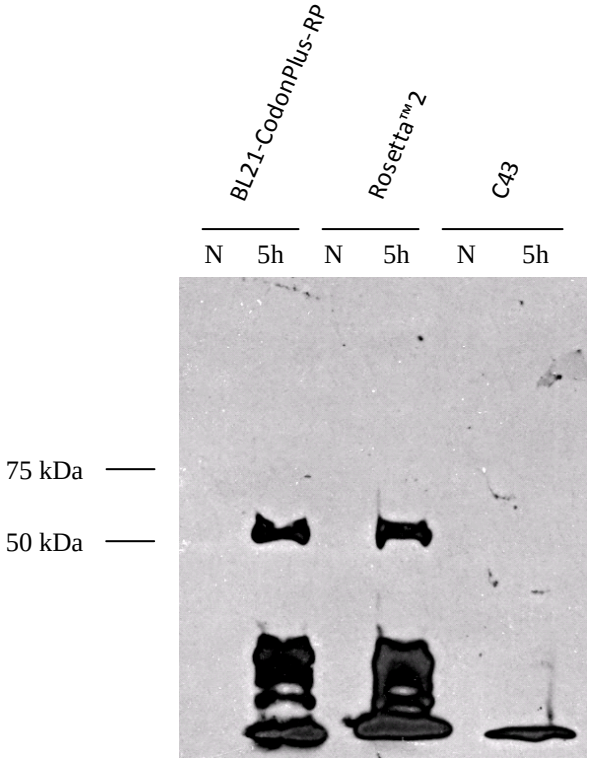


Figure 5.9: Comparison of the level of expression of the His-tagged recombinant Cops-1 protein among the three specialized strains (BL21-CodonPlus-RP, Rosetta™2 and C43 strains) 5h post induction. BL21-CodonPlus-RP and Rosetta™2 strains exhibited the same level of expression, while C43 strains failed to express the recombinant protein.

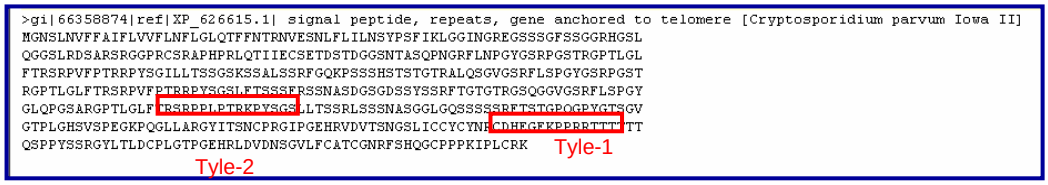


Figure 5.10: Location and sequence of the two peptides Tyle-1 and Tyle-2 assayed for the production of peptide antibody at CRB. The peptide sequences were synthesized at high purity and used for mice immunization.

5.3.11 Immunolocalization of Cops-1

The 9E1 monoclonal antibody was used to stain fixed parasites as described in 2.9 to identify the location of Cops-1 protein. In addition, a FITC-conjugated antibody that stains the oocyst cell wall of *Cryptosporidium* species and DAPI staining of the sporozoites nuclei were used as counter stains. For *C. parvum* oocysts (Iowa and Moredun), the 9E1 monoclonal antibody clearly recognized the oocyst content (Figure 5.11) and the stain partially colocalized with DAPI staining of sporozoites, 9E1 stained intact oocysts, broken oocysts still preserving their sporozoites and free sporozoites (Figure 5.12). No 9E1 staining of empty oocysts was observed. 9E1 antibody enabled staining of intact oocyst contents when permeabilized. For impermeabilized oocysts (Paraformaldehyde fixation), oocyst content was stained only when the oocyst wall was broken. This staining pattern strongly implies that cops-1 is a surface sporozoite protein, which is not inconsistent with the protein being secreted.

9E1 monoclonal antibody was also used to stain *C. hominis* purified oocyst suspensions from clinical sources (Ch3, Ch4) as described in 2.9.1. *C. hominis* oocysts showed weak staining of the oocyst content in comparison with the level of staining of *C. parvum* oocysts, which can be difficult to observe (Figure 5.13). Several secondary antibodies (FITC Goat Anti-Mouse IgG, Cy™5- conjugated Donkey Anti-Mouse IgG, Alexa Fluor®546 Goat Anti-Mouse IgG) were tried as described in 2.9.1. The level of staining of *C. hominis* oocysts was similar for all these secondary antibodies. Subsequently, a two-step staining method using Biotinylated Rat Anti-Mouse IgG revealed with streptavidin substrate conjugated to Alexa Fluor®568 was used to improve the binding specificity. Even with the use of these reagents, residual staining of *C. hominis* oocysts was still observed and persisted even after adding a blocking step using a Streptavidin/Biotin Blocking kit.

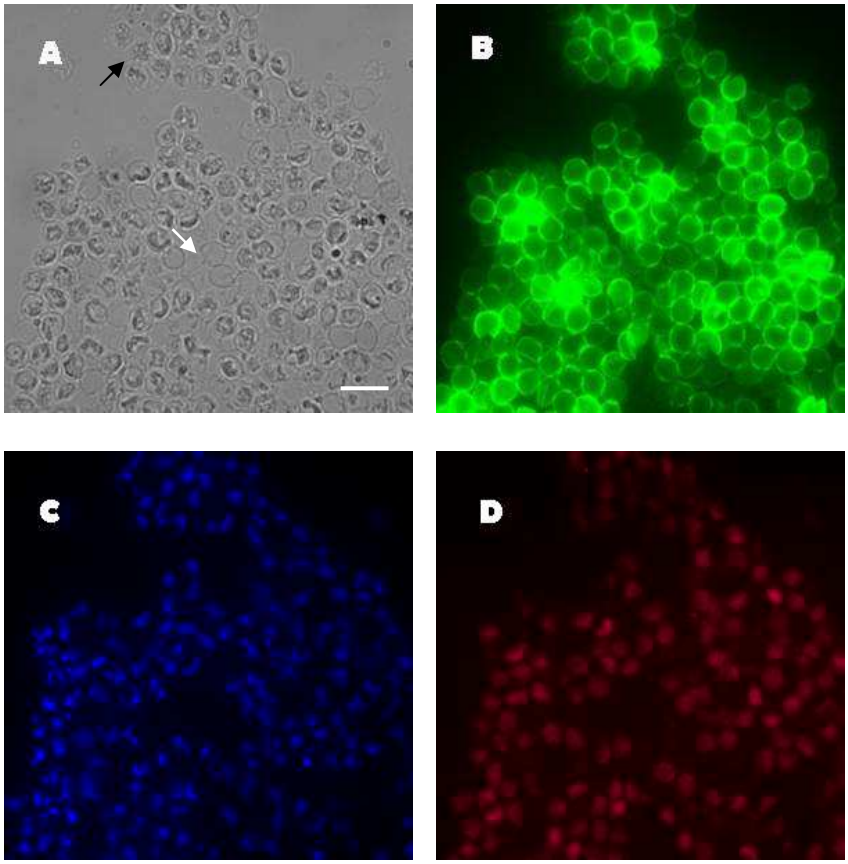


Figure 5.11: Staining of purified *C. parvum* oocysts to identify the location of Cops-1 protein. **A**: DIC (Differential Interference Contrast) microscopy showed both empty (white arrowhead) and intact (black arrowhead) oocysts. **B**: FITC oocyst wall staining using Crypto-cell commercial antibody. **C**: DAPI staining of sporozoite nuclei. **D**: 9E1 staining of oocyst content revealed with Cy™5-conjugated Donkey Anti-Mouse IgG. Scale bar= 10 µm.

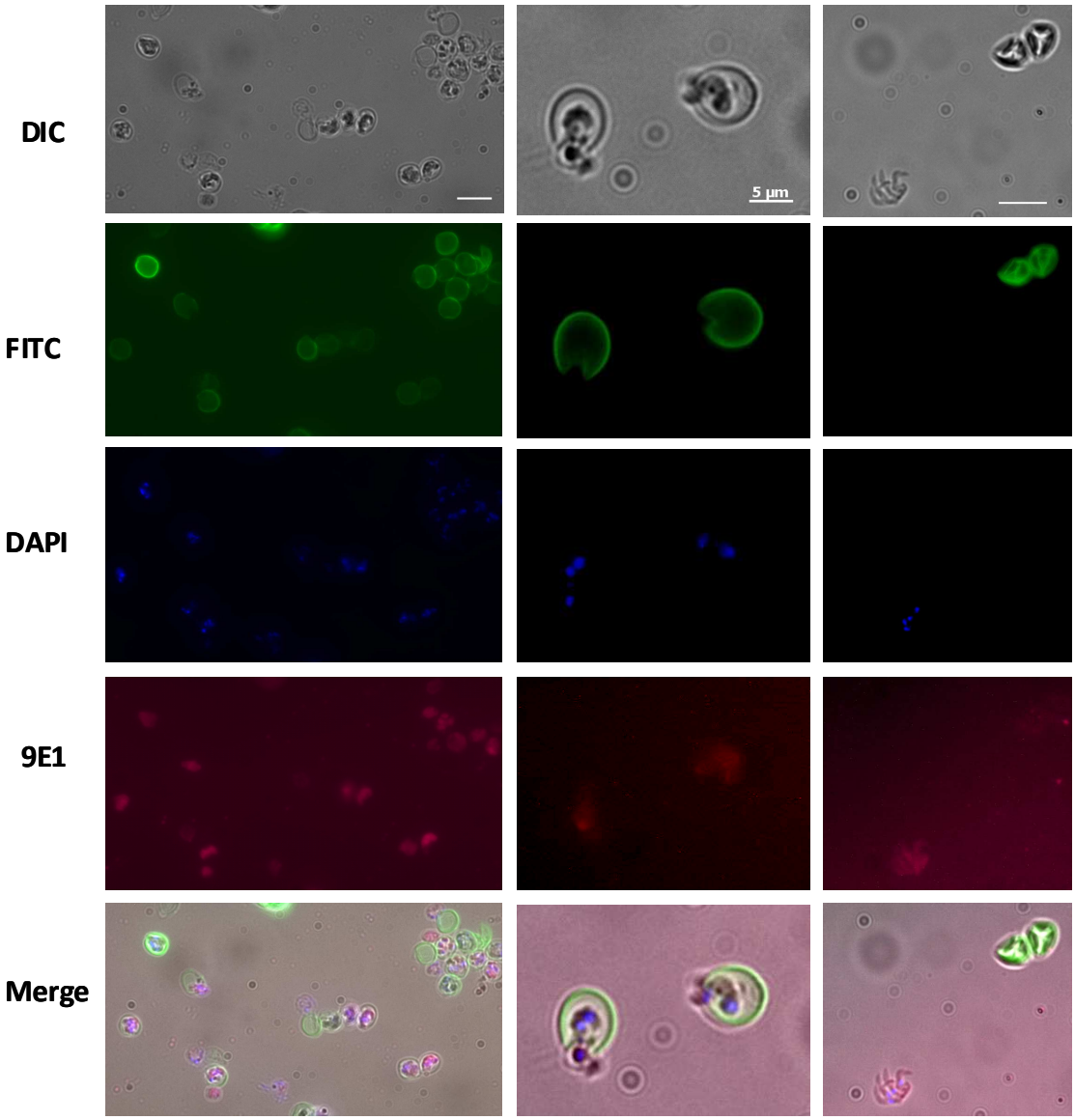


Figure 5.12: Microscopic observation of stained *C. parvum* oocysts and sporozoites using DIC, FITC, DAPI and Alexa Fluor®546 filters. The bottom picture is a merge of the 4 fields. 9E1 staining colocalized with DAPI staining of sporozoites in intact and ruptured oocysts and in free sporozoites. Scale bar= 10 µm (unless stated otherwise).

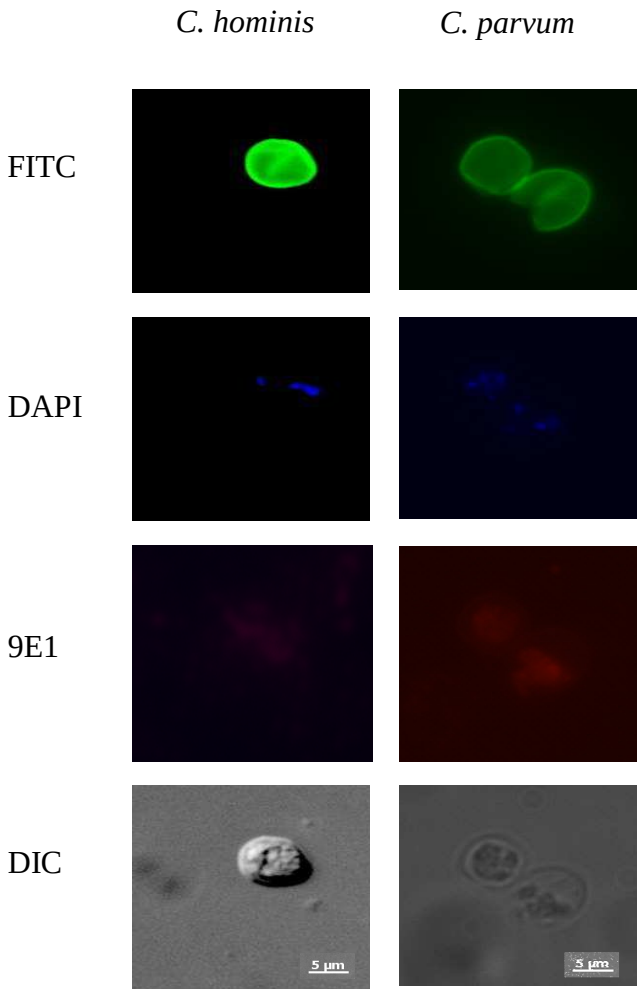


Figure 5.13: Comparison of the level of staining of 9E1 antibody of *C. hominis* and *C. parvum* purified oocysts. Microscopic observation of stained oocysts using DIC, FITC, DAPI and Alexa Fluor®546 filters. 9E1 antibody stained oocyst content, while *C. hominis* staining was very faint.

The alternative of using a directly conjugated antibody offers the advantage of eliminating possible cross reaction of conjugated secondary antibody with the sample. Directly conjugated 9E1 antibody was tested to improve staining of *C. hominis* oocysts. Direct conjugation of 9E1 was performed using commercial direct conjugation kits. These kits are subtype specific, therefore, the class and subclass of 9E1 antibody was determined using Iso-Gold™ Rapid Mouse-Monoclonal Isotyping Kit as described in 2.9.2. The 9E1 monoclonal antibody is IgG1 as shown on Figure 5.14. Subsequently, Zenon® Alexa Fluor®594 Mouse IgG1 Labelling kit was used to produce directly conjugated 9E1 antibody. However, the staining profile using this antibody was similar to those obtained with secondary antibodies previously used. Alternatively, Alexa Fluor®594 Monoclonal Antibody Labelling kit was also used for direct conjugation of 9E1. The direct conjugated antibody also showed residual staining of *C. hominis* oocysts. The weak and diffuse staining of *C. hominis* oocysts has been consistent using different secondary antibodies and two direct conjugation kits.

5.3.12 Gliding Assay

It has been demonstrated that some antibodies raised against sporozoites surface proteins were able to stain gliding trails formed by sporozoites after excystation when they are left to glide against glass or plastic surfaces. In order to assess if Cops-1 is likely to be a sporozoite surface protein shed in trail, a gliding assay was performed as described in 2.10 and the slides stained with 9E1. In addition, a monoclonal antibody 4C1 was used as a positive control. 4C1 is an anti-*C. parvum* monoclonal antibody raised against sporozoite proteins shed during gliding motility of *Cryptosporidium* sporozoites (Feng *et al.*, 2006). This 4C1 antibody stained trails left by *Cryptosporidium* sporozoites when gliding on the slide as shown in Figure 5.15. However, no FITC staining of gliding trails was visible after staining with 9E1 antibody (not shown).

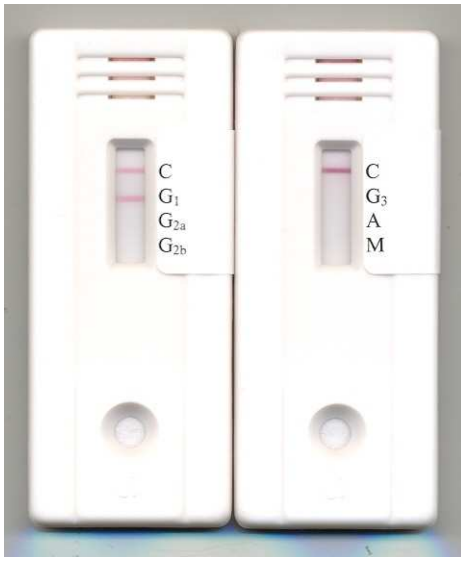


Figure 5.14: Typing of 9E1 monoclonal antibody using the Iso-Gold™ Rapid Mouse-Monoclonal Isotyping Kit, showing that 9E1 subclass is IgG1. C is a control band for each cassette.

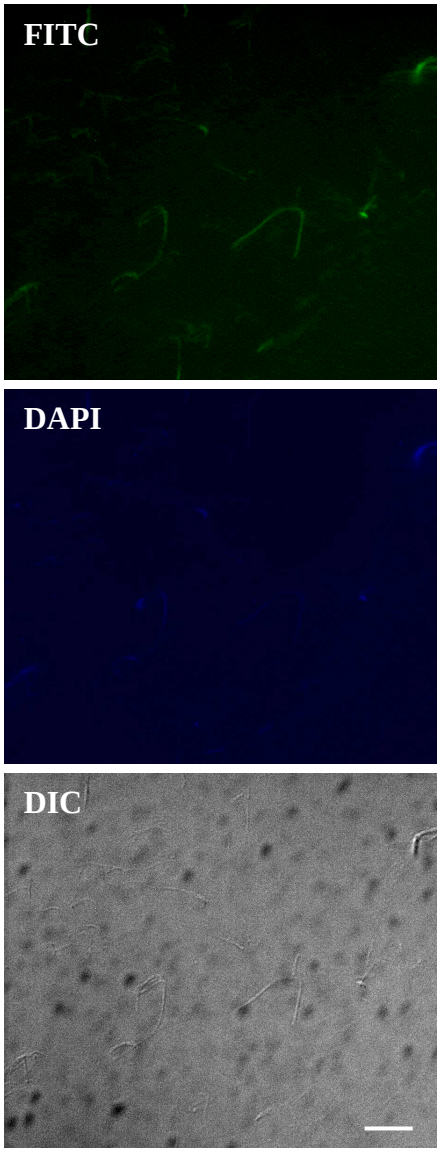


Figure 5.15: Gliding trails of *Cryptosporidium* revealed using 4C1 monoclonal antibody. Scale bar= 10 µm.

5.3.13 Coculture of *Cryptosporidium* oocysts with Caco-2 cell monolayers

Cryptosporidium oocysts and sporozoites were used to infect Caco-2 cell monolayers as described in 2.11. The excysted oocysts were added to the cells, after incubation, the non-adherent parasites were washed in PBS and supplemented DMEM medium was added to the cells. The cultures were subcultured for further 24h. *Cryptosporidium* intracellular life stages were visible in cell monolayers after Giemsa staining (Figure 5.16).

5.3.14 Effect of 9E1 on *Cryptosporidium* invasion of Caco-2 cells *in vitro*

The investigation of a potential role of Cops-1 protein in *Cryptosporidium* attachment and invasion of host cells was undertaken using an *in vitro* blocking assay as described in 2.11.7. Pre-incubation of the parasite with a neutralising blocking antibody was shown to inhibit host cell invasion by *Neospora caninum* (Nishikawa *et al.*, 2000). A similar blocking assay was developed for *Cryptosporidium*. Briefly, *Cryptosporidium* oocysts were pre-incubated with 9E1 monoclonal antibody at 1mg/ml and 0.1 mg/ml. In addition, a mouse IgG1 antibody (0.1 mg/ml) was used as a negative control. As a positive control, oocysts were incubated in PBS. For each condition, the mixture was used to infect Caco-2 cell monolayers as described previously. After 24h subculturing, the numbers of parasites was counted in 20 random fields as shown in Table 5.2. The mean number of parasites per field was then determined as shown in Figure 5.17 to allow comparison between the different conditions. Pre-incubation of *Cryptosporidium* oocysts with 9E1 did not influence Caco-2 cells invasion. The number of parasites detected in cell monolayers was comparable when the parasites were incubated with PBS or with 0.1 or 1 mg/ml 9E1: 113 versus 122 and 117, respectively (Figure 5.17).

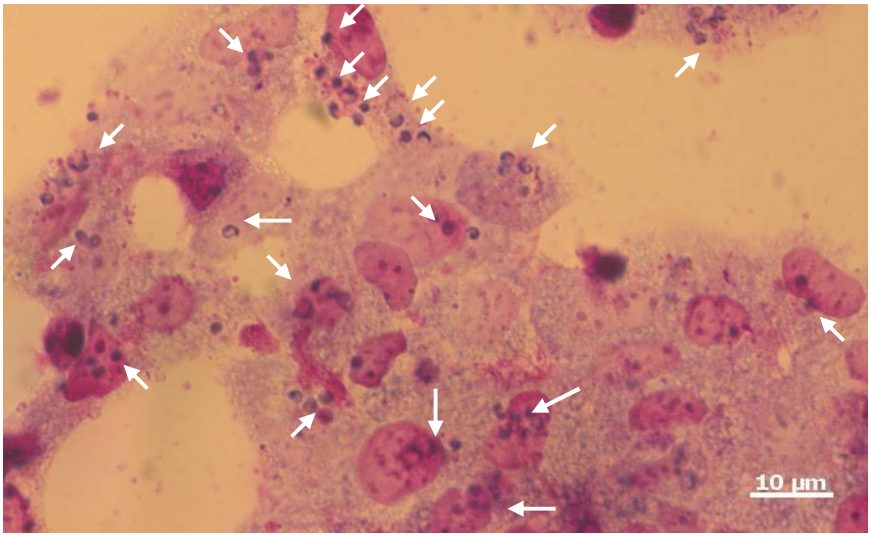


Figure 5.16: Intracellular life stages of *Cryptosporidium* visualized after Giemsa staining of infected Caco-2 cell monolayers.

No antibody	Mouse IgG (0.1mg/ml)	9E1 (0.1 mg/ml)	9E1 (1 mg/ml)
101	36	145	153
83	58	60	178
127	77	69	119
78	77	104	116
144	75	100	53
89	111	114	108
187	101	152	67
95	67	196	79
86	93	101	101
91	43	110	139
63	159	135	95
58	85	76	45
131	103	59	137
102	100	125	153
85	102	125	153
90	88	121	83
135	140	215	64
148	66	136	149
211	73	114	155
161	124	184	197

Table 5.2: Number of parasites detected for each randomly selected field for each of the different conditions tested. Parasites were pre-incubated with 9E1 antibody at 2 different concentrations, with PBS as a positive control and a mouse IgG antibody as a negative control before infection of Caco-2 cell monolayers.

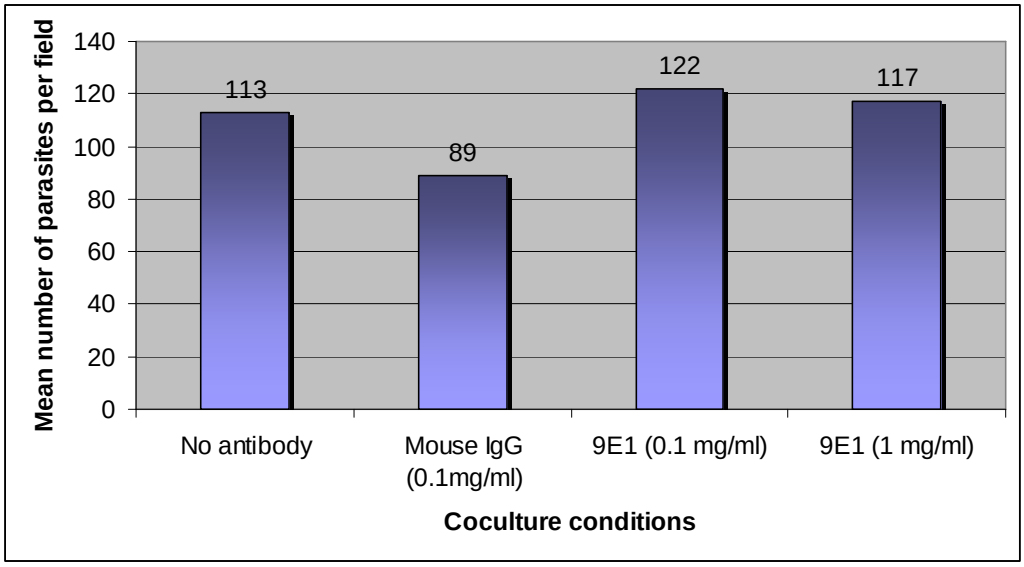


Figure 5.17: Mean number of parasites per field for the different coculture conditions. Using 2 different concentration of 9E1 monoclonal antibody did not influence the level of invasion of Caco-2 cells by *Cryptosporidium* oocysts.

5.3.15 Screening of recombinant and native antigen preparations using sera from natural *Cryptosporidium* infection and 9E1 monoclonal antibody

Recombinant protein preparations from BL21-CodonPlus-RP and Rosetta™2 strains and native antigen preparations from *C. parvum* and *C. hominis* oocysts prepared as described in 2.12.1 were tested by Western Blot revealed using sera from patients infected with *C. parvum* and *C. hominis*. These sera originated from cryptosporidiosis cases, previously genotyped at CRU. In addition, a negative human serum for *Cryptosporidium* infection was tested as negative control.

For the recombinant protein preparations, several bands were revealed with the polyclonal goat anti-human IgG secondary antibody. Both induced and non-induced bacterial lysates were tested to confirm the detection of recombinant Cops-1 protein. There were no distinct profiles between the non-induced and the induced protein preparations for the two strains (Figure 5.18). A strong band of 50 kDa was visible for Rosetta™2 protein preparation, but it was also present in the non-induced bacterial lysate. This protein was absent from the induced lysate when the blot was treated with the negative serum. A 70 kDa protein was also detected with *C. parvum* infection serum from induced and non-induced protein preparations but was also visible with the negative serum. *C. hominis* serum did not reveal any band. The native antigen preparations were also tested using patient's sera. *C. parvum* serum detected a 75 kDa band from *C. parvum* and *C. hominis* antigen preparations; however, this band was also detected with the negative serum (Figure 5.18). This was also the case for a 20 kDa protein detected from *C. hominis* proteins.

Interestingly, a protein of 50 kDa was detected from *C. parvum* native antigen preparation but not from *C. hominis* when the blot was treated with *C. parvum* infection serum. This protein was not detected with the negative serum. It is possible that the 50 kDa protein detected from *C. parvum* native antigen preparations with a *C. parvum* infected patient is Cops-1.

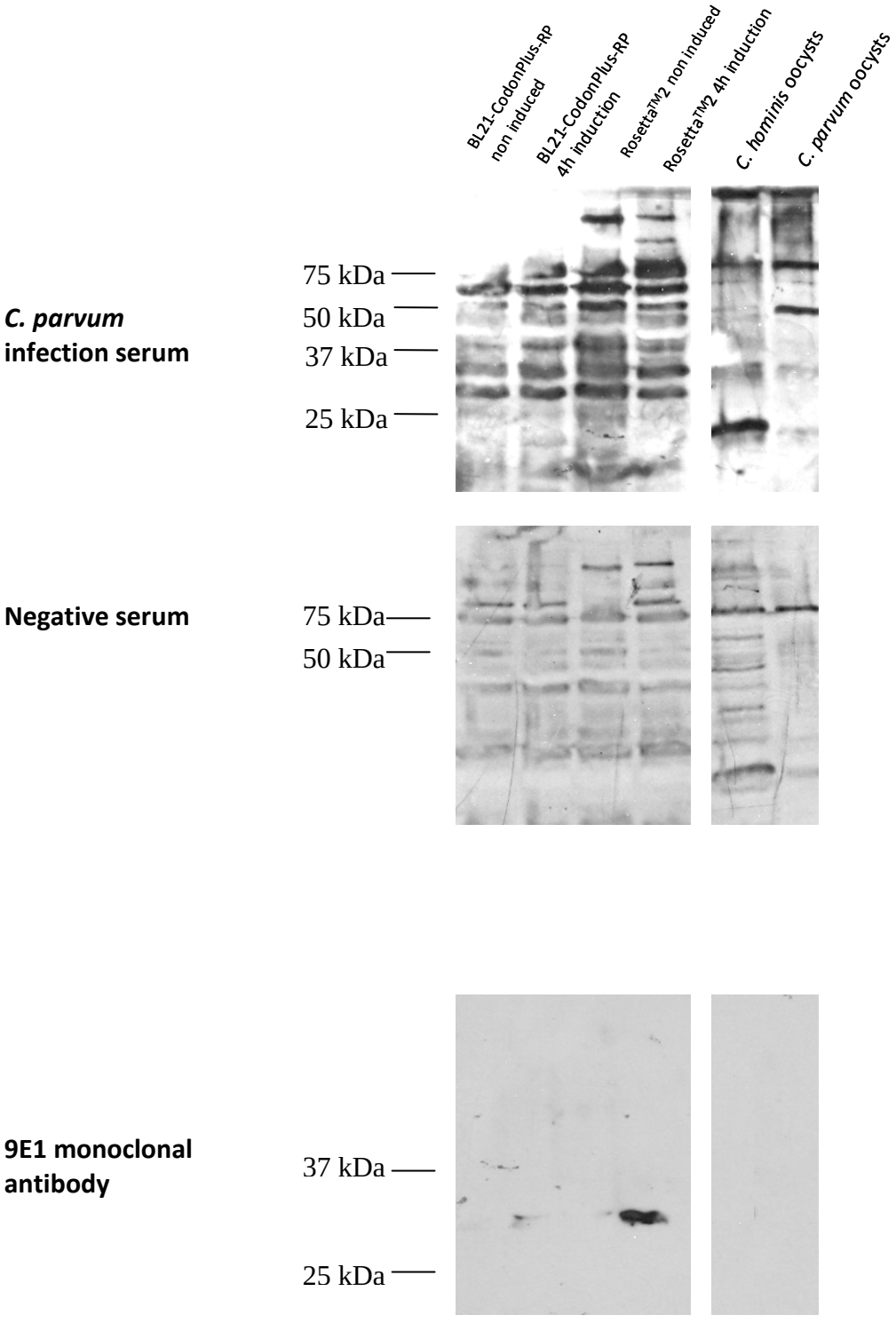


Figure 5.18: Immunoblot results of recombinant protein preparations from Rosetta™2 and BL21-CodonPlus-RP strains and native antigen preparations from *C. hominis* and *C. parvum* oocysts revealed with 9E1 monoclonal antibody, natural *C. parvum* infection serum and a negative serum.

The 9E1 antibody did not detect a 50 kDa protein from induced bacterial lysates and native antigen preparations. Only a band of 30 kDa from induced Rosetta ™2 bacterial preparation was detected, which is likely to be a product of degradation of the full length recombinant Cops-1 protein. No such band was detected from BL21-CodonPlus_RP induced lysate, despite a 30 kDa band being detected using the anti-His tagged antibody.

5.4 Discussion

The characterization of the Cops-1 gene on the genetic and molecular level was undertaken to achieve an improved understanding of the protein characteristics and function. Cops-1 was first thought to be *C. parvum* specific, as the right size PCR product was generated only from *C. parvum* DNA, which has lead to its re-labelling as *Cryptosporidium parvum* specific gene (Cops-1). Subsequently, sequencing of a smaller band of 200 bp generated using Cgd2_4380 F and R primers from both *C. parvum* and *C. hominis* showed high sequence similarity to Cgd2_4380 gene. This result suggested that Cops-1 has an ortholog in *C. hominis*. The two genes were named CpCops-1 and ChCops-1, for *C. parvum* and *C. hominis*, respectively.

In order to follow up this discovery, a new pair of primers (based on the retrieved gene sequence) was designed and tested. These primers (Cgd2_4380_2F and 2R) allowed amplification of two bands from *C. hominis* and three bands from *C. parvum*, thus increasing the amount of known sequence for ChCops-1. These results were supported by BLAST analysis showing three potential primer binding sites for Cgd2_4380_2R primer. The last binding site (generating the highest molecular weight band) seemed to be absent in *C. hominis,* thus supporting the hypothesis of partial gene truncation in *C. hominis*. The recurrent primer binding sites were due to the presence of repeats in the Cgd2_4380 as was suggested by the gene annotation. Interestingly, PCR products using Cgd2_4380_2F and 2R primers were obtained only from clinical isolates of *C. hominis* but not from the reference strain TU502.

This might be explained by high sequence variability between the clinical isolates and reference strain of *C. hominis* and could justify the absence of the Cgd2_4380 gene from the *C. hominis* genome sequence.

Sequence analysis of Cgd2_4380 F and R primers PCR products showed that clinical isolates of *C. parvum* had an identical sequence to the published gene sequence. The anthroponotic group had four specific SNPs. *C. meleagridis* DNA gave the correct size PCR product (655 bp) using Cgd2_4380 F and R primers, suggesting that Cops-1 is not exclusively a *C. parvum* gene. This result supports the phylogenetic relatedness of *C. meleagridis* to *C. parvum* as discussed in Chapter 3. The *C. meleagridis* PCR product had only one SNP. *C. hominis* PCR products, despite being shorter, showed some level of sequence variability as four species-specific SNPs were detected from the 200 bp product. 75% of the SNPs detected were non synonymous. This proportion of non synonymous SNPs per gene is within the range (0%-83.4%) detected in the other genetic loci as shown in Chapter 3.

The identification of the full ChCops-1 gene sequence was undertaken using a primer walking approach. Not all primer combinations allowed DNA amplification from *C. hominis* isolates. This is probably due to targeting conserved versus deleted regions of the gene. Surprisingly, some primer combinations worked with the Ch3 isolate only. Thus, the determination of the full gene sequence was possible only from Ch3. ChCops-1 seems to be truncated to 1263 bp in this isolate. The sequence showed 78.8% sequence identity to the published Cgd2_4380 gene. Alignment of CpCops-1 and ChCops-1 sequences showed high sequence variability, 101 SNPs were detected in addition to few nucleotide insertions and deletions. On average this corresponds to 1 SNP for every 13 nucleotides for *C. hominis* and 1 SNP for every 15 nucleotides for *C. parvum*. This is much higher than the most variable gene detected during this study Cgd2_2430 gene, for which an average of 1 SNP per 30 nucleotides was detected (Chapter 3). Despite this high sequence variability, the proportion of non-

synonymous SNPs for the whole gene sequence was 78.3% which is within the range detected in the other genetic loci tested.

Based on the ChCops-1 sequence, showing a conserved 5' end, a reverse primer targeting the 3' of the gene (Cops-1_*C. hominis*_R) was designed. PCR using Cgd2_4380 FF and Cops-1_*C. hominis*_R primers did not allow amplification from Ch3 or any other *C. hominis* isolates. At this stage, a technical limitation was encountered, testing a newly extracted Ch3 isolate DNA (from the same oocyst suspension stored at CRU) did not allow the previous results to be reproduced. This new sample differed in DNA quantity and quality, which may have influenced the PCR results. In addition, degradation of oocyst suspensions is likely after long term conservation at 4°C, but PCR reactions using other targets was positive. A new pair of primers aiming to amplify the full ChCops did not allow DNA amplification from Ch3 or any other *C. hominis* isolate. This suggests that the ChCops-1 may not be conserved amongst *C. hominis* clinical and reference strains. This is supported by the high level of sequence variability that was detected. However, it is clear that the internal part of the Cgd2_4380 gene is conserved among *C. hominis* clinical isolates since internal primers showed consistent amplification of DNA from *C. hominis*. The 5' end of the gene seems also to be conserved in *C. hominis* isolates as confirmed by sequencing. The determination of the ChCops-1 3' end, however, proved more challenging. Because of these limitations, the full length ChCops-1 sequence requires further validation and it is still unclear if Cops-1 is functional in *C. hominis* species.

The potential of Cops-1 gene to be a species determinant was considered largely on account of the initial PCR results. In addition, the predicted protein features are highly suggestive of a role in host-parasite interaction. Particularly, the telomeric location indicates that the gene is prone to higher recombination rates and it is likely to be a contingency gene. Such genes were shown to be involved in host pathogen interaction and parasite survival in the host, such as VSG gene of *T. brucei*, which undergoes antigenic variation to evade the host immune responses and allow parasite survival (Yang *et al.,* 2009, Barry *et al.,* 2003). The

predicted protein has at its N terminus a signal peptide, a transmembrane domain and a myristoylation motif implying that Cops-1 encodes a secreted protein likely to be involved in host-pathogen interaction. Myristoylation of protein is a translational event catalyzed by the enzyme N-myristoyl transferase (NMT), which attaches myristic acid to the N-terminal glycine residues of eukaryotic and viral proteins (Farazi *et al.,* 2001, Poli *et al.*, 1991, Towler *et al.* 1988). Myristoylated proteins aid subcellular targeting, protein-protein interaction and are involved in signal transduction cascades. NMTs have been characterized in several protozoan parasites. In *Leishmania* major and *Trypanosoma brucei*, NMT is a 48.5 kDa protein that localizes to both membrane and cytoplasmic fractions, is expressed in all life stages and was shown to be essential for viability (Price *et al.,* 2003). NMT genes were detected in *C. parvum*, *C. hominis* and *C. muris* genome sequences.

Cops-1 Protein BLAST showed low sequence identity to the *Drosophila melanogaster* gene "*shot*" (short stop). The *shot* gene encodes actin-binding proteins and is predicted to contain, at the N terminus, an actin-binding domain and at the C terminus two EF-hand calcium-binding motifs (Lee *et al.,* 2000 b). *Shot* has been shown to coordinate actin and microtubule dynamics, which is important for cellular motility and morphogenesis (Lee and Kolodziej, 2002). Strangely, no EF hand or actin binding domain were detected by the analyse software. This suggests that Cops-1 protein may not share functional features with *shot*.

In order to characterize the Cops-1 encoded protein, cloning of the full gene length from *C. parvum* Iowa DNA and expression of the recombinant protein in bacterial strains was performed. Cloning of the Cops-1 gene into various cloning and expression vectors was successful; however, the expression of the recombinant protein was less straightforward. Despite the stability of the recombinant plasmid in the bacterial strain, no expression of the recombinant protein was initially observed on SDS-PAGE or by Western Blot. This observation was suggestive of a translation malfunction, restraining the expression of the

recombinant protein. This could be driven by the differences in codon usage between the eukaryotic machinery of *Cryptosporidium* and the prokaryotic system of *E. coli*, which is particularly relevant for predominant codons in gene sequence. The codon composition for Cgd2_4380 gene was calculated and showed that 17 codons occurred at a frequency greater than 2%. Of the 17, 11 (64.7%) had at least 20% difference in the average codon usage between *C. parvum* and *E. coli*. A biased codon usage in *Cryptosporidium* genes has been reported and was mainly attributed to the A+T rich genome (Grocock and Sharp, 2001, Char *et al.,* 1996). The codon bias has been anticipated by Fayer (1997) to limit the performance of *E. coli* recombinant expression libraries, as many *Cryptosporidium* DNA sequences may not be translated at detectable levels because the tRNAs necessary for efficient expression are present at too low concentration. Therefore, strains having tRNAs for rare codons (Rosetta™2 and BL21-CodonPlus-RP) were used and enabled expression of a 50 kDa recombinant protein. Despite this successful outcome, the expression of the recombinant Cops-1 protein remained at low level and limited stability. The recombinant His-tagged protein seems to be either degraded as soon as expressed or partially produced. The degradation persisted after increasing the amount of protease inhibitors and adopting more stringent handling procedures, suggesting that the degradation process is more likely to be intrinsic to the bacterial strain. Alternative expression systems such as yeast, insect or mammalian cells could be used for expression of a stable recombinant Cops-1 protein. Such systems are very useful for expression of eukaryotic proteins because they perform many posttranslational modifications such as translocation and glycosylation necessary for production of active proteins (Weidner *et al.,* 2010, Rothblatt and Meyer, 1986).

In order to investigate the biological function of Cops-1 protein, a monoclonal peptide antibody was produced. The highly reactive 9E1 hybridoma supernatant was used for immunolocalization studies. On purified *C. parvum* oocysts and sporozoites, the monoclonal antibody clearly recognized the contents of the oocyst when permeabilized and was able to stain free sporozoites, implying that

the protein is either a surface or secreted sporozoite protein. These characteristics are, along with the telomeric location, consistent with a role for Cgd2_4380 in host adaptation and virulence. On *C. hominis* oocysts and sporozoites, 9E1 showed a weaker internal staining of the oocysts. This level of staining persisted with several secondary antibody and also directly conjugated antibodies, suggesting that the staining is likely to be specific. This result is surprising because the Tyle-2 peptide (used to immunize mice) seems to be absent from the ChCops-1 sequence retrieved by sequencing. This could be explained by cross reactivity of the antibody with either a homologous region of the *C. hominis* ortholog or another *C. hominis* antigen. In addition, *C. hominis* oocyst suspensions were from clinical sources and showed high background staining by immunofluorescence which could partially explain the cross reactivity observed. Future studies would benefit from a polyclonal antibody directed against the whole recombinant protein for an improved determination of the localization of the Cops-1 protein and to confirm whether the gene is expressed in *C. hominis*.

The assessment of the role of Cops-1 protein in host-cell infection was undertaken *in vitro* using a blocking assay in coculture of *Cryptosporidium* oocysts and Caco-2 cells. The pre-incubation of the parasite suspension with 9E1 antibody did not influence the infectivity of *Cryptosporidium* when compared to controls as evidenced by comparable parasite counts. Although this result may suggest that Cops-1 protein does not have a role in host-parasite interaction, it could be that Tyle-2 epitope does not correspond to the region interacting with the intestinal cell receptors. The validation of this hypothesis would require the use of a polyclonal antibody targeting the whole recombinant Cgd2_4380 protein.

The immunogenicity of the Cops-1 protein was tested by Western Blot, treated with sera from *C. parvum* and *C. hominis* natural human infection. In recombinant preparations, Western Blot analysis with *C. parvum* serum showed a band of 50 kDa from Rosetta™2 lysates and to a lesser extent from BL21-

CodonPlus_RP lysates. This protein, however, was visible in both induced and non-induced lysate. Interestingly, when native *C. parvum* and *C. hominis* antigen preparations were tested by Western Blot using the same serum, a 50 kDa band was detected from *C. parvum* but not *C. hominis*. This result was obtained using a single *C. parvum* infected patient serum; testing of other *C. parvum* sera or serum pools would be valuable to confirm this observation. Nevertheless, this is a first report of a *C. parvum*-specific immunodominant protein. This protein could be Cops-1, but this could not be confirmed because of the poor reactivity of the 9E1 antibody in immunoblotting. A possible explanation for this poor reactivity could be the alteration or degradation of the linear epitope when tested by Western Blot. The likelihood of this 50 kDa *C. parvum* protein being Cops-1 are mainly based on size. There was only one report of a protein of similar size Cp47 (47 kDa) by Nesterenko and colleagues (1999). Cp47 is membrane-associated protein, which possesses significant binding affinity for the surface of both human and animal intestinal cells. The gene encoding Cp47 protein has not been discovered yet. Cp47 was identified by Western Blot using anti-*C. parvum* sporozoite membrane associated protein and is implicated in host-cell attachment (Nesterenko *et al.,* 1999). An anti-Cp47 antibody or a polyclonal anti-Cops-1 recombinant protein is required to determine the identity of the 50 kDa protein.

These preliminary data suggest that a *C. parvum* specific immunodominant protein of ~ 50 kDa is differentially expressed between *C. parvum* and *C. hominis* and could be the Cops-1 protein. This result, if confirmed, is the first evidence that Cops-1 is *C. parvum*-specific protein and would have several applications as diagnostic marker, species determinant and vaccine candidate.

5.5 Summary

One gene, Cops-1, emerged from PCR based screening as putatively specific for *C. parvum*. This gene was positioned telomerically and its product is annotated as secreted, serine rich, containing internal repeats and bioinformatic analysis suggests that the protein may be myristoylated at the N terminus, which itself is hydrophobic and capable of membrane insertion. The predicted protein may be N-glycosylated and under appropriate conditions could be phosphorylated heavily. In addition, it shows moderate sequence identity to cytoskeletal interacting proteins such as *shot* gene of *Drosophila melanogaster*.

The initial studies identified an orthologous gene in *C. hominis* (ChCops-1), but one which is considerably different in sequence and which is foreshortened lacking a repeat. This is suggestive of rapid evolution, therefore Cops-1 is a possible virulence factor candidate.

Cops-1 PCR product sequence analysis showed that the variability associated with this short fragment has a good discriminatory power and allowed clustering of *Cryptosporidium* genotypes and subtypes, which was consistent with the polymorphism seen in the other genetic loci tested by MLA. This result suggests that Cops-1 is also a good marker for taxonomic analysis.

A monoclonal peptide antibody 9E1 directed against a *C. parvum*-specific epitope clearly recognized the contents of the oocyst and stained free sporozoites, implying that the protein could be a surface or secreted sporozoite protein. These characteristics are, along with the telomeric location, consistent with a role for Cops-1 in host adaptation and virulence.

The protein encoded by CpCops-1 was expressed as a recombinant protein in *E. coli* strains having tRNAs for rare codons because of the codon bias of *Cryptosporidium* and a particular Cops-1 gene composition. The recombinant

protein was expressed at low levels and had limited stability. Optimization of the expression of the recombinant protein in other expression systems is required for the production of a full length stable recombinant protein, which will be useful for further characterization studies.

Native antigen preparations, from *C. parvum* and *C. hominis* oocysts, tested by immunoblot with *C. parvum* infected patient serum showed that a 50 kDa band was present in *C. parvum* but not *C. hominis* preparations. This is the first report of species-specific immunodominant protein for *Cryptosporidium,* which may be Cops-1 and would have several applications as diagnostic target, species determinant and vaccine candidate.

CHAPTER 6:

General discussion &

Future research

6.1 General discussion

This thesis has exploited genomic data from the relatively recent sequencing of *C. parvum* and *C. hominis* genomes in order to identify species-specific *Cryptosporidium* genes. The data presented demonstrating that the minimal genetic variability existing between *C. parvum* and *C. hominis* (<5%) is likely to include interesting genetic determinants with predicted features suggesting a role in host-pathogen interactions. Hitherto, no similar analyses mining genetic variation between novel gene sequences for these two *Cryptosporidium* species has been undertaken. This lack of interest in species-specific genes reflects the focus of *Cryptosporidium* researchers on further characterization of established genes of interest and genetic loci commonly used for *Cryptosporidium* typing. However, recognition of the potential of proteins, with no known homology to polypeptides with known function, in mediating host-parasite interactions and parasite-specific metabolic pathways has been recently acknowledged (Tomley, 2009).

6.1.1 Identification of species-specific genes by comparative genomics

Genome sequences of *C. parvum* and *C. hominis* were published in the same year (Abrahamsen *et al.,* 2004, Xu *et al.,* 2004). Genome sizes were comparable and only 3–5% sequence divergence was observed between the two species. The authors concluded that the phenotypic differences between *C. hominis* and *C. parvum* are caused by polymorphisms in coding regions and differences in gene regulation (Pain *et al.,* 2005, Xu *et al.,* 2004). In this study, attempts to identify genetic determinants that account for this minimal genetic variability were undertaken. Using Reciprocal BLAST, each gene was given a similarity score. Genes with <10% sequence identity were considered putatively species-specific. The highly variable genes were estimated at 93 for *C. hominis* and 211 genes for *C. parvum*. A subset of these genes was tested by PCR in a collection of *Cryptosporidium* clinical isolates and reference strains. Approximately 90% of the

genes tested were present in both *C. hominis and C. parvum*, indicating that the vast majority of the predicted genes seem to be common to both species. These results question either the accuracy of the *in silico* prediction or the quality of the published genome sequences. The latter explanation is most likely due to the incomplete status of the published *C. hominis* genome. Xu and colleagues (2004) reported over 240 gaps and no further progress of the genome sequence has been published to date. This partial completion impaired the accuracy of prediction of putative species-specific genes; especially for genes corresponding to sequence gaps. Therefore, the majority of the genes identified during this study (and shown in Appendix II) are not likely to be specific but are homologous genes that eluded the genome sequence due to incomplete coverage. An improved comparative genomic analysis is likely to be possible with fast progress being made towards the completion of the *C. muris* genome. The genomic sequences are being made accessible online prior to publication (Widmer *et al.,* 2007). At the time of writing this chapter, 8.9 Mb from the *C. muris* genome are available for download from CryptoDB, of which 7.2 Mb correspond to coding sequences. 58.4% of the subset of genes tested by PCR appear to have orthologs in *C. muris*. This result further supports that the majority of the genes tested are common to *Cryptosporidium* species more generally. Comparative genomics of *C. parvum*, *C. hominis* and *C. muris* genomes is expected to improve the understanding of *Cryptosporidium* species evolution and host adaptation (Widmer *et al.,* 2007). A similar reciprocal BLAST analysis, performed when the data are complete, will add considerable power to the analysis, enabling further refining of the process of discovery for genetic determinants of *Cryptosporidium* host tropism.

6.1.2 Sequence analysis of novel genetic loci and phylogenetic applications

The ten genetic loci identified as described above were common to both *Cryptosporidium* species. Nevertheless, they are novel genetic determinants never before investigated. Sequence analysis of these genes revealed interesting genetic polymorphism: 78 SNPs were detected, 78.3% (61) of which were

species-specific. The presence of species-specific SNPs was reported previously for several genetic markers and have been exploited for *Cryptosporidium* genotyping and subtyping (Sulaiman *et al.,* 1999). Several of the newly identified SNPs were confirmed by PCR-RFLP and showed a similar discriminatory power to the extensively used markers. Interestingly, this level of SNP frequency per gene is higher than in the routinely used *Cryptosporidium* loci. The majority (64.2%) of the SNPs were synonymous, suggesting a minimal effect on the protein function. However, it has been documented that synonymous mutations can also alter the structure, function and expression level of the protein by affecting messenger RNA splicing, stability, protein folding and structure (Hunt *et al.,* 2009). Thus the actual effect of these SNPs on gene function can only be assessed experimentally.

Sequence analysis of these novel genetic loci allowed the construction of a robust and novel MLA. The Neighbor-Joining phylogenetic tree constructed clearly grouped and discriminated with high bootstrap values the previously described lineage of anthroponotic *C. parvum* from *C. parvum* isolates and the rabbit genotype from *C. hominis* which has been recently proposed as a new *C. hominis* subspecies on the basis of this and other data (Chalmers *et al.,* 2009 b). The discriminatory power of these loci is powerful because not only does it allow discrimination of *Cryptosporidium* species/genotypes but also subtypes. This level of discrimination (for anthroponotic *C. parvum* and rabbit genotype) was only previously achieved using GP60 subtyping.

There is a debate in the literature about the advantages and limitations of GP60 subtyping of *Cryptosporidium* isolates. GP60 is undeniably the most commonly used population marker for *Cryptosporidium* spp (Jex and Gasser, 2010). This locus is hypervariable and isolates are classified in subtypes according to the number of repeats in a nomenclature proposed by Sulaiman and colleagues (2005). Subsequent studies showed that the level of discrimination is not in accordance with multi-locus subtyping and GP60 genotype by itself is difficult to reconcile with the concept of a subtype defined as a genetically distinct

population within a species (Widmer, 2009). Thus, Widmer has stated that GP60 may not be a reliable marker of *C. parvum* and *C. hominis* population structure (Widmer, 2009). In addition, Hunter and colleagues (2008) pointed out that it is still unclear if the identification of different subtypes in outbreak settings represented different lineages or evolution of strains during the outbreaks. Another limitation of GP60 is that the high level of variability is under selective pressure as this locus was shown to be involved in sporozoite adhesion to host cells (Okhuysen and Chappell, 2002). In addition, Jex and Gasser (2010) reported that the genetic richness of *Cryptosporidium* isolates was associated with a surprisingly low diversity. In fact, despite the report of high number of subtypes (richness), the relative prevalence of each distinct type (diversity) was low. The authors reported that human cryptosporidiosis is mainly associated with six dominant GP60 subtypes.

In this study, comparative genomics were used to identify novel genetic loci. These loci were common to *C. hominis* and *C. parvum* and some were also present in *C. meleagridis*. Sequence analysis showed interesting genetic polymorphism and excellent phylogenetic potential. One limitation of this approach is the high number of loci used to achieve high discriminatory power, which would limit its use for routine subtyping. However, a reduction of the number of loci while keeping the discriminatory power can be achieved by selecting a subset of highly variable genes, in which polymorphisms reflect the ecological adaptation of divergent subpopulations. The usefulness of these novel genetic loci for phylogenetic analysis needs to be validated by testing higher numbers of subtypes and isolates from different geographical locations. Moreover, further PCR testing of the remaining loci identified by comparative genomics will likely yield additional genes encoding proteins with roles in host-parasite interaction in general and host range determinants and virulence factors in particular.

In the current situation, typing of *Cryptosporidium* isolates is highly desirable because it provides insights about population structure and virulence of

particular subtypes and it allows source tracking for epidemiological investigation of outbreaks (Chalmers *et al.,* 2008, Hunter *et al.,* 2007). However, the identification of an adequate locus (or loci in the case of multi-locus typing) is dependent on the satisfaction of several evaluation criteria that still have to be delineated in the case of *Cryptosporidium*. To address the issues of *Cryptosporidium* typing, Medini and colleagues (2008) suggested that new typing systems need to incorporate whole genome sequences. In addition, Jex and Gasser (2010) recommended that future studies should use next generation sequencing to conduct comparative genome sequence surveys to test the validity of current genetic classifications based on gp60 data, which would overcome the limitations of classification based solely on the analysis of a small number of genetic loci.

A less ambitious but probably more realistic approach is the hypothesis that the identification of new genetic markers can be drawn, using comparative genomic techniques, from the vast amount of genetic data generated by the three *Cryptosporidium* genome projects. Data presented in this thesis (Chapter 3) clearly showed the feasibility of this approach; in fact, this is the first report of genetic loci identified using a comparative genomics approach, which possess genotyping and subtyping discriminatory power.

6.1.3 Whole genome amplification for generating practically unlimited quantities of isolate specific DNA

A major barrier to the extensive testing of *Cryptosporidium* isolate DNA by typing and sequence analysis is the limited amount of material available (Smith *et al.,* 2006). This is particularly relevant for unculturable or fastidious microorganisms such as *Cryptosporidium*. Whole genome amplification (WGA) can be used to increase the amount of nucleic acid materials of waterborne pathogens isolated from clinical and environmental samples as reviewed by Bouzid and colleagues (2008). The study described in this thesis (Chapter 4) trialling three commercial WGA kits, showed that one MDA based kit: illustra GenomiPhi WGA kit had the

best performance amplifying *Cryptosporidium* DNA with 90.9% success rate, generating high concentration of high molecular weight DNA with 100% fidelity. Real-time PCR result showed over 30-fold increase specifically of *Cryptosporidium* DNA. The amplified DNA was also shown to be suitable for downstream genotyping and subtyping applications. However, a validation of the integrity and fidelity of amplification of highly polymorphic genetic loci (such as GP60) should be carried out to fully validate the usefulness of the amplified DNA for downstream applications. In addition, use of WGA amplified DNA as PCR template could increase PCR sensitivity from clinical samples. Similar findings were reported for the detection of *Trypanosoma* species from blood samples (Pinchbeck *et al.,* 2008). Despite these advantages, the use of WGA is still limited in diagnosis and research laboratories and the WGA related publication is scattered. Several investigators agree about the usefulness of the technique but implementing it for routine and extensive use may require further demonstration of its utility.

6.1.4 *Cryptosporidium* species-specific genes

Although the genomes of *C. hominis* and *C. parvum* are 97% identical, their host range is strikingly different. *C. parvum* is the zoonotic species, while *C. hominis* normally only infects humans. The accepted consensus is that the two genomes are homologous and have essentially the same gene content (Xu *et al.,* 2004, Abrahamsen *et al.,* 2004). It has been postulated that the variable regions correspond primarily to the presence of micro or minisatellites (Tanriverdi and Widmer, 2006) and that the phenotypic differences between *C. hominis* and *C. parvum* are caused by polymorphisms in coding regions and differences in gene regulation (Pain *et al.,* 2005, Xu *et al.,* 2004). However, this minimal genetic divergence may include hitherto undetected genetic determinants of host specificity and virulence. Data presented in this thesis (Chapter 3) showed that despite the majority of the predicted specific genes being common to both *Cryptosporidium* species, evidence was found that one *C. parvum* gene (Cops-1) and one *C. hominis* gene (Chos-1) generated appropriate PCR product from only

one species. This result suggests that within the predicted genes, a small proportion may be species-specific. Those initial results suggest that these two loci could be used for diagnosis purposes to discriminate between *C. hominis* and *C. parvum* in a robust duplex PCR assay which reflects actual gene repertoire and is not simply dependent on maintenance of few single nucleotide polymorphisms. Sensitivity and specificity assays should be undertaken to validate the diagnostic potential of this test.

6.1.5 *C. parvum* specific gene (Cops-1)

The potential of Cops-1 as a species determinant and a virulence factor was reinforced by the predicted features of the gene and its encoded protein. Cops-1 is positioned telomerically and annotated as a secreted, serine rich protein, containing internal repeats. The predicted protein may be N-glycosylated and under appropriate conditions could be phosphorylated heavily. It shows moderate sequence identity to cytoskeletal interacting proteins such as shot gene of *Drosophila melanogaster*. The telomeric location of the gene indicates that the gene is prone to higher recombination rates and it is likely to be a contingency gene. Such genes were shown to be involved in host-pathogen interaction and parasite survival in the host, in *T. brucei* it includes the genes of the subtelomerically positioned expression site (ESAGS) and notably the VSG gene, which undergoes antigenic variation to evade the host immune responses and allow parasite survival (Yang *et al.,* 2009, Barry *et al.,* 2003) and similarly, the var genes of *Plasmodium falciparum* coding for erythrocyte protein-1, which likewise are the subjects of antigenic variation (Kyes *et al.,* 2001). Genetic determinants of host tropism are also likely to be among these contingency genes and such a role has been proposed for both the transferrin receptor encoded as a heterodimer by ESAG6 and ESAG7 and the serum resistance antigen (SRA) which confers resistance to human sera for *T. brucei rhodesiense*.

The predicted Cops-1 protein has at its N terminus a signal peptide, a transmembrane domain and a myristoylation motif implying that Cops-1 encodes

a secreted protein involved in host pathogen interaction. Several *Cryptosporidium* surface and apical complex proteins are predicted to have a signal peptide (Gp900, gp40, Cpa135, Cp2, TRAP-C1) (Wanyiri and Ward, 2006). These proteins are potential virulence factors and have been shown to mediate host-cell attachment and invasion. One of these proteins, gp900 has a signal peptide and a transmembrane domain and was shown to mediate invasion of host cells (Barnes *et al.,* 1998). The myristoylation motif mediates the covalent attachment of myristic acid to the N terminal glycine of eukaryotic and viral proteins and this process is catalyzed by an enzyme N-myristoyl transferase (NMT) (Farazi *et al.,* 2001, Poli *et al.*, 1991, Towler *et al.* 1988). Myristoylated proteins aid subcellular targeting, protein-protein interaction and are involved in signal transduction cascades. NMTs were characterized in several protozoan parasites. In *Leishmania* major and *Trypanosoma brucei*, NMT is a 48.5 kDa protein that localizes to both membrane and cytoplasmic fractions and is expressed in all life stages and was shown to be essential for viability (Price *et al.,* 2003).

Cops-1 has a potential as a diagnostic target amplifying a 655 bp band from *C. parvum* isolates. In addition, a 200 bp band was present in PCR products from *C. hominis* and *C. parvum* DNA. This band could serve as a useful internal control for the presence of human infective *Cryptosporidium*. This assay as it stands types to the species level. Sequencing of the 200 bp band showed that Cops-1 seems to have an ortholog in *C. hominis*. PCR product sequences were used to build a Neighbor-Joining Tree, which showed good discrimination of *C. parvum*, *C. parvum* anthroponotic subtype, *C. hominis* and the rabbit genotype. The phylogenetic tree showed that the variability associated with this short fragment has a good discriminatory power and allowed for subtyping of *Cryptosporidium* isolates in a manner consistent with the multi-locus analysis from Chapter 3.

The identification of the full ChCops-1 gene sequence was undertaken using a primer walking approach. Not all primer combinations allowed DNA amplification from *C. hominis* isolates. Surprisingly, some primer combinations worked with

Ch3 isolate only. Thus, the determination of the full gene sequence was possible only from Ch3. ChCops-1 seems to be truncated to 1263 bp in this isolate. The retrieved sequence showed 78.8% sequence identity to the published Cgd2_4380 gene. Interestingly, TU502 DNA did not generate PCR products using the majority of the primers tested. This result could possibly explain why Cops-1 gene had eluded the genome sequence. A set of primers based on the predicted ChCops-1 gene sequence failed to generate DNA amplification from any isolate. Therefore, full length ChCops-1 sequence requires further investigation and validation. However, the 5' end of the gene is conserved as shown by several primer combinations. Therefore, CpCops-1 and ChCops-1 have a conserved N terminus, suggesting that some of the Cops-1 characteristic features are likely maintained. These initial studies have identified an orthologous gene in *C. hominis* (ChCops-1), but one which is considerably different in sequence and which is foreshortened lacking a repeat. These findings support the rapid evolution of this gene and the candidature for the protein encoded by Cgd2_4380 as a virulence factor or host determinant.

6.1.6 Cloning and expression of CpCops-1 protein

The full length CpCops-1 gene from *C. parvum* Iowa DNA was successfully cloned into cloning and expression vectors. Despite maintenance of the plasmids in the bacterial hosts, no expression of the recombinant protein was detected by SDS-PAGE or Western Blot. This might be due to differences in codon usage between the eukaryotic machinery of *Cryptosporidium* and the prokaryotic system of *E. coli*. Codon bias has been reported by Fayer (1997) to potentially limit the performance of *E. coli* as *Cryptosporidium* DNA sequences may not be translated at detectable levels because the tRNAs necessary for efficient expression are present at too low concentration. As a consequence of the codon bias, a solution was sought in using bacterial strains having tRNAs for rare codons: Rosetta™2 and BL21-CodonPlus-RP. The full-length recombinant His-tagged protein was expressed, but only at low levels and this has hindered further purification and characterization of the recombinant protein. Alternative expression systems such

as yeast, insect or mammalian cells should be assessed for the expression of stable recombinant protein, especially because they allow posttranslational modifications (Weidner *et al.,* 2010, Rothblatt and Meyer, 1986). In addition, gene synthesis using commercially available services (such as Genscript) claiming to produce high amounts of purified recombinant protein are also an option and could be investigated in the future.

6.1.7 9E1 a monoclonal peptide antibody anti-CpCops-1

A monoclonal peptide antibody directed against Cops-1 protein was produced and used to determine the localization of the protein on *Cryptosporidium* oocysts and sporozoites. The monoclonal antibody clearly recognized the contents of *C. parvum* oocysts and was able to stain free sporozoites, implying that the protein is either a surface or secreted sporozoite protein. On *C. hominis* oocysts and sporozoites, 9E1 showed a much weaker internal staining of the oocysts. The cross reactivity of the monoclonal antibody is surprising because the Tyle-2 peptide (used to immunize mice) seems to be absent from *C. hominis* Cops-1 sequence retrieved by sequencing. This could be explained by cross reactivity of the antibody with either a homologous region of the *C. hominis* ortholog or another *C. hominis* antigen entirely. In addition, *C. hominis* oocyst suspensions were from clinical sources and not purified, they showed high background staining by immunofluorescence which could partially explain the cross reactivity observed. Future studies would benefit from a polyclonal antibody directed against the whole recombinant protein for an improved determination of the localization of the Cops-1 protein and to confirm whether the gene is expressed in *C. hominis*.

Cops-1 predicted characteristics, its telomeric location and the sporozoite staining pattern are all consistent with a role of the protein in host-parasite interaction. Several other sporozoite proteins involved in host-cell attachment and invasion have been shown to be deposited in trails during gliding motility (Wanyiri and Ward, 2006). 9E1 staining of glass slides, on which *Cryptosporidium*

sporozoites were left to glide after excystation, did not show staining of the trails, unlike the slides stained with 4C1 control antibody clearly showing fluorescent stain of the gliding trails as previously reported (Feng *et al.,* 2006). This preliminary result suggests that Cops-1 protein is unlikely to be shed in trails. However, future confirmation would rely on colocalization studies of 4C1 with 9E1 and additional Cops-1 specific antibody.

An attempt to investigate the potential role of Cops-1 protein in host-cell attachment and invasion was undertaken *in vitro* using a blocking assay with the 9E1 monoclonal antibody in coculture with Caco-2 cell monolayers. The pre-incubation of the parasite suspension with 9E1 antibody did not influence the infectivity of *Cryptosporidium* when compared to controls as evidenced by comparable parasite counts. Although this result may suggest that Cgd2_4380 protein does not have a role in host-parasite interaction, it could be that Tyle-2 epitope does not correspond to the region interacting with the intestinal cell receptors. 9E1 seems to be a non blocking and non neutralizing antibody. *Cryptosporidium* antibodies against putative virulence factors described in the literature which showed blocking of infection *in vitro* and *in vivo* were directed against whole recombinant proteins (Boulter-Bitzer *et al.,* 2007). The production of a polyclonal antibody raised to the whole recombinant Cops-1 protein may better demonstrate if Cops-1 protein is involved in *Cryptosporidium* pathogenesis and host-cell interaction.

6.1.8 Immunogenicity of Cops-1 protein

Immunogenicity of Cops-1 protein was investigated using both native and recombinant antigen preparations revealed by 9E1 and *Cryptosporidium* natural infection sera. Interestingly, sera from *C. parvum* infected patients recognized a 50 kDa protein from native *C. parvum* antigen preparations but not *C. hominis*. This is the first report of a *C. parvum* specific immunodominant protein. This protein may be Cops-1, however, this could not be confirmed because of the poor reactivity of 9E1 monoclonal antibody in immunoblotting. A likely

explanation for this poor reactivity is the alteration or degradation of the linear epitope tested by Western Blot, especially given that 9E1 was reactive and stained oocyst content by IFA.

6.2 Summary and Future work

Despite recent advances in genomics, proteomics, biology and pathogenesis of *Cryptosporidium*, important questions remain unresolved. These include methods to discriminate rapidly, reliably and inexpensively between morphologically undistinguishable human infective and non infective oocysts, especially those found during filtration of drinking water supplies. Similarly, the search for loci with the ability to adequately type and subtype isolates in order to inform risk assessment is far from over. Finally, our understanding of host range and virulence of *Cryptosporidium* species is still in its infancy for this under researched but important pathogen.

This thesis has exploited genomic data from *C. parvum* and *C. hominis* genome sequences to identify genetic determinants associated with host tropism. The majority of the loci identified by *in silico* screen were shown experimentally to be common to both species. This illustrates the pitfalls of some forms of comparative analyses in "finished" eukaryotic genomes, which are not actually fully completed and assembled. Unfortunately, when looking for contingency genes, repetitive regions such as the telomeres are often amongst the most difficult regions of the genome for complete assembly and it is likely that important pathogenicity factors may continue to lurk undiscovered in such areas. Nevertheless, sequence analysis of the genes highlighted by *in silico* screening showed interesting genetic polymorphism and excellent phylogenetic potential.

In these studies, a *C. parvum* gene, Cops-1, was identified *in silico* and confirmed experimentally to be species-specific. A fragment from the 5' end of the gene, which appears to be more conserved, amplified from all *C. hominis* isolates and a

truncated ortholog was identified in at least one isolate of *C. hominis*. Phylogenetic analysis of sequence from the fragments generated permitted subtyping of both *C. hominis* and *C. parvum* with a high degree of confidence in a manner consistent with multi-locus and microsatellite analyses and in a way that the hypervariable GP60 was not able to do. A direct comparison of divergence in Cops-1 sequence between the GP60 subtypes is a priority in the evaluation of Cops-1 as a candidate marker for subtyping. It is worth considering that single locus typing can bias results in organisms like *Cryptosporidium*, where sexual exchange occurs and this may limit the use of Cops-1 as a single marker for typing. In the future, whole genome sequencing from key subtypes will better define the genetic diversity of *Cryptosporidium* and enable better evaluation of the efficacy of Cops-1 as a subtyping marker.

Cops-1 PCR based assay can be used for *Cryptosporidium* genotyping with no requirement for subsequent RFLP as with the routinely used targets. Subtyping is also possible by RFLP, sequencing and real-time PCR analysis to exploit the SNPs described. In addition, Cops-1 could be used as a target for loop-mediated isothermal amplification procedure (LAMP) of particular use for field studies and in modestly equipped laboratories.

Cops-1 protein predicted features are highly suggestive of a potential role in virulence and host-parasite interaction. Unfortunately, the preliminary characterization studies faced several limitations. First, the full gene length was determined in only one *C. hominis* clinical isolate and was not subsequently amplifiable by PCR from this or other *C. hominis* isolates. Second, the expression of the recombinant protein was of low level and limited stability and finally, no evidence of Cops-1 being deposited in trails or mediating host-cell invasion were established. Nevertheless, an immunodominant 50 kDa protein was detected from *C. parvum* native antigen preparations and not *C. hominis* consistent with the idea that Cops-1, differentially expressed in *C. parvum* and *C. hominis,* may be discriminated serologically, if this is indeed the case. The recombinant Cops-1 protein has the potential for use in serodiagnostic assays and to form the basis of

seroepidemiological studies based on comparison between groups with previous exposure to *C. parvum* compared with *C. hominis* as determined by ELISA and Western Blot.

The first monoclonal antibody raised to a linear epitope of Cops-1 interacted well with the peptide to which it was raised and gave good immunofluorescence from purified *C. parvum* parasites. Unfortunately the antibody performed poorly on immunoblot against recombinant protein and appeared to give high background fluorescence from components of faecal samples suggesting that although it may be able to discriminate species in a single oocyst, it may not be an ideal diagnostic tool in its own right. Production of better antibody with higher specificity perhaps raised to whole recombinant proteins or other epitopes raises the prospect of a reagent which can be used to directly discriminate species in clinical and environmental isolates.

The data presented in this thesis has laid foundation for future work, which includes applied and basic scientific research. Comparative genomics have been shown to be effective in discovering genetic determinants involved in virulence and host pathogen interaction. Genome comparison of *C. parvum*, *C. hominis* and *C. muris* is likely to improve the understanding of the biology, pathogenesis and evolution of *Cryptosporidium* species and is likely to shed light on the processes that had mediated host tropism and niche adaptation. Extension of comparative genomics approaches can also be usefully utilized to identify orthologous genes known to have roles in host pathogen interactions and which could be conserved functionally amongst apicomplexan parasites or orthologous genes in closely related species with highly divergent domains, which may be interesting biologically and useful for subtyping.

Data presented in this thesis showed evidence of one *C. hominis* specific gene (Chos-1). Both Cops-1 and Chos-1 are predicted to have a signal peptide, thus they are likely to be secreted and might be involved in host-parasite interactions. Chos-1 characterization should be undertaken in the future both on the genetic

and molecular level. Interestingly, neither the Cops-1 nor Chos-1 gene has appeared in the published secretome of *C. parvum* although there is evidence for the transcription of these genes in the form of ESTs. In the future, platform technologies such as proteomics and transcriptomics may give additional insight into the likely function and partner proteins of these molecules.

For Cops-1 though, ensuring full characterization is dependent on relating DNA sequence to structure and ultimately to function. Functional biology assays *in vitro* and *in vivo* should be performed to confirm the importance of this interesting protein in *Cryptosporidium* biology and pathogenesis. In addition, experimental evaluation of Cops-1 predicted features should be carried out. Myristoylation assays can be undertaken using radio-labelled myristic acid *in vitro* and *in vivo*. Glycosylated proteins can be purified using a pull down assay and the effect of deglycosylation (glycosidase) or glycosylation inhibitors can be assessed by western blot and *in vitro*. Trans-expression of GFP into *Cryptosporidium* was reported using a unique viral transfection system and could be used for trans-expression of tagged full length Cops-1. Perhaps more straightforwardly, the protein could be trans-expressed in *Toxoplasma gondii* to consider its trafficking pathway and further functional assays of host-parasite interaction. With the recombinant Cops-1 protein, it would be possible to evaluate its role in host-cell attachment, eliciting host-cell signalling and invasion by evaluation of coated bead attachment and uptake *in vitro*. Currently, genetic manipulations mainly “knock out” or “knock down” are not feasible in *Cryptosporidium*, but these trans-expression and coated bead systems would enable evaluation of the effect of targeted mutations on the protein function, once these assays are established. Such mutation could primarily target the predicted features, for instance myristoylation and glycosylation sites. Identification of proteins interacting with Cops-1 is likely to assist uncovering its function and could be investigated using yeast two-hybrid assay or co-precipitation. Finally, the evaluation of the protective potential of Cops-1 can be assessed *in vivo* by immunization of mice using the recombinant protein and comparing the severity of symptoms and oocyst shedding after *Cryptosporidium*

infection to a non immunized control group. Furthermore, the immune response could be targeted by oral dose with viral or bacterial carriers to elicit a mucosal immune response in the gut. A protective vaccine against *C. parvum* is highly desirable for livestock. The evaluation of Cops-1 as a vaccine candidate using this approach would likely be *C. parvum* specific and may or may not be effective in humans. Nevertheless, animal vaccines are commercially viable and could dramatically reduce the number of human outbreaks derived from zoonotic transmission.

Appendix I

Appendix I: Part of the work in this thesis has been published or presented at these conferences

Papers

1- **Bouzid M**, Heavens D, Elwin K, Chalmers RM, Hadfield SJ, Hunter PR, Tyler KM. (2010). Whole genome amplification (WGA) for archiving and genotyping of clinical isolates of *Cryptosporidium* species. Parasitology **137**(1):27-36.
2- **Bouzid M**, Steverding D, Tyler KM. Detection and surveillance of waterborne protozoan parasites. (2008). Curr Opin Biotechnol. **19**(3):302-6.
3- **Bouzid M**, Tyler K.M, Christen R, Chalmers R. M, Elwin K. and Hunter P. R. Multi-locus analysis of human infective *Cryptosporidium* species and subtypes using ten novel genetic loci. Submitted to BMC Microbiol. May 2010.

Abstracts

1- American Society for Microbiology General Meeting. May 2010. San Diego, USA. Investigation of putative species specific genes for anthroponotic *Cryptosporidium* species identified by comparative genomics approach. **Maha Bouzid**, Louise Crawley, Audrey Dubourg, Rachel M. Chalmers, Paul R. Hunter, Kevin M. Tyler. ***Poster presentation*.**
2- British Society for Parasitology meeting. March 2010. Cardiff UK. Investigation of putative species specific markers of *C. hominis* and *C. parvum* identified by comparative genomics tools. **Maha Bouzid**, Rachel Chalmers, Paul R. Hunter, Kevin Tyler. ***Oral presentation*.**
3- III International *Giardia* and *Cryptosporidium* conference. October 2009. Orvieto, Italy. Investigation of putative species-specific markers of *C. hominis* and *C. parvum* identified by comparative genomics tools. **Maha Bouzid**, Kirstin Elwin, Steve Hadfield, Darren Heavens, Richard Christen, Rachel Chalmers, Paul R. Hunter, Kevin Tyler. ***Oral presentation*.**
4- III international *Giardia* and *Cryptosporidium* conference. October 2009. Orvieto, Italy. Comparative biology of *Cryptosporidium hominis* and the *Cryptosporidium* rabbit genotype. Rachel Chalmers, Guy Robinson, Steve Wright, Paul Hunter, Kristin Elwin, **Maha Bouzid**, Lee Innes, Steve Hadfield, Kevin Tyler, Frank Katzer. ***Poster presentation*.**
5- 15th Health Related Water Microbiology Symposium. May 2009. Naxos, Greece. Investigation of the genetic variability of anthroponotic *Cryptosporidium* species. **Maha Bouzid**, Rachel Chalmers, Paul Hunter, Kevin Tyler. ***Poster presentation*.**

6- British Society for Protist Biology meeting. April 2009. Norwich, UK. Cops-1: a species determinant for anthroponotic *Cryptosporidium species*? **Maha Bouzid**, Rachel Chalmers, Paul Hunter, Kevin Tyler. ***Oral presentation*.**

7- British Society for Parasitology spring meeting. March 2008. Newcastle, UK. Investigation of the genetic divergence between clinical isolates of anthroponotic species of *Cryptosporidium*: *C. hominis* and *C. parvum*. **Maha Bouzid**, Rachel Chalmers, Paul Hunter, Kevin Tyler. ***Poster presentation*.**

Appendix II

Appendix II: Results of the reciprocal BLAST applying a 10% sequence identity threshold. All coding sequences from *C. parvum* were blasted against *C. hominis* genomic data. For each gene, the best target in *C. hominis* was indicated. The number of identical nucleotides and the percentage of sequence coverage were presented. Similarly, every coding sequence of *C. hominis* was blasted to the genomic sequences of *C. parvum*. 272 and 117 putative species-specific genes were identified for *C. parvum* and *C. hominis*, respectively.

Accession Number	*C. parvum* gene name	Gene annotation	Best *C. hominis* target	% identity	Query length (bp)	Number of identical nucleotides	Sequence coverage (%)
AAEE01000002	cgd6_5320	hypothetical protein with similarity to many plasmodium proteins	Chro.60073	0	2811	19	0
AAEE01000013	cgd2_80	ABC transporter family protein, 2x AAA domain	Chro.80487	0	4431	35	1
AAEE01000007	cgd5_330	hypothetical protein	Chro.70096	0	3504	23	0
AAEE01000007	cgd5_1000	hypothetical protein	Chro.60190	0	2112	20	0
AAEE01000010	cgd5_4500	possible vacuolar protein sorting associated protein (VPS)	Chro.10424	1	1200	19	1
AAEE01000002	cgd6_660	secreted pepsinogen like aspartyl protease having a signal peptide	Chro.60138	1	1902	34	1
AAEE01000008	cgd4_1740	possible HSMGG motif (esterase?)	Chro.80330	1	3078	51	3
AAEE01000009	cgd4_1330	NADPH-dependent FMN FAD containing oxidoreductase , transcripts identified by EST	Chro.60070	1	2460	39	2
AAEE01000009	cgd4_290	putative protein phosphatase 2A regulatory B subunit, highly conserved but no plasmodium hits	Chro.80116	1	2076	33	2
AAEE01000004	cgd3_500	putative fucose translocator with 8 transmembrane domains, within locus of 3 paralogous genes	Chro.40126	1	1185	17	1
AAEE01000004	cgd3_2870	similarity of possible bacterial origin within discrete region similar to long-chain fatty acid acyl-CoA synthetases	Chro.00004	1	2334	37	2
AAEE01000002	cgd6_440	PP2Cc like protein phosphatase	Chro.80459	1	1602	22	1
AAEE01000011	cgd8_680	large low complexity protein with repeats	Chro.30010	1	3441	68	4
AAEE01000006	cgd1_110	predicted secreted protein, signal peptide, paralogs	Chro.10158	1	1221	20	1
AAEE01000002	cgd6_540	Ser/Thr protein kinase	Chro.10402	1	2841	54	3
AAEE01000008	cgd4_1570	hypothetical protein, possible conserved	Chro.50401	1	1224	21	1
AAEE01000004	cgd3_4240	insulinase like peptidase	Chro.50235	1	3342	63	4
AAEE01000004	cgd3_2640	hypothetical protein with conserved domain	Chro.40078	1	1311	19	1
AAEE01000010	cgd5_4190	hypothetical protein	Chro.20329	1	2247	38	1
AAEE01000009	cgd4_1130	PhnP like hydrolase of the metallobetalactamase fold	Chro.10012	1	1065	19	1
AAEE01000005	cgd2_4380	signal peptide, repeats, gene anchored to telomere	Chro.10099	1	1434	19	1

AAEE01000010	cgd5_4000	hypothetical protein	Chro.40042	1	2034	40	2
AAEE01000004	cgd3_510	putative fucose translocatorw ith 8 transmembrane domains, within locus of 3 paralogous genes	Chro.30381	1	1170	17	1
AAEE01000014	cgd6_320	hypothetical protein with signal peptide and 8 transmembrane domains, amino terminl region conserved in plasmodium, transcripts identified by EST	Chro.80433	1	2427	38	2
AAEE01000013	cgd2_10	hypothetical protein	Chro.50504	1	1338	20	1
AAEE01000003	cgd8_2370	adenosine kinase like ribokinase	Chro.40163	1	1188	21	1
AAEE01000002	cgd6_670	extacellular protein with a signal peptide, clostripain like caspase/hemoglobinase domain, notch domain and 2 EGF domains	Chro.70432	1	4824	54	2
AAEE01000006	cgd1_360	similar to Noc2p N-terminus	Chro.60406	1	2511	28	1
AAEE01000008	cgd4_1780	ring domain protein	Chro.50317	1	1434	21	1
AAEE01000007	cgd5_540	hypothetical protein	Chro.50091	1	1524	24	1
AAEE01000002	cgd6_720	unknown protein	Chro.50392	1	3267	51	3
AAEE01000012	cgd4_4100	hypothetical protein	Chro.60053	1	939	18	1
AAEE01000007	cgd5_2970	hypothetical protein	Chro.20342	1	1869	37	3
AAEE01000004	cgd3_3670	putative protein kinase CK2 regulatory subunit CK2B1	Chro.70481	2	759	19	2
AAEE01000002	cgd6_5120	protein with two PHD Zn fingers that is probably involved in chromatin function	Chro.40494	2	687	20	2
AAEE01000012	cgd4_3990	hypothetical protein with possible transmembrane domain near N-terminus	Chro.70342	2	1743	42	3
AAEE01000002	cgd6_1130	hypothetical protein with signal peptide	Chro.40330	2	1827	38	3
AAEE01000016	cgd6_5510	telomeric insulinase-like protease with signal peptide	Chro.40328	2	1473	33	3
AAEE01000005	cgd2_3800	hypothetical protein	Chro.60478	2	1110	33	4
AAEE01000004	cgd3_630	hypothetical protein with signal peptide	Chro.80019	2	948	24	2
AAEE01000010	cgd5_3770	putative arginine N-methyltransferase	Chro.80547	2	1998	40	2
AAEE01000018	cgd5_4510	hypothetical protein	Chro.40216	2	789	20	2
AAEE01000007	cgd5_320	carboxylesterase , lysophospholipase, signal peptide	Chro.10070	2	1422	33	3
AAEE01000005	cgd2_2430	ximpact ortholog conserved protein seen in bacteria and eukaryotes	Chro.10042	2	873	21	2
AAEE01000012	cgd4_4090	putative Sec14d	Chro.50438	2	3192	74	2
AAEE01000013	cgd2_100	hypothetical protein	Chro.30403	2	723	18	2
AAEE01000004	cgd3_1090	Rrp9p/U3-55K-family snoRNP-associated protein with several WD40 repeats	Chro.50400	2	1374	30	2
AAEE01000003	cgd8_3150	hypothetical conserved protein, transcripts identified by EST	Chro.30367	2	582	17	2
AAEE01000009	cgd4_1510	similar to undecaprenyl pyrophosphate synthetase	Chro.80332	2	759	21	2
AAEE01000010	cgd5_4490	hypothetical protein, low complexity, limited similarity to gi	Chro.60107	2	8076	213	12

AAEE01000012	cgd4_4160	possible carboxypeptidase	Chro.20426	2	1371	39	2
AAEE01000006	cgd1_960	RING finger containing protein	Chro.20379	2	1458	31	2
AAEE01000007	cgd5_3290	transcription elongation factor TFIIS	Chro.60122	2	702	19	2
AAEE01000005	cgd2_3970	RRM domain containing protein	Chro.40100	2	597	17	2
AAEE01000011	cgd8_620	hypothetical protein	Chro.20223	2	783	19	2
AAEE01000002	cgd6_2530	hypothetical protein	Chro.70569	2	858	24	2
AAEE01000002	cgd6_1810	hypothetical protein	Chro.10285	2	675	18	2
AAEE01000004	cgd3_680	cathepsin like thiol protease possibly membrane associated, putative	Chro.70394	2	897	20	2
AAEE01000004	cgd3_2720	putative topoisomerase VIA	Chro.20351	2	708	18	2
AAEE01000001	cgd7_4890	hypothetical protein	Chro.30428	2	1200	25	2
AAEE01000003	cgd8_2990	hypothetical protein	Chro.50389	2	543	16	2
AAEE01000005	cgd2_1820	hypothetical protein	Chro.10109	2	603	17	2
AAEE01000003	cgd8_4700	hypothetical protein	Chro.60225	2	654	17	2
AAEE01000016	cgd6_5520	peptidase'insulinase like peptidase'	Chro.40162	2	1713	48	4
AAEE01000009	cgd4_100	hypothetical protein	Chro.50507	2	681	18	2
AAEE01000002	cgd6_5020	protein with WD40 repeats	Chro.50196	3	651	20	3
AAEE01000007	cgd5_3250	secreted lipopolysaccharide sugar transferase like family 8 glycosyltransferase	Chro.10377	3	477	18	3
AAEE01000007	cgd5_190	plasmodium conserved protein	Chro.20379	3	1041	37	5
AAEE01000012	cgd4_3920	DinB/family X-type DNA polymerase	Chro.30486	3	1077	37	5
AAEE01000009	cgd4_90	hypothetical protein, similarity to Rrp7p	Chro.70603	3	531	18	3
AAEE01000006	cgd1_3280	hypothetical protein	Chro.40376	3	939	36	6
AAEE01000007	cgd5_2260	Low complexity hypothetical protein	Chro.60427	3	924	34	5
AAEE01000014	cgd6_220	hypothetical protein, possible G-patch domain	Chro.30092	3	534	21	3
AAEE01000001	cgd7_730	hypothetical protein	Chro.40021	3	441	17	3
AAEE01000002	cgd6_3440	hypothetical protein	Chro.30421	3	570	18	3
AAEE01000010	cgd5_4460	hypothetical protein	Chro.20065	3	1023	36	3
AAEE01000004	cgd3_990	CG6144-like AlkB	Chro.30338	3	543	17	3
AAEE01000011	cgd8_210	transcription factor TFIID, TBP	Chro.60267	3	588	21	3
AAEE01000009	cgd4_20	hypothetical protein having a signal peptide, telomeric gene	Chro.30460	3	1008	32	5
AAEE01000007	cgd5_210	hypothetical protein	Chro.50157	3	3258	105	5
AAEE01000007	cgd5_2840	hypothetical protein with 4 transmembrane domains near carboxy terminus	Chro.50091	3	7938	266	6

AAEE01000012	cgd4_4000	possible ribosomal-protein-alanine acetyltransferase	Chro.30486	3	459	18	3
AAEE01000007	cgd5_3420	*cryptosporidium* TRAP-C2 extracellular protein with the following domain organization: signal peptide-Tox1-notch-TSP(13)-Tox1-notch-Tox1-notch-Tox1- notch-Tox1-Tox1-notch-Tox1(4)-sushi(4)	Chro.50029	3	11610	359	3
AAEE01000007	cgd5_290	hypothetical protein, 6 transmembrane domains	Chro.10356	3	762	25	3
AAEE01000003	cgd8_3690	mitosis protein DIM1	Chro.70416	3	429	17	3
AAEE01000004	cgd3_2660	hypothetical protein	Chro.30288	3	612	19	3
AAEE01000005	cgd2_2250	putative integral membrane protein	Chro.10095	3	495	18	3
AAEE01000007	cgd5_140	ARF GTPase activating protein, putative	Chro.10023	3	579	18	3
AAEE01000002	cgd6_1650	conserved hypothetical protein	Chro.30100	3	576	21	3
AAEE01000003	cgd8_5070	hypothetical protein	Chro.10357	3	429	17	3
AAEE01000002	cgd6_2020	BT1 family protein	Chro.80548	3	432	17	3
AAEE01000005	cgd2_2970	DNA-directed RNA polymerase II, putative	Chro.70455	3	537	17	3
AAEE01000002	cgd6_3000	ferredoxin-like protein Fd1, putative	Chro.30168	3	504	16	3
AAEE01000006	cgd1_440	hypothetical protein	Chro.80486	3	438	17	3
AAEE01000001	cgd7_2500	Skp1 family protein, putative	Chro.80163	3	489	19	3
AAEE01000001	cgd7_2840	ubiquitin conjugating enzyme, putative	Chro.70611	3	480	17	3
AAEE01000003	cgd8_1040	hypothetical protein	Chro.60374	3	651	26	3
AAEE01000006	cgd1_2310	unnamed protein product	Chro.20466	3	468	18	3
AAEE01000003	cgd8_3630	Similar to CGI-126 protein, putative	Chro.80363	3	468	16	3
AAEE01000002	cgd6_970	hypothetical protein	Chro.40119	3	525	18	3
AAEE01000003	cgd8_4250	hypothetical protein	Chro.40348	3	723	23	3
AAEE01000006	cgd1_3470	fork head domain protein, putative	Chro.20302	3	525	18	3
AAEE01000002	cgd6_1390	ribosomal protein S23	Chro.60392	3	486	16	3
AAEE01000003	cgd8_4900	clathrin assembly protein, putative	Chro.60195	3	477	19	3
AAEE01000004	cgd3_4320	possible similarity to RNA polymerase Rpb4	Chro.50268	4	375	17	4
AAEE01000004	cgd3_1410	small hypothetical protein with transcripts identified by EST	Chro.60411	4	333	16	4
AAEE01000005	cgd2_4000	hypothetical protein	Chro.40338	4	405	17	4
AAEE01000014	cgd6_200	*cryptosporidium* oocyst wall protein 8, CpCOWP8, signal peptide	Chro.60138	4	1374	65	12
AAEE01000002	cgd6_4980	hypothetical protein	Chro.80548	4	384	17	4
AAEE01000004	cgd3_3760	small protein with possible EF hand domains, calmodulin like	Chro.60242	4	432	18	4

AAEE01000009	cgd4_1010	hypothetical protein	Chro.40121	4	5589	239	8
AAEE01000003	cgd8_5100	hypothetical protein	Chro.30413	4	525	21	4
AAEE01000002	cgd6_380	hypothetical protein with signal peptide and 4 transmembrane domains, possible apicomplexan conserved	Chro.10229	4	444	20	4
AAEE01000008	cgd4_2400	60S ribosomal protein L31, transcript identified by EST	Chro.50110	4	348	16	4
AAEE01000002	cgd6_930	Dim1p-like ERMB/KSGA methylase	Chro.70119	4	1158	47	8
AAEE01000007	cgd5_2430	similar to prefoldin	Chro.50088	4	492	20	4
AAEE01000018	cgd5_4530	hypothetical protein with signal peptide and 2 *cryptosporidium*-specific paralogs	Chro.20423	4	819	36	6
AAEE01000008	cgd4_2290	hypothetical protein	Chro.70304	4	417	18	4
AAEE01000007	cgd5_260	hypothetical protein	Chro.60196	4	2478	117	18
AAEE01000014	cgd6_330	hypothetical protein	Chro.40297	4	525	23	4
AAEE01000006	cgd1_3770	hypothetical protein	Chro.10098	4	303	15	4
AAEE01000001	cgd7_2120	hypothetical protein	Chro.70293	4	411	20	4
AAEE01000007	cgd5_2650	hypothetical protein	Chro.50182	4	423	18	4
AAEE01000002	cgd6_1240	hypothetical protein	Chro.50381	4	468	19	4
AAEE01000006	cgd1_1050	hypothetical protein with predicted 3x transmembrane domains, similarity to KOG1726 (HVA22/DP1 gene product)	Chro.50331	4	438	20	4
AAEE01000001	cgd7_5070	small conserved protein	Chro.70226	4	387	18	4
AAEE01000002	cgd6_910	protein with possible 2 transmembrane domains, possible ER retention signal, similar to Sec20 is a membrane glycoprotein associated with secretory pathway	Chro.30424	4	405	19	4
AAEE01000004	cgd3_2290	possible domain AAA, ATpase family	Chro.20457	4	393	18	4
AAEE01000006	cgd1_430	possible emp24/gp25L/p24 family protein, transmembrane domain, trancript detected by EST	Chro.70470	4	393	18	4
AAEE01000018	cgd5_4550	hypothetical protein	Chro.80514	4	426	18	4
AAEE01000007	cgd5_570	hypothetical protein, 5 transmembrane domains near N-terminus	Chro.80285	4	408	18	4
AAEE01000003	cgd8_4590	hypothetical protein	Chro.10011	4	675	33	7
AAEE01000006	cgd1_2270	40S ribosomal protein S26	Chro.10099	4	324	15	4
AAEE01000009	cgd4_770	Low complexity protein with large Glu repeat	Chro.50438	4	4320	203	33
AAEE01000002	cgd6_1850	anaphase promoting complex subunit 10, putative	Chro.70143	4	399	17	4
AAEE01000005	cgd2_2440	hypothetical protein	Chro.10417	4	366	18	4
AAEE01000005	cgd2_2740	hypothetical protein	Chro.80387	4	363	17	4
AAEE01000005	cgd2_2820	transcription factor, putative	Chro.70453	4	375	17	4
AAEE01000002	cgd6_2400	hypothetical protein	Chro.30048	4	303	15	4

AAEE01000004	cgd3_2430	hypothetical protein	Chro.20064	4	390	18	4
AAEE01000005	cgd2_4330	hypothetical protein	Chro.60160	4	411	17	4
AAEE01000001	cgd7_2800	ubiquitin-conjugating enzyme E2, putative	Chro.20030	4	414	18	4
AAEE01000001	cgd7_4370	RNA polymerase II, putative	Chro.20066	4	348	17	4
AAEE01000001	cgd7_4870	hypothetical protein	Chro.50468	4	414	17	4
AAEE01000006	cgd1_3270	hypothetical protein	Chro.40076	4	1815	87	11
AAEE01000011	cgd8_580	ubiquitin-conjugating enzyme E2, putative	Chro.80357	4	423	18	4
AAEE01000007	cgd5_2160	hypothetical protein	Chro.50015	4	414	17	4
AAEE01000003	cgd8_2140	hypothetical protein	Chro.60133	5	312	17	5
AAEE01000006	cgd1_900	hypothetical protein	Chro.50153	5	318	16	5
AAEE01000007	cgd5_3440	hypothetical protein	Chro.10387	5	312	18	5
AAEE01000002	cgd6_4100	hypothetical protein	Chro.80087	5	306	17	5
AAEE01000010	cgd5_4200	transcription initiation factor IIA	Chro.80451	5	336	20	5
AAEE01000001	cgd7_1880	60S ribosomal protein L44	Chro.10216	5	315	17	5
AAEE01000003	cgd8_1130	hypothetical protein	Chro.80542	5	384	23	5
AAEE01000002	cgd6_5220	cysteine-rich protein with zinc finger	Chro.30098	5	1929	111	15
AAEE01000004	cgd3_3400	similar to hypothetical protein, domain with limited similarity to a domain RPA14, replication protein A (RPA), subunit RPA14	Chro.60246	5	333	19	5
AAEE01000004	cgd3_1080	hypothetical protein	Chro.10084	5	345	20	5
AAEE01000004	cgd3_1040	hypothetical protein	Chro.70399	5	345	18	5
AAEE01000005	cgd2_3560	hypothetical protein with 8 transmembrane domains	Chro.30110	5	1356	76	16
AAEE01000006	cgd1_3140	P-loop nucleotide (UMP) kinase	Chro.80460	5	618	35	9
AAEE01000006	cgd1_3560	hypothetical protein	Chro.80398	5	381	21	5
AAEE01000009	cgd4_1260	possible tRNA-INTRON ENDONUCLEASE	Chro.40032	5	477	25	5
AAEE01000004	cgd3_2550	hypothetical protein with carboxy terminus motif shared with DNA-directed RNA polymerase subunit and TFIIS	Chro.80053	5	612	31	8
AAEE01000004	cgd3_3890	60S ribosomal protein L30, pelota RNA binding domain containing protein	Chro.40217	5	336	18	5
AAEE01000001	cgd7_1070	snRNP core protein homolog Sm-X5. SM domain containing protein.	Chro.70451	5	324	19	5
AAEE01000003	cgd8_1710	hypothetical protein	Chro.70483	5	375	21	5
AAEE01000001	cgd7_1870	hypothetical protein	Chro.70265	5	312	18	5
AAEE01000001	cgd7_690	small nuclear ribonucleoprotein	Chro.30192	5	288	16	5
AAEE01000009	cgd4_1250	hypothetical protein	Chro.80610	5	306	18	5

AAEE01000009	cgd4_1500	hypothetical protein	Chro.10355	5	378	22	5
AAEE01000007	cgd5_3390	very large hypothetical protein	Chro.50033	5	11130	585	5
AAEE01000003	cgd8_1930	large protein with a GCN1 domain	Chro.80225	5	10188	581	5
AAEE01000008	cgd4_2030	centromeric histone h3-like protein, cse4 like	Chro.80310	5	297	15	5
AAEE01000004	cgd3_2380	hypothetical protein	Chro.30018	5	366	19	5
AAEE01000006	cgd1_1190	hypothetical protein	Chro.70205	5	324	17	5
AAEE01000008	cgd4_2830	Mra1/NEP1 like protein, involved in pre-rRNA processing, adjacent genes putative paralogs	Chro.80070	5	651	35	8
AAEE01000005	cgd2_3910	hypothetical protein	Chro.20131	5	321	17	5
AAEE01000002	cgd6_4900	acylphosphatase, putative	Chro.20235	5	369	21	5
AAEE01000003	cgd8_3410	hypothetical protein	Chro.70268	5	309	17	5
AAEE01000008	cgd4_3760	hypothetical protein	Chro.50015	5	336	17	5
AAEE01000009	cgd4_1050	hypothetical protein	Chro.60222	5	420	23	5
AAEE01000007	cgd5_200	hypothetical protein	Chro.60165	6	342	21	6
AAEE01000004	cgd3_2340	hypothetical protein	Chro.50232	6	309	21	6
AAEE01000013	cgd2_70	ABC transporter, with 12 x transmembrane domains and 2x AAA domains, transcripts identified by EST	Chro.20017	6	4764	333	38
AAEE01000006	cgd1_300	40S ribosomal protein S21	Chro.80601	6	243	17	6
AAEE01000006	cgd1_2880	13 kda membrane protein subunit [*cryptosporidium parvum*, sporozoites, peptide, 119 aa], transcript identified by EST	Chro.40430	6	234	16	6
AAEE01000014	cgd6_340	hypothetical protein	Chro.50342	6	333	22	6
AAEE01000004	cgd3_2360	putative nucleoporin, FG-rich motifs within N-terminal region	Chro.30276	6	5418	364	13
AAEE01000010	cgd5_3670	hypothetical protein	Chro.50486	6	1215	80	6
AAEE01000004	cgd3_1070	synaptobrevin like SNARE	Chro.70137	6	312	20	6
AAEE01000006	cgd1_720	apicomplexan-conserved protein	Chro.80467	6	636	44	14
AAEE01000003	cgd8_4560	hypothetical protein	Chro.70144	6	486	30	9
AAEE01000004	cgd3_1420	conserved small protein, transcript identified by EST	Chro.80150	6	231	16	6
AAEE01000001	cgd7_170	hypothetical protein	Chro.20017	6	285	18	6
AAEE01000007	cgd5_470	hypothetical protein, transcripts identified by EST	Chro.30098	6	534	33	6
AAEE01000004	cgd3_2250	60S ribosomal protein L37A, transcripts identified by EST	Chro.80242	6	264	17	6
AAEE01000007	cgd5_430	70 kDa peptidylprolyl isomerase, putative	Chro.10018	6	237	16	6
AAEE01000004	cgd3_2310	hypothetical protein	Chro.80319	6	258	16	6
AAEE01000001	cgd7_1520	peptidylprolyl isomerase, putative	Chro.40479	6	252	17	6

AAEE01000001	cgd7_4210	Sec61-gamma subunit of protein translocation complex, putative	Chro.80306	6	264	17	6
AAEE01000007	cgd5_3000	hypothetical protein	Chro.10328	6	285	18	6
AAEE01000008	cgd4_1660	small nuclear ribonucleoprotein, putative	Chro.10062	6	264	17	6
AAEE01000009	cgd4_890	hypothetical protein	Chro.70285	6	258	17	6
AAEE01000008	cgd4_3840	Ank, ankyrin repeats containing protein with 9 transmembrane domains at C-terminus	Chro.40438	7	6564	492	7
AAEE01000001	cgd7_3520	hypothetical protein	Chro.80114	7	222	16	7
AAEE01000003	cgd8_3550	hypothetical protein	Chro.40249	7	309	22	7
AAEE01000002	cgd6_1790	hypothetical protein	Chro.60216	7	2880	202	7
AAEE01000003	cgd8_4050	ribosomal protein S29	Chro.40360	7	195	15	7
AAEE01000006	cgd1_3410	possible domain similar to DUF392, domain of unknown function (DUF392), probable DNA replication complex GINS protein PSF2	Chro.30126	7	273	21	7
AAEE01000004	cgd3_2770	hypothetical protein	Chro.30315	7	2931	209	7
AAEE01000005	cgd2_620	similar to protein translocation complex beta; protein transport protein SEC61 beta subunit	Chro.50406	7	201	16	7
AAEE01000002	cgd6_5110	large hypothetical protein with possible signal peptide	Chro.60590	7	8136	630	7
AAEE01000009	cgd4_1290	small hypothetical protein, possible conserved	Chro.70065	7	240	17	7
AAEE01000001	cgd7_4990	ubiquitin ligase with a HECT domain at the C-terminus	Chro.70558	7	20043	1524	22
AAEE01000007	cgd5_1430	hypothetical protein	Chro.60269	7	387	30	12
AAEE01000006	cgd1_3430	hypothetical protein	Chro.70011	7	312	22	7
AAEE01000006	cgd1_3370	similarity to domain KOG2265, KIAA1068 protein and nuclear distribution protein NUDC	Chro.20426	7	636	45	7
AAEE01000004	cgd3_2670	protein with UBC domain, ubiquitin conjugating enzyme E2	Chro.60479	7	453	36	7
AAEE01000005	cgd2_2150	hypothetical protein	Chro.10022	7	408	30	11
AAEE01000010	cgd5_3810	hypothetical protein with a signal peptide	Chro.50472	7	1044	78	7
AAEE01000003	cgd8_1010	hypothetical protein	Chro.80553	7	453	35	12
AAEE01000013	cgd2_320	similar to uncharacterized expressed protein	Chro.80073	7	405	30	11
AAEE01000004	cgd3_1950	BIS(5'-nucleosyl)-tetraphosphatase (diadenosine tetraphosphatase), putative	Chro.20448	7	288	22	7
AAEE01000001	cgd7_700	N-acetylglucosaminyl-phosphatidylinositolde-N-acetylase	Chro.30199	7	528	39	12
AAEE01000008	cgd4_3260	putative DNA-directed RNA polymerases I, II, and III 8.3 kda polypeptide	Chro.30046	7	219	16	7
AAEE01000007	cgd5_280	signal peptide containing protein with 6 transmembrane domains	Chro.50351	8	963	80	18
AAEE01000008	cgd4_1810	hypothetical protein	Chro.40032	8	354	29	12
AAEE01000002	cgd6_3490	RNA-binding protein, putative	Chro.10322	8	462	39	14
AAEE01000001	cgd7_510	RNA polymerase III subunit C11	Chro.50080	8	189	17	8

AAEE01000004	cgd3_370	hypothetical protein	Chro.40348	8	315	27	8
AAEE01000002	cgd6_3710	40S ribosomal protein S30, transcripts identified be EST	Chro.40296	8	180	16	8
AAEE01000003	cgd8_5360	hypothetical protein with signal peptide, within telomeric locus of *cryptosporidium*-specific predicted secreted proteins	Chro.50317	8	786	70	23
AAEE01000009	cgd4_190	large low complexity protein	Chro.40033	8	6300	542	8
AAEE01000010	cgd5_4060	protein with N-terminal region ZnF U1 domain, similar to U1 snRNP-specific protein	Chro.80265	8	597	50	18
AAEE01000013	cgd2_140	similar to IMP4 family, transcript identified by EST	Chro.60522	8	198	16	8
AAEE01000006	cgd1_840	similar to GAJ protein	Chro.80146	8	201	17	8
AAEE01000004	cgd3_710	large hypothetical protein with signal peptide	Chro.30095	8	3963	350	8
AAEE01000004	cgd3_3730	hypothetical protein	Chro.30098	8	369	32	14
AAEE01000002	cgd6_500	membrane associated HD superfamily cyclic nucleotide phosphodiesterase domain containing protein	Chro.60068	8	5904	476	15
AAEE01000008	cgd4_2210	N-terminal region similar to putative epsilon-adaptin, probable adaptin	Chro.40252	8	2733	238	8
AAEE01000001	cgd7_4550	ubiquitin-like protein nedd8 homologue, putative	Chro.30042	8	237	19	8
AAEE01000003	cgd8_2360	hypothetical protein	Chro.30090	8	192	16	8
AAEE01000003	cgd8_2510	Cyclin dependent kinase regulatory subunit, putative	Chro.40307	8	258	21	8
AAEE01000013	cgd2_120	ribosomal protein L29	Chro.20151	8	207	17	8
AAEE01000012	cgd4_4070	hypothetical protein	Chro.20236	9	315	30	15
AAEE01000005	cgd2_3690	WD repeat protein	Chro.20394	9	11847	1093	9
AAEE01000012	cgd4_4170	hypothetical protein	Chro.40476	9	462	45	16
AAEE01000003	cgd8_2730	multidomain chromatinic protein with the following architecture: 3x PHD-bromo-3xPHD-SET domain and associated cysteine cluster at the C-terminus	Chro.80318	9	6735	617	9
AAEE01000005	cgd2_3280	hypothetical protein	Chro.40348	9	375	36	9
AAEE01000006	cgd1_1920	large protein with a SPRY domain and HECT domain	Chro.10218	9	13839	1378	9
AAEE01000008	cgd4_1750	hypothetical protein	Chro.70464	9	333	30	14
AAEE01000005	cgd2_2870	similar to 40S ribosomal protein S28, no good start Met	Chro.10245	9	207	19	9
AAEE01000001	cgd7_2440	giant membrane protein with homologs only in plasmodium	Chro.70280	9	5370	526	9
AAEE01000002	cgd6_3320	ypothetical protein	Chro.10343	9	321	29	14
AAEE01000006	cgd1_860	ubiquitin-like protein, putative	Chro.70152	9	219	20	9
AAEE01000004	cgd3_3550	hypothetical protein with signal peptide	Chro.30400	10	3348	339	10
AAEE01000013	cgd2_350	60S ribosomal protein L39, transcripts identified be EST	Chro.80185	10	168	17	10
AAEE01000001	cgd7_2770	uncharacterized low complexity protein	Chro.70314	10	4554	498	10

AAEE01000002	cgd6_1940	hypothetical protein	Chro.60229	10	2838	301	10
AAEE01000005	cgd2_3870	putative ABC transporter with 2x AAA and 11+ transmembrane domains	Chro.20413	10	5787	612	20
AAEE01000003	cgd8_1510	large protein with possibly 18 transmembrane domains	Chro.80178	10	4815	521	10
AAEE01000004	cgd3_3810	similar to clathrin adaptor complex, small subunit	Chro.40020	10	420	45	22
AAEE01000006	cgd1_1400	hypothetical protein	Chro.10162	10	2889	294	10
AAEE01000001	cgd7_2530	large membrane protein with signal peptide and transmembrane domain near carboxy terminus	Chro.70289	10	7164	737	10
AAEE01000002	cgd6_2630	hypothetical protein	Chro.60304	10	4830	486	10
AAEE01000002	cgd6_4310	hypothetical low complexity	Chro.60493	10	5796	594	10
AAEE01000003	cgd8_4420	WD40 repeat and RING finger domain-containing protein	Chro.80507	10	3129	332	10
AAEE01000006	cgd1_1770	domain similar to KOG3415, putative Rab5-interacting protein	Chro.70125	10	288	30	15
AAEE01000008	cgd4_2050	hypothetical protein	Chro.20017	10	402	42	22
AAEE01000003	cgd8_1610	sacsin like HSP90 chaperone domain, likely plant origin	Chro.80189	11	6363	729	11

Accession number	*C. hominis* gene name	Gene Annotation	Best *C. parvum* target	% identity	Query length (bp)	Number of identical nucleotides	Sequence coverage (%)
AAEL01000003	Chro.70187	hypothetical protein	cgd1_1330	0	3096	23	0
AAEL01000717	Chro.00003	sensor histidine kinase	cgd8_3370	0	1647	16	0
AAEL01000018	Chro.40268	hypothetical protein	cgd5_3560	0	2721	19	0
AAEL01000017	Chro.70405	hypothetical protein	cgd7_5010	1	1242	18	1
AAEL01000004	Chro.70479	hypothetical protein	cgd4_2720	1	1029	18	1
AAEL01000170	Chro.80591	hypothetical protein	cgd5_3850	1	1992	24	1
AAEL01000059	Chro.80288	hypothetical protein	cgd1_1390	1	1785	19	1
AAEL01000028	Chro.70434	hypothetical protein	cgd6_3390	1	1374	19	1
AAEL01000267	Chro.50010	hypothetical protein	cgd2_690	1	2082	38	2
AAEL01000452	Chro.50317	RNA polymerase A/beta'/A" subunit	cgd8_4860	1	1419	25	1
AAEL01000256	Chro.30271	hypothetical protein	cgd6_3330	2	675	19	2
AAEL01000369	Chro.50330	leucyl tRNA synthetase (134.5 kD) (lrs-1)	cgd6_3190	2	681	18	2
AAEL01000428	Chro.50457	erythrocyte membrane-associated antigen	cgd8_4810	2	1080	22	2
AAEL01000251	Chro.60010	hypothetical protein	cgd4_1590	2	969	27	2
AAEL01000267	Chro.00007	hypothetical protein	cgd6_4830	2	942	21	2
AAEL01000065	Chro.20156	hypothetical protein	cgd8_2790	2	795	22	2
AAEL01000301	Chro.50310	hypothetical protein	cgd8_5420	2	828	20	2
AAEL01000444	Chro.60598	hypothetical protein	cgd2_3870	2	1755	37	3
AAEL01000368	Chro.40321	hypothetical protein	cgd8_2550	2	732	21	2
AAEL01000117	Chro.70274	hypothetical protein	cgd6_2540	2	621	18	2
AAEL01000413	Chro.50011	hypothetical protein	cgd8_20	3	1470	45	3
AAEL01000001	Chro.30236	hypothetical protein	cgd5_2230	3	543	17	3
AAEL01000232	Chro.30149	ubiquitin-protein ligase 1	cgd8_1200	3	1500	57	8
AAEL01000330	Chro.80592	hypothetical protein	cgd6_4070	3	624	19	3
AAEL01000045	Chro.20364	hypothetical protein	cgd5_2840	3	447	17	3
AAEL01000410	Chro.70032	hypothetical protein	cgd6_3460	3	558	19	3
AAEL01000081	Chro.30091	hypothetical protein	cgd5_40	3	2991	114	11
AAEL01000063	Chro.30044	hypothetical protein	cgd2_3440	4	720	33	7

AAEL01000687	Chro.60599	hypothetical protein	cgd4_380	4	573	23	4
AAEL01000458	Chro.50090	hypothetical protein	cgd2_2600	4	459	19	4
AAEL01000031	Chro.70013	hypothetical protein	cgd7_90	4	714	32	7
AAEL01000068	Chro.30290	hypothetical protein	cgd3_4140	4	423	19	4
AAEL01000813	Chro.10385	hypothetical protein	cgd7_4990	4	429	21	4
AAEL01000668	Chro.30347	hypothetical protein	cgd3_3070	4	1470	64	4
AAEL01000519	Chro.30030	hypothetical protein	cgd2_1920	4	410	19	4
AAEL01000057	Chro.60252	transmembrane protein	cgd6_3860	4	789	32	6
AAEL01000023	Chro.60543	hypothetical protein	cgd8_1150	5	336	17	5
AAEL01000657	Chro.00006	senescence-associated protein	cgd4_3820	5	321	19	5
AAEL01000239	Chro.20422	hypothetical protein	cgd8_2800	5	276	16	5
AAEL01000412	Chro.60093	hypothetical protein	cgd1_3360	5	558	31	8
AAEL01000256	Chro.30272	hypothetical protein	cgd4_3560	5	297	17	5
AAEL01000010	Chro.10301	hypothetical protein	cgd6_270	5	288	17	5
AAEL01000538	Chro.50065	hypothetical protein	cgd2_3510	6	258	17	6
AAEL01000090	Chro.40235	hypothetical protein	cgd2_3510	6	264	17	6
AAEL01000036	Chro.80427	hypothetical protein	cgd8_2490	6	276	18	6
AAEL01000103	Chro.80041	hypothetical protein	cgd1_2670	6	279	17	6
AAEL01001081	Chro.60044	hypothetical protein	cgd5_3460	6	252	17	6
AAEL01000180	Chro.80545	hypothetical protein	cgd1_3250	6	288	20	6
AAEL01000474	Chro.80550	hypothetical protein	cgd2_3580	6	267	18	6
AAEL01000005	Chro.80379	hypothetical protein	cgd4_1340	6	243	17	6
AAEL01000812	Chro.40398	hypothetical protein	cgd8_2630	6	246	16	6
AAEL01000119	Chro.80356	hypothetical protein	cgd3_3230	6	252	17	6
AAEL01000007	Chro.50283	hypothetical protein	cgd5_1030	6	243	17	6
AAEL01000015	Chro.80606	hypothetical protein	cgd5_2790	6	237	16	6
AAEL01000090	Chro.40237	hypothetical protein	cgd5_4000	6	282	18	6
AAEL01000139	Chro.80064	hypothetical protein	cgd6_4460	6	246	16	6
AAEL01000332	Chro.50315	hypothetical protein	cgd8_1610	6	273	18	6
AAEL01000440	Chro.10077	hypothetical protein	cgd5_540	6	252	17	6
AAEL01000124	Chro.60431	hypothetical protein	cgd5_4000	6	237	16	6

AAEL01000122	Chro.80417	hypothetical protein	cgd7_3670	6	270	17	6
AAEL01001123	Chro.70543	hypothetical protein	cgd8_490	6	249	17	6
AAEL01000117	Chro.70268	hypothetical protein	cgd8_3470	6	300	18	6
AAEL01000142	Chro.70073	hypothetical protein	cgd2_2400	6	246	17	6
AAEL01000153	Chro.80322	hypothetical protein	cgd1_2740	6	246	17	6
AAEL01000382	Chro.80170	hypothetical protein	cgd4_580	7	267	21	7
AAEL01000238	Chro.60358	hypothetical protein	cgd2_1160	7	237	18	7
AAEL01000429	Chro.60361	hypothetical protein	cgd7_2260	7	246	18	7
AAEL01000025	Chro.60126	cAMP-dependent protein kinase	cgd3_1340	7	255	18	7
AAEL01000114	Chro.30365	hypothetical protein	cgd6_1940	7	213	17	7
AAEL01000003	Chro.70195	hypothetical protein	cgd8_3700	7	273	20	7
AAEL01000060	Chro.60547	hypothetical protein	cgd4_3520	7	237	17	7
AAEL01000013	Chro.50199	hypothetical protein	cgd8_3260	7	297	21	7
AAEL01000039	Chro.40229	hypothetical protein	cgd2_1440	7	237	17	7
AAEL01000499	Chro.10090	hypothetical protein	cgd1_3220	7	234	18	7
AAEL01000039	Chro.40228	hypothetical protein	cgd2_1430	7	282	22	7
AAEL01000362	Chro.30104	hypothetical protein	cgd8_4030	7	291	21	7
AAEL01000017	Chro.70398	hypothetical protein	cgd5_2440	7	228	16	7
AAEL01000261	Chro.40211	hypothetical protein	cgd6_610	7	207	16	7
AAEL01000353	Chro.50226	hypothetical protein	cgd1_1690	7	225	16	7
AAEL01000003	Chro.70179	hypothetical protein	cgd8_550	7	234	17	7
AAEL01000347	Chro.10127	hypothetical protein	cgd4_1390	7	270	21	7
AAEL01000114	Chro.30364	hypothetical protein	cgd3_50	7	240	18	7
AAEL01000382	Chro.80169	hypothetical protein	cgd4_650	8	204	18	8
AAEL01000009	Chro.20113	hypothetical protein	cgd5_210	8	207	18	8
AAEL01000066	Chro.80037	hypothetical protein	cgd5_3420	8	222	18	8
AAEL01000137	Chro.60099	hypothetical protein	cgd8_3030	8	216	19	8
AAEL01000058	Chro.40286	hypothetical protein	cgd7_5480	8	222	18	8
AAEL01000127	Chro.40012	hypothetical protein	cgd2_1380	8	225	20	8
AAEL01000233	Chro.80531	hypothetical protein	cgd3_400	8	231	19	8
AAEL01000536	Chro.40447	hypothetical protein	cgd5_500	8	213	18	8

AAEL01000050	Chro.60575	hypothetical protein	cgd4_1330	8	213	19	8
AAEL01001242	Chro.70522	hypothetical protein	cgd2_470	8	225	19	8
AAEL01000003	Chro.70158	hypothetical protein	cgd4_770	8	378	32	13
AAEL01000620	Chro.10340	hypothetical protein	cgd4_1360	8	222	19	8
AAEL01000116	Chro.80349	hypothetical protein	cgd8_4410	8	210	17	8
AAEL01000179	Chro.50216	hypothetical protein	cgd8_3220	8	273	22	8
AAEL01000021	Chro.60179	hypothetical protein	cgd3_4180	8	216	18	8
AAEL01000679	Chro.40091	hypothetical protein	cgd6_2260	8	183	16	8
AAEL01000181	Chro.10351	hypothetical protein	cgd5_4130	8	207	17	8
AAEL01000291	Chro.30130	hypothetical protein	cgd3_260	8	228	20	8
AAEL01000611	Chro.80489	hypothetical protein	cgd2_1980	8	240	21	8
AAEL01000079	Chro.80248	hypothetical protein	cgd8_2060	8	222	18	8
AAEL01000506	Chro.30028	hypothetical protein	cgd6_3160	8	207	18	8
AAEL01000337	Chro.40251	hypothetical protein	cgd6_5110	8	318	28	8
AAEL01000127	Chro.40016	hypothetical protein	cgd6_1920	8	243	20	8
AAEL01000004	Chro.70478	hypothetical protein	cgd8_620	8	213	18	8
AAEL01000192	Chro.70515	hypothetical protein	cgd3_600	9	222	20	9
AAEL01000082	Chro.60476	hypothetical protein	cgd4_3930	9	222	22	9
AAEL01000237	Chro.60454	hypothetical protein	cgd6_3690	9	234	23	9
AAEL01000031	Chro.70026	hypothetical protein	cgd4_820	9	216	21	9
AAEL01000031	Chro.70011	hypothetical protein	cgd1_3430	9	243	22	9
AAEL01000101	Chro.20319	hypothetical protein	cgd7_3810	9	207	19	9
AAEL01000434	Chro.50247	hypothetical protein	cgd5_4190	10	312	34	17
AAEL01000234	Chro.80118	hypothetical protein	cgd2_1590	10	228	24	10
AAEL01000298	Chro.10105	hypothetical protein	cgd8_1060	10	297	30	16
AAEL01000453	Chro.80410	hypothetical protein	cgd1_2850	10	300	30	16
AAEL01000091	Chro.10038	hypothetical protein	cgd8_1510	11	243	29	18

Appendix III

Appendix III: Identification of putative species-specific genes for *C. parvum* and *C. hominis* based on Reciprocal BLAST, individual BLAST and ortholog identification on CryptoDB database. 93 and 211genes are putatively specific for *C. hominis* and *C. parvum*, respectively. The selected genes show low level of sequence identity (<10%) to genes from the other species.

A: *C. hominis* specific genes

Reciprocal BLAST				Data from CryptoDB and NCBI BLAST					
							BLAST		
C. hominis gene	Function	*C. parvum* target	% identity	Sequence length	*C hominis*	*C parvum*	% identity	Probability	Homologous gene(s)
Chro.00003	sensor histidine kinase	cgd8_3370	0	1647 bp / 548 AAs	yes	no	0%		
Chro.40268	hypothetical protein	cgd5_3560	0	2721 bp / 906 AAs	yes	no	0%		
Chro.70405	hypothetical protein	cgd7_5010	1	1242 bp / 413 AAs	yes	no	0%		
Chro.70479	hypothetical protein	cgd4_2720	1	1029 bp / 342 AAs	yes	no	0%		
Chro.80591	hypothetical protein	cgd5_3850	1	1992 bp / 663 AAs	yes	no	0%		
Chro.80288	hypothetical protein	cgd1_1390	1	1785 bp / 594 AAs	yes	no	0%		
Chro.70434	hypothetical protein	cgd6_3390	1	1374 bp / 457 AAs	yes	no	0%		
Chro.50010	hypothetical protein	cgd2_690	1	2082 bp / 693 AAs	yes	no	8%	9e-09	C *hominis* Chro.50011
Chro.50317	RNA polymerase A/beta'/A'' subunit	cgd8_4860	1	1419 bp / 472 AAs	yes	no	0%		
Chro.30271	hypothetical protein	cgd6_3330	2	675 bp / 224 AAs	yes	no	0%		
Chro.50330	leucyl tRNA synthetase (134.5 kD) (lrs-1)	cgd6_3190	2	681 bp / 226 AAs	yes	no	0%		
Chro.50457	erythrocyte membrane-associated antigen	cgd8_4810	2	1080 bp / 359 AAs	yes	no	0%		
Chro.60010	hypothetical protein	cgd4_1590	2	969 bp / 322 AAs	yes	no	0%		C *hominis* Chro.00007
Chro.00007	hypothetical protein	cgd6_4830	2	942 bp / 313 AAs	yes	no	0%		C *hominis* Chro.60010
Chro.20156	hypothetical protein	cgd8_2790	2	795 bp / 264 AAs	yes	no	0%		
Chro.40321	hypothetical protein	cgd8_2550	2	732 bp / 243 AAs	yes	no	0%		
Chro.70274	hypothetical protein	cgd6_2540	2	621 bp / 206 AAs	yes	no	0%		
Chro.50011	hypothetical protein	cgd8_20	3	1470 bp / 489 AAs	yes	no	9%	7e-09	C *hominis* Chro.50010
Chro.30236	hypothetical protein	cgd5_2230	3	543 bp / 180 AAs	yes	no	0%		
Chro.30149	ubiquitin-protein ligase 1	cgd8_1200	3	1500 bp / 499 AAs	yes	no	0%		
Chro.80592	hypothetical protein	cgd6_4070	3	624 bp / 207 AAs	yes	no	0%		
Chro.20364	hypothetical protein	cgd5_2840	3	447 bp / 148 AAs	yes	no	0%		
Chro.70032	hypothetical protein	cgd6_3460	3	558 bp / 185 AAs	yes	no	0%		

Chro.30091	hypothetical protein	cgd5_40	3	2991 bp / 996 AAs	yes	no	0%		
Chro.30044	hypothetical protein	cgd2_3440	4	720 bp / 239 AAs	yes	no	0%		
Chro.50090	hypothetical protein	cgd2_2600	4	459 bp / 152 AAs	yes	no	0%		
Chro.70013	hypothetical protein	cgd7_90	4	714 bp / 237 AAs	yes	no	0%		
Chro.30290	hypothetical protein	cgd3_4140	4	423 bp / 140 AAs	yes	no	0%		
Chro.10385	hypothetical protein	cgd7_4990	4	429 bp / 142 AAs	yes	no	0%		
Chro.30030	hypothetical protein	cgd2_1920	4	410 bp / 407/3 AAs	yes	no	0%		
Chro.20422	hypothetical protein	cgd8_2800	5	276 bp / 91 AAs	yes	no	0%		
Chro.60093	hypothetical protein	cgd1_3360	5	558 bp / 185 AAs	yes	no	0%		
Chro.30272	hypothetical protein	cgd4_3560	5	297 bp / 98 AAs	yes	no	0%		
Chro.10301	hypothetical protein	cgd6_270	5	288 bp / 95 AAs	yes	no	0%		
Chro.50065	hypothetical protein	cgd2_3510	6	258 bp / 85 AAs	yes	no	0%		
Chro.80427	hypothetical protein	cgd8_2490	6	276 bp / 91 AAs	yes	no	0%		
Chro.80041	hypothetical protein	cgd1_2670	6	279 bp / 92 AAs	yes	no	0%		
Chro.60044	hypothetical protein	cgd5_3460	6	252 bp / 83 AAs	yes	no	0%		
Chro.80550	hypothetical protein	cgd2_3580	6	267 bp / 88 AAs	yes	no	0%		
Chro.80379	hypothetical protein	cgd4_1340	6	243 bp / 80 AAs	yes	no	0%		
Chro.40398	hypothetical protein	cgd8_2630	6	246 bp / 81 AAs	yes	no	0%		
Chro.80356	hypothetical protein	cgd3_3230	6	252 bp / 83 AAs	yes	no	0%		
Chro.50283	hypothetical protein	cgd5_1030	6	243 bp / 80 AAs	yes	no	13%	2e-07	C *hominis* Chro.50284
Chro.80606	hypothetical protein	cgd5_2790	6	237 bp / 78 AAs	yes	no	0%		
Chro.40237	hypothetical protein	cgd5_4000	6	282 bp / 93 AAs	yes	no	0%		
Chro.80064	hypothetical protein	cgd6_4460	6	246 bp / 81 AAs	yes	no	0%		
Chro.50315	hypothetical protein	cgd8_1610	6	273 bp / 90 AAs	yes	no	0%		
Chro.10077	hypothetical protein	cgd5_540	6	252 bp / 83 AAs	yes	no	0%		
Chro.80417	hypothetical protein	cgd7_3670	6	270 bp / 89 AAs	yes	no	0%		
Chro.70543	hypothetical protein	cgd8_490	6	249 bp / 82 AAs	yes	no	0%		
Chro.70268	hypothetical protein	cgd8_3470	6	300 bp / 99 AAs	yes	no	0%		
Chro.70073	hypothetical protein	cgd2_2400	6	246 bp / 81 AAs	yes	no	0%		
Chro.80322	hypothetical protein	cgd1_2740	6	246 bp / 81 AAs	yes	no	0%		
Chro.30365	hypothetical protein	cgd6_1940	7	213 bp / 70 AAs	yes	no	0%		
Chro.70195	hypothetical protein	cgd8_3700	7	273 bp / 90 AAs	yes	no	0%		
Chro.50199	hypothetical protein	cgd8_3260	7	297 bp / 98 AAs	yes	no	0%		

Chro.40229	hypothetical protein	cgd2_1440	7	237 bp / 78 AAs	yes	no	0%
Chro.10090	hypothetical protein	cgd1_3220	7	234 bp / 77 AAs	yes	no	0%
Chro.40228	hypothetical protein	cgd2_1430	7	282 bp / 93 AAs	yes	no	0%
Chro.30104	hypothetical protein	cgd8_4030	7	291 bp / 96 AAs	yes	no	0%
Chro.70398	hypothetical protein	cgd5_2440	7	228 bp / 75 AAs	yes	no	0%
Chro.40211	hypothetical protein	cgd6_610	7	207 bp / 68 AAs	yes	no	0%
Chro.70179	hypothetical protein	cgd8_550	7	234 bp / 77 AAs	yes	no	0%
Chro.10127	hypothetical protein	cgd4_1390	7	270 bp / 89 AAs	yes	no	0%
Chro.30364	hypothetical protein	cgd3_50	7	240 bp / 79 AAs	yes	no	0%
Chro.80169	hypothetical protein	cgd4_650	8	204 bp / 67 AAs	yes	no	0%
Chro.20113	hypothetical protein	cgd5_210	8	207 bp / 68 AAs	yes	no	0%
Chro.80037	hypothetical protein	cgd5_3420	8	222 bp / 73 AAs	yes	no	0%
Chro.40286	hypothetical protein	cgd7_5480	8	222 bp / 73 AAs	yes	no	0%
Chro.40012	hypothetical protein	cgd2_1380	8	225 bp / 74 AAs	yes	no	0%
Chro.80531	hypothetical protein	cgd3_400	8	231 bp / 76 AAs	yes	no	0%
Chro.40447	hypothetical protein	cgd5_500	8	213 bp / 70 AAs	yes	no	0%
Chro.70522	hypothetical protein	cgd2_470	8	225 bp / 74 AAs	yes	no	0%
Chro.70158	hypothetical protein	cgd4_770	8	378 bp / 125 AAs	yes	no	0%
Chro.10340	hypothetical protein	cgd4_1360	8	222 bp / 73 AAs	yes	no	0%
Chro.80349	hypothetical protein	cgd8_4410	8	210 bp / 69 AAs	yes	no	0%
Chro.50216	hypothetical protein	cgd8_3220	8	273 bp / 90 AAs	yes	no	0%
Chro.40091	hypothetical protein	cgd6_2260	8	183 bp / 60 AAs	yes	no	0%
Chro.10351	hypothetical protein	cgd5_4130	8	207 bp / 68 AAs	yes	no	0%
Chro.30130	hypothetical protein	cgd3_260	8	228 bp / 75 AAs	yes	no	0%
Chro.80489	hypothetical protein	cgd2_1980	8	240 bp / 79 AAs	yes	no	0%
Chro.80248	hypothetical protein	cgd8_2060	8	222 bp / 73 AAs	yes	no	0%
Chro.30028	hypothetical protein	cgd6_3160	8	207 bp / 68 AAs	yes	no	0%
Chro.40251	hypothetical protein	cgd6_5110	8	318 bp / 105 AAs	yes	no	0%
Chro.70478	hypothetical protein	cgd8_620	8	213 bp / 70 AAs	yes	no	0%
Chro.70515	hypothetical protein	cgd3_600	9	222 bp / 73 AAs	yes	no	0%
Chro.70026	hypothetical protein	cgd4_820	9	216 bp / 71 AAs	yes	no	0%
Chro.70011	hypothetical protein	cgd1_3430	9	243 bp / 80 AAs	yes	no	0%
Chro.20319	hypothetical protein	cgd7_3810	9	207 bp / 68 AAs	yes	no	0%

Chro.50247	hypothetical protein	cgd5_4190	10	312 bp / 103 AAs	yes	no	0%
Chro.80118	hypothetical protein	cgd2_1590	10	228 bp / 75 AAs	yes	no	0%
Chro.80410	hypothetical protein	cgd1_2850	10	300 bp / 99 AAs	yes	no	0%

B: *C. parvum* specific genes

Reciprocal BLAST				Data from CryptoDB and NCBI Blast					
							BLAST		
***C. parvum* gene**	**Function**	***C. hominis* target**	**% identity**	**Sequence length**	***C hominis***	***C parvum***	**% identity**	**Probability**	**Homologous gene(s)**
cgd6_5320	hypothetical protein with similarity to many plasmodium proteins	Chro.60073	0	2811 bp / 936 AAs	no	yes	0%		
cgd2_80	ABC transporter family protein, 2x AAA domain	Chro.80487	0	4431 bp / 1476 AAs	no	yes	0%		
cgd5_330	hypothetical protein	Chro.70096	0	3504 bp / 1167 AAs	no	yes	0%	0.91	C *hominis* TU502 hypothetical protein Chro.70096
cgd5_1000	hypothetical protein	Chro.60190	0	2112 bp / 703 AAs	no	yes	0%		
cgd4_1740	possible HSMGG motif (esterase?)	Chro.80330	1	3078 bp / 1025 AAs	no	yes	0%		
cgd4_1330	NADPH-dependent FMN FAD containing oxidoreductase , transcripts identified by EST	Chro.60070	1	2460 bp / 819 AAs	no	yes	0%		
cgd4_290	putative protein phosphatase 2A regulatory B subunit, highly conserved but no plasmodium hits	Chro.80116	1	2076 bp / 691 AAs	no	yes	0%		
cgd3_500	putative fucose translocator with 8 transmembrane domains, within locus of 3 paralogous genes	Chro.40126	1	1185 bp / 394 AAs	no	yes	0%		
cgd3_2870	similarity of possible bacterial origin within discrete region similar to long-chain fatty acid acyl-CoA synthetases	Chro.00004	1	2334 bp / 777 AAs	no	yes	0%		
cgd6_440	PP2Cc like protein phosphatise	Chro.80459	1	1602 bp / 533 AAs	no	yes	0%		
cgd1_110	predicted secreted protein, signal peptide, paralogs	Chro.10158	1	1221 bp / 406 AAs	no	yes	0%		
cgd6_540	Ser/Thr protein kinase	Chro.10402	1	2841 bp / 946 AAs	no	yes	0%		
cgd4_1570	hypothetical protein, possible conserved	Chro.50401	1	1224 bp / 407 AAs	no	yes	0%		
cgd3_4240	insulinase like peptidase	Chro.50235	1	3342 bp / 1113 AAs	no	yes	0%	0.87	C *parvum* Chr 6
cgd3_2640	hypothetical protein with conserved domain	Chro.40078	1	1311 bp / 436 AAs	no	yes	0%		
cgd5_4190	hypothetical protein	Chro.20329	1	2247 bp / 748 AAs	no	yes	0%		

cgd4_1130	PhnP like hydrolase of the metallobetalactamase fold	Chro.10012	1	1065 bp / 354 AAs	no	yes	0%		
cgd2_4380	signal peptide, repeats, gene anchored to telomere	Chro.10099	1	1434 bp / 477 AAs	no	yes	0%		
cgd5_4000	hypothetical protein	Chro.40042	1	2034 bp / 677 AAs	no	yes	0%		
cgd3_510	putative fucose translocatorw ith 8 transmembrane domains, within locus of 3 paralogous genes	Chro.30381	1	1170 bp / 389 AAs	no	yes	0%		
cgd6_320	hypothetical protein with signal peptide and 8 transmembrane domains, amino terminl region conserved in plasmodium, transcripts identified by EST	Chro.80433	1	2427 bp / 808 AAs	no	yes	0%		
cgd2_10	hypothetical protein	Chro.50504	1	1338 bp / 445 AAs	no	yes	0%		
cgd8_2370	adenosine kinase like ribokinase	Chro.40163	1				97%	0.0	C *parvum* adenosine kinase (AK) gene
				1188 bp / 395 AAs	no	yes	1%	0.31	C *parvum* Iowa II hypothetical protein cgd8_4780
cgd6_670	extacellular protein with a signal peptide, clostripain like caspase/hemoglobinase domain, notch domain and 2 EGF domains	Chro.70432	1	4824 bp / 1607 AAs	no	yes	0%		
cgd1_360	similar to Noc2p N-terminus	Chro.60406	1	2511 bp / 836 AAs	no	yes	0%		
cgd4_1780	ring domain protein	Chro.50317	1	1434 bp / 477 AAs	no	yes	0%		
cgd6_5120	protein with two PHD Zn fingers that is probably involved in chromatin function	Chro.40494	2	687 bp / 228 AAs	no	yes	0%		
cgd4_3990	hypothetical protein with possible transmembrane domain near N-terminus	Chro.70342	2	1743 bp / 580 AAs	no	yes	0%		
cgd6_1130	hypothetical protein with signal peptide	Chro.40330	2	1827 bp / 608 AAs	no	yes	0%		
cgd6_5510	telomeric insulinase-like protease with signal peptide	Chro.40328	2						C *parvum* isolate Ontario zinc protease telomerase C *parvum* isolate Michigan zinc protease telomerase
				1473 bp / 490 AAs	no	yes	41%	0.0	C *parvum* isolate Georgia zinc protease telomerase
cgd2_3800	hypothetical protein	Chro.60478	2	1110 bp / 369 AAs	no	yes	2%	4.5	C *parvum* Iowa II DHHC family palmitoyl transferase cgd8_4730
cgd5_4510	hypothetical protein	Chro.40216	2	789 bp / 262 AAs	no	yes	0%		
cgd5_320	carboxylesterase , lysophospholipase, signal peptide	Chro.10070	2	1422 bp / 473 AAs	no	yes	0%		

cgd2_2430	ximpact ortholog conserved protein seen in bacteria and eukaryotes	Chro.10042	2	873 bp / 290 AAs	no	yes	0%		
cgd2_100	hypothetical protein	Chro.30403	2	723 bp / 240 AAs	no	yes	0%		
cgd3_1090	Rrp9p/U3-55K-family snoRNP-associated protein with several WD40 repeats	Chro.50400	2	1374 bp / 457 AAs	no	yes	0%		
cgd8_3150	hypothetical conserved protein, transcripts identified by EST	Chro.30367	2	582 bp / 193 AAs	no	yes	0%		
cgd4_4160	possible carboxypeptidase	Chro.20426	2	1371 bp / 456 AAs	no	yes	0%		
cgd1_960	RING finger containing protein	Chro.20379	2	1458 bp / 485 AAs	no	yes	0%		
cgd5_3290	transcription elongation factor TFIIS	Chro.60122	2	702 bp / 233 AAs	no	yes	0%		
cgd2_3970	RRM domain containing protein	Chro.40100	2	670 bp / 222 AAs	no	yes	0%		
cgd8_620	hypothetical protein	Chro.20223	2	783 bp / 260 AAs	no	yes	0%		
cgd6_1810	hypothetical protein	Chro.10285	2	675 bp / 224 AAs	no	yes	0%		
cgd6_5520	peptidase'insulinase like peptidase'	Chro.40162	2	1713 bp / 570 AAs	no	yes	0%		
cgd6_5020	protein with WD40 repeats	Chro.50196	3	651 bp / 216 AAs	no	yes	0%		
cgd5_3250	secreted lipopolysaccharide sugar transferase like family 8 glycosyltransferase	Chro.10377	3	477 bp / 158 AAs	no	yes	0%		
cgd5_190	plasmodium conserved protein	Chro.20379	3	1041 bp / 346 AAs	no	yes	0%		
cgd4_3920	DinB/family X-type DNA polymerase	Chro.30486	3	1077 bp / 358 AAs	no	yes	4%	3e-04	C *parvum* Iowa II archeo-eukaryotic exosomal RNAse cgd4_3930
cgd4_90	hypothetical protein, similarity to Rrp7p	Chro.70603	3	531 bp / 176 AAs	no	yes	0%		
cgd1_3280	hypothetical protein	Chro.40376	3	939 bp / 312 AAs	no	yes	0%		
cgd6_220	hypothetical protein, possible G-patch domain	Chro.30092	3	534 bp / 177 AAs	no	yes	3%	2.1	C *parvum* Iowa II hypothetical protein cgd5_260
cgd7_730	hypothetical protein	Chro.40021	3	441 bp / 146 AAs	no	yes	0%		

cgd6_3440	hypothetical protein	Chro.30421	3	570 bp / 189 AAs	no	yes	0%		
cgd5_4460	hypothetical protein	Chro.20065	3	1023 bp / 340 AAs	no	yes	0%		
cgd3_990	CG6144-like AlkB	Chro.30338	3	543 bp / 180 AAs	no	yes	0%		
cgd8_210	transcription factor TFIID, TBP	Chro.60267	3	588 bp / 195 AAs	no	yes	0%		
cgd4_20	hypothetical protein having a signal peptide, telomeric gene	Chro.30460	3	1008 bp / 335 AAs	no	yes	0%		
cgd5_210	hypothetical protein	Chro.50157	3	3258 bp / 1085 AAs	no	yes	0%		
cgd4_4000	possible ribosomal-protein-alanine acetyltransferase	Chro.30486	3	459 bp / 152 AAs	no	yes	0%		
cgd5_290	hypothetical protein, 6 transmembrane domains	Chro.10356	3	762 bp / 253 AAs	no	yes	0%		
cgd8_3690	mitosis protein DIM1	Chro.70416	3	429 bp / 142 AAs	no	yes	0%		
cgd3_2660	hypothetical protein	Chro.30288	3	612 bp / 203 AAs	no	yes	0%		
cgd6_1650	conserved hypothetical protein	Chro.30100	3	618 bp / 205 AAs	no	yes	0%		
cgd8_5070	hypothetical protein	Chro.10357	3	487 bp / 161 AAs	no	yes	0%		
cgd6_3000	ferredoxin-like protein Fd1, putative	Chro.30168	3	937 bp / 311 AAs	no	yes	0%		
							4%	0.13	C *parvum* Iowa II protein kinase cgd1_400
cgd6_970	hypothetical protein	Chro.40119	3	607 bp / 201 AAs	no	yes	6%	2.1	C *parvum* Iowa II putative peptidase family C54 cgd3_2920
cgd3_4320	possible similarity to RNA polymerase Rpb4	Chro.50268	4	427 bp / 141 AAs	no	yes	0%		
cgd3_1410	small hypothetical protein with transcripts identified by EST	Chro.60411	4	333 bp / 110 AAs	no	yes	0%		
cgd2_4000	hypothetical protein	Chro.40338	4	405 bp / 134 AAs	no	yes	0%		
cgd6_200	*cryptosporidium* oocyst wall protein 8, CpCOWP8, signal peptide	Chro.60138	4	1374 bp / 457 AAs	no	yes	0%		
cgd6_4980	hypothetical protein	Chro.80548	4	384 bp / 127 AAs	no	yes	0%		

cgd3_3760	small protein with possible EF hand domains, calmodulin like	Chro.60242	4	432 bp / 143 AAs	no	yes	0%
cgd8_5100	hypothetical protein	Chro.30413	4	525 bp / 174 AAs	no	yes	0%
cgd6_380	hypothetical protein with signal peptide and 4 transmembrane domains, possible apicomplexan conserved	Chro.10229	4	444 bp / 147 AAs	no	yes	0%
cgd4_2400	60S ribosomal protein L31, transcript identified by EST	Chro.50110	4	426 bp / 141 AAs	no	yes	0%
cgd6_930	Dim1p-like ERMB/KSGA methylase	Chro.70119	4	1158 bp / 385 AAs	no	yes	0%
cgd5_4530	hypothetical protein with signal peptide and 2 *cryptosporidium*-specific paralogs	Chro.20423	4	819 bp / 272 AAs	no	yes	0%
cgd4_2290	hypothetical protein	Chro.70304	4	417 bp / 138 AAs	no	yes	0%
cgd5_260	hypothetical protein	Chro.60196	4	2478 bp / 825 AAs	no	yes	0%
cgd1_3770	hypothetical protein	Chro.10098	4	303 bp / 100 AAs	no	yes	0%
cgd7_2120	hypothetical protein	Chro.70293	4	411 bp / 136 AAs	no	yes	0%
cgd5_2650	hypothetical protein	Chro.50182	4	423 bp / 140 AAs	no	yes	0%
cgd6_1240	hypothetical protein	Chro.50381	4	468 bp / 155 AAs	no	yes	0%
cgd7_5070	small conserved protein	Chro.70226	4	387 bp / 128 AAs	no	yes	0%
cgd6_910	protein with possible 2 transmembrane domains, possible ER retention signal, similar to Sec20 is a membrane glycoprotein associated with secretory pathway	Chro.30424	4	405 bp / 134 AAs	no	yes	0%
cgd3_2290	possible domain AAA, ATpase family	Chro.20457	4	393 bp / 130 AAs	no	yes	0%
cgd1_430	possible emp24/gp25L/p24 family protein, transmembrane domain, trancript detected by EST	Chro.70470	4	393 bp / 130 AAs	no	yes	0%
cgd5_4550	hypothetical protein	Chro.80514	4	426 bp / 141 AAs	no	yes	0%
cgd5_570	hypothetical protein, 5 transmembrane domains near N-terminus	Chro.80285	4	408 bp / 135 AAs	no	yes	0%
cgd8_4590	hypothetical protein	Chro.10011	4	675 bp / 224 AAs	no	yes	0%

cgd1_2270	40S ribosomal protein S26	Chro.10099	4	324 bp / 107 AAs	no	yes	0%
cgd6_1850	anaphase promoting complex subunit 10, putative	Chro.70143	4	441 bp / 146 AAs	no	yes	0%
cgd6_2400	hypothetical protein	Chro.30048	4	355 bp / 117 AAs	no	yes	0%
cgd8_2140	hypothetical protein	Chro.60133	5	312 bp / 103 AAs	no	yes	0%
cgd1_900	hypothetical protein	Chro.50153	5	318 bp / 105 AAs	no	yes	0%
cgd5_3440	hypothetical protein	Chro.10387	5	312 bp / 103 AAs	no	yes	0%
cgd6_4100	hypothetical protein	Chro.80087	5	306 bp / 101 AAs	no	yes	0%
cgd5_4200	transcription initiation factor IIA	Chro.80451	5	336 bp / 111 AAs	no	yes	0%
cgd7_1880	60S ribosomal protein L44	Chro.10216	5	315 bp / 104 AAs	no	yes	0%
cgd8_1130	hypothetical protein	Chro.80542	5	384 bp / 127 AAs	no	yes	0%
cgd6_5220	cysteine-rich protein with zinc finger	Chro.30098	5	1929 bp / 642 AAs	no	yes	0%
cgd3_3400	similar to hypothetical protein, domain with limited similarity to a domain RPA14, replication protein A (RPA), subunit RPA14	Chro.60246	5	333 bp / 110 AAs	no	yes	0%
cgd3_1040	hypothetical protein	Chro.70399	5	345 bp / 114 AAs	no	yes	0%
cgd1_3140	P-loop nucleotide (UMP) kinase	Chro.80460	5	618 bp / 205 AAs	no	yes	0%
cgd1_3560	hypothetical protein	Chro.80398	5	381 bp / 126 AAs	no	yes	0%
cgd4_1260	possible tRNA-INTRON ENDONUCLEASE	Chro.40032	5	477 bp / 158 AAs	no	yes	0%
cgd3_2550	hypothetical protein with carboxy terminus motif shared with DNA-directed RNA polymerase subunit and TFIIS	Chro.80053	5	612 bp / 203 AAs	no	yes	0%
cgd3_3890	60S ribosomal protein L30, pelota RNA binding domain containing protein	Chro.40217	5	336 bp / 111 AAs	no	yes	0%
cgd7_1070	snRNP core protein homolog Sm-X5. SM domain containing protein.	Chro.70451	5	324 bp / 107 AAs	no	yes	0%
cgd8_1710	hypothetical protein	Chro.70483	5	375 bp / 124 AAs	no	yes	0%

cgd7_1870	hypothetical protein	Chro.70265	5	312 bp / 103 AAs	no	yes	0%
cgd7_690	small nuclear ribonucleoprotein	Chro.30192	5	288 bp / 95 AAs	no	yes	0%
cgd4_1250	hypothetical protein	Chro.80610	5	306 bp / 101 AAs	no	yes	0%
cgd4_1500	hypothetical protein	Chro.10355	5	378 bp / 125 AAs	no	yes	0%
cgd4_2030	centromeric histone h3-like protein, cse4 like	Chro.80310	5	297 bp / 98 AAs	no	yes	0%
cgd3_2380	hypothetical protein	Chro.30018	5	366 bp / 121 AAs	no	yes	0%
cgd1_1190	hypothetical protein	Chro.70205	5	324 bp / 107 AAs	no	yes	0%
cgd4_2830	Mra1/NEP1 like protein, involved in pre-rRNA processing, adjacent genes putative paralogs	Chro.80070	5	651 bp / 216 AAs	no	yes	0%
cgd6_4900	acylphosphatase, putative	Chro.20235	5	417 bp / 138 AAs	no	yes	0%
cgd3_2340	hypothetical protein	Chro.50232	6	309 bp / 102 AAs	no	yes	0%
cgd1_300	40S ribosomal protein S21	Chro.80601	6	243 bp / 80 AAs	no	yes	0%
cgd1_2880	13 kda membrane protein subunit [*cryptosporidium parvum*, sporozoites, peptide, 119 aa], transcript identified by EST	Chro.40430	6	234 bp / 77 AAs	no	yes	0%
cgd6_340	hypothetical protein	Chro.50342	6	333 bp / 110 AAs	no	yes	0%
cgd3_1070	synaptobrevin like SNARE	Chro.70137	6	312 bp / 103 AAs	no	yes	0%
cgd8_4560	hypothetical protein	Chro.70144	6	486 bp / 161 AAs	no	yes	0%
cgd3_1420	conserved small protein, transcript identified by EST	Chro.80150	6	231 bp / 76 AAs	no	yes	0%
cgd7_170	hypothetical protein	Chro.20017	6	285 bp / 94 AAs	no	yes	0%
cgd5_470	hypothetical protein, transcripts identified by EST	Chro.30098	6	534 bp / 177 AAs	no	yes	0%
cgd3_2250	60S ribosomal protein L37A, transcripts identified by EST	Chro.80242	6	264 bp / 87 AAs	no	yes	0%
cgd7_3520	hypothetical protein	Chro.80114	7	222 bp / 73 AAs	no	yes	0%
cgd8_3550	hypothetical protein	Chro.40249	7	309 bp / 102 AAs	no	yes	0%

cgd8_4050	ribosomal protein S29	Chro.40360	7	195 bp / 64 AAs	no	yes	0%		
cgd1_3410	possible domain similar to DUF392, domain of unknown function (DUF392), probable DNA replication complex GINS protein PSF2	Chro.30126	7	273 bp / 90 AAs	no	yes	0%		
cgd2_620	similar to protein translocation complex beta; protein transport protein SEC61 beta subunit	Chro.50406	7	201 bp / 66 AAs	no	yes	0%		
cgd5_1430	hypothetical protein	Chro.60269	7	387 bp / 128 AAs	no	yes	0%		
cgd1_3430	hypothetical protein	Chro.70011	7	312 bp / 103 AAs	no	yes	0%		
cgd1_3370	similarity to domain KOG2265, KIAA1068 protein and nuclear distribution protein NUDC	Chro.20426	7	636 bp / 211 AAs	no	yes	0%		
cgd2_2150	hypothetical protein	Chro.10022	7	408 bp / 135 AAs	no	yes	0%		
cgd2_320	similar to uncharacterized expressed protein	Chro.80073	7	481 bp / 159 AAs	no	yes	0%		
cgd5_280	signal peptide containing protein with 6 transmembrane domains	Chro.50351	8	963 bp / 320 AAs	no	yes	2%	0.25	C *parvum* Iowa II mannose-1-phosphate guanylyltransferase cgd2_1770
cgd4_1810	hypothetical protein	Chro.40032	8	354 bp / 117 AAs	no	yes	0%		
cgd7_510	RNA polymerase III subunit C11	Chro.50080	8	189 bp / 62 AAs	no	yes	0%		
cgd3_370	hypothetical protein	Chro.40348	8	315 bp / 104 AAs	no	yes	0%		
cgd6_3710	40S ribosomal protein S30, transcripts identified be EST	Chro.40296	8	180 bp / 59 AAs	no	yes	0%		
cgd5_4060	protein with N-terminal region ZnF U1 domain, similar to U1 snRNP-specific protein	Chro.80265	8	597 bp / 198 AAs	no	yes	0%		
cgd2_140	similar to IMP4 family, transcript identified by EST	Chro.60522	8	321 bp / 106 AAs	no	yes	0%		
cgd3_3730	hypothetical protein	Chro.30098	8	369 bp / 122 AAs	no	yes	5%	5.6	C *parvum* Iowa II RAD1, NH2 terminal ERCC4-like helicase domain cgd4_2970
cgd4_4070	hypothetical protein	Chro.20236	9	315 bp / 104 AAs	no	yes	0%		
cgd4_4170	hypothetical protein	Chro.40476	9	462 bp / 153 AAs	no	yes	10%	9e-18	C *parvum* Iowa II hypothetical protein cgd4_4180
cgd2_3280	hypothetical protein	Chro.40348	9	375 bp / 124 AAs	no	yes	0%		

cgd4_1750	hypothetical protein	Chro.70464	9	333 bp / 110 AAs	no	yes	0%
cgd2_2870	similar to 40S ribosomal protein S28, no good start Met	Chro.10245	9	207 bp / 68 AAs	no	yes	0%
cgd6_3320	ypothetical protein (user's comment: partially encodes an MHC binding peptide)	Chro.10343	9	321 bp / 106 AAs	no	yes	0%
cgd2_350	60S ribosomal protein L39, transcripts identified be EST	Chro.80185	10	168 bp / 55 AAs	no	yes	0%
cgd3_3810	similar to clathrin adaptor complex, small subunit	Chro.40020	10	420 bp / 139 AAs	no	yes	0%
cgd1_1770	domain similar to KOG3415, putative Rab5-interacting protein	Chro.70125	10	288 bp / 95 AAs	no	yes	0%

Appendix IV

Appendix IV: Alignment of PCR product sequences of *Cryptosporidium* DNA from clinical isolates and reference strains amplified using newly designed primers. Ch2, Ch3, Ch4, TU502 are *C. hominis* isolates. Cp2, Cp3, Cp4, Iowa, Moredun are *C. parvum* isolates. W65, W66, W67, W70 are anthroponotic *C. parvum* isolates. Rabbit genotype isolates are labelled rabbit 1,2,3 and 4.

Cgd2_80 gene PCR products

```
                                1                                                                 60
          ch2_cgd2_80     (1) TGATATGTTGATTATAGATACATGTATTGTGAAGTCAATTGTTTCCAGTTTAGCACCAGT
          ch3_cgd2_80     (1) TGATATGTTGATTATAGATACATGTATTGTGAAGTCAATTGTTTCCAGTTTAGCACCAGT
          ch4_cgd2_80     (1) TGATATGTTGATTATAGATACATGTATTGTGAAGTCAATTGTTTCCAGTTTAGCACCAGT
        TU502_cgd2_80     (1) TGATATGTTGATTATAGATACATGTATTGTGAAGTCAATTGTTTCCAGTTTAGCACCAGT
      rabbit1_cgd2_80     (1) TGATATGTTGATTATAGATACATGTATTGTGAAATCAATTGTTTCCAGTTTAGCACCAGT
      rabbit2_cgd2_80     (1) TGATATGTTGATTATAGATACATGTATTGTGAAATCAATTGTTTCCAGTTTAGCACCAGT
      rabbit3_cgd2_80     (1) TGATATGTTGATTATAGATACATGTATTGTGAAATCAATTGTTTCCAGTTTAGCACCAGT
      rabbit4_cgd2_80     (1) TGATATGTTGATTATAGATACATGTATTGTGAAATCAATTGTTTCCAGTTTAGCACCAGT
          cp2_cgd2_80     (1) TGATATGTTGATTATAGATACATGTATTGTGAAGTCAATTATTTCCAGTTTAGCACCAGT
          cp3_cgd2_80     (1) TGATATGTTGATTATAGATACATGTATTGTGAAGTCAATTATTTCCAGTTTAGCACCAGT
          cp4_cgd2_80     (1) TGATATGTTGATTATAGATACATGTATTGTGAAGTCAATTATTTCCAGTTTAGCACCAGT
         Iowa_cgd2_80     (1) TGATATGTTGATTATAGATACATGTATTGTGAAGTCAATTATTTCCAGTTTAGCACCAGT
      Moredun_cgd2_80     (1) TGATATGTTGATTATAGATACATGTATTGTGAAGTCAATTATTTCCAGTTTAGCACCAGT
          w65_cgd2_80     (1) TGATATGTTGATTATAGATACATGTATTGTGAAGTCAATTATTTCCAGTTTAGCACCAGT
          w66_cgd2_80     (1) TGATATGTTGATTATAGATACATGTATTGTGAAGTCAATTATTTCCAGTTTAGCACCAGT
          w67_cgd2_80     (1) TGATATGTTGATTATAGATACATGTATTGTGAAGTCAATTATTTCCAGTTTAGCACCAGT
          w70_cgd2_80     (1) TGATATGTTGATTATAGATACATGTATTGTGAAGTCAATTATTTCCAGTTTAGCACCAGT
  c.meleagridis_cgd2_80   (1) TGATATGTTGATTATAGATACATGTATTGTGAAGTCAATTATTTCCAGTTTAGCACCAGT
                               61                                                                120
          ch2_cgd2_80    (61) ATTGACAATCATAGGTCAGACAATATTTATTGTATATACATTTCCTTATTTTACGCCATT
          ch3_cgd2_80    (61) ATTGACAATCATAGGTCAGACAATATTTATTGTATATACATTTCCTTATTTTACGCCATT
          ch4_cgd2_80    (61) ATTGACAATCATAGGTCAGACAATATTTATTGTATATACATTTCCTTATTTTACGCCATT
        TU502_cgd2_80    (61) ATTGACAATCATAGGTCAGACAATATTTATTGTATATACATTTCCTTATTTTACGCCATT
      rabbit1_cgd2_80    (61) ATTGACAATCATAGGTCAGACAATATTTATTGTATATACATTTCCTTATTTTACGCCATT
      rabbit2_cgd2_80    (61) ATTGACAATCATAGGTCAGACAATATTTATTGTATATACATTTCCTTATTTTACGCCATT
      rabbit3_cgd2_80    (61) ATTGACAATCATAGGTCAGACAATATTTATTGTATATACATTTCCTTATTTTACGCCATT
      rabbit4_cgd2_80    (61) ATTGACAATCATAGGTCAGACAATATTTATTGTATATACATTTCCTTATTTTACGCCATT
          cp2_cgd2_80    (61) ATTGACAATCATAGGTCAGACAATATTTATTGTATATACATTTCCTTATTTTACACCATT
          cp3_cgd2_80    (61) ATTGACAATCATAGGTCAGACAATATTTATTGTATATACATTTCCTTATTTTACACCATT
          cp4_cgd2_80    (61) ATTGACAATCATAGGTCAGACAATATTTATTGTATATACATTTCCTTATTTTACACCATT
         Iowa_cgd2_80    (61) ATTGACAATCATAGGTCAGACAATATTTATTGTATATACATTTCCTTATTTTACACCATT
      Moredun_cgd2_80    (61) ATTGACAATCATAGGTCAGACAATATTTATTGTATATACATTTCCTTATTTTACACCATT
          w65_cgd2_80    (61) ATTGACAATCATAGGTCAGACAATATTTATTGTATATACATTTCCTTATTTTACACCATT
          w66_cgd2_80    (61) ATTGACAATCATAGGTCAGACAATATTTATTGTATATACATTTCCTTATTTTACACCATT
          w67_cgd2_80    (61) ATTGACAATCATAGGTCAGACAATATTTATTGTATATACATTTCCTTATTTTACACCATT
          w70_cgd2_80    (61) ATTGACAATCATAGGTCAGACAATATTTATTGTATATACATTTCCTTATTTTACACCATT
  c.meleagridis_cgd2_80  (61) ATTGACAATCATAGGTCAGACAATATTTATTGTATATACATTTCCTTATTTTACACCATT
                               121                                                               180
          ch2_cgd2_80   (121) TTTTCTTATATGGATAATGCTTATTATAAAACCAATATGTCTCAAATTTATTTCATCTTA
          ch3_cgd2_80   (121) TTTTCTTATATGGATAATGCTTATTATAAAACCAATATGTCTCAAATTTATTTCATCTTA
          ch4_cgd2_80   (121) TTTTCTTATATGGATAATGCTTATTATAAAACCAATATGTCTCAAATTTATTTCATCTTA
        TU502_cgd2_80   (121) TTTTCTTATATGGATAATGCTTATTATAAAACCAATATGTCTCAAATTTATTTCATCTTA
      rabbit1_cgd2_80   (121) TTTTCTTATATGGATAATGCTTATTATAAAACCAATATGTCTCAAATTTATTTCATCTTA
      rabbit2_cgd2_80   (121) TTTTCTTATATGGATAATGCTTATTATAAAACCAATATGTCTCAAATTTATTTCATCTTA
      rabbit3_cgd2_80   (121) TTTTCTTATATGGATAATGCTTATTATAAAACCAATATGTCTCAAATTTATTTCATCTTA
      rabbit4_cgd2_80   (121) TTTTCTTATATGGATAATGCTTATTATAAAACCAATATGTCTCAAATTTATTTCATCTTA
          cp2_cgd2_80   (121) TTTTCTTATATGGATAATGCTTATTATAAAGCCGATATGCCTTAAATTTATTTCATCTTA
          cp3_cgd2_80   (121) TTTTCTTATATGGATAATGCTTATTATAAAGCCGATATGCCTTAAATTTATTTCATCTTA
          cp4_cgd2_80   (121) TTTTCTTATATGGATAATGCTTATTATAAAGCCGATATGCCTTAAATTTATTTCATCTTA
         Iowa_cgd2_80   (121) TTTTCTTATATGGATAATGCTTATTATAAAGCCGATATGCCTTAAATTTATTTCATCTTA
      Moredun_cgd2_80   (121) TTTTCTTATATGGATAATGCTTATTATAAAGCCGATATGCCTTAAATTTATTTCATCTTA
          w65_cgd2_80   (121) TTTTCTTATATGGATAATGCTTATTATAAAGCCGATATGCCTTAAATTTATTTCATCTTA
          w66_cgd2_80   (121) TTTTCTTATATGGATAATGCTTATTATAAAGCCGATATGCCTTAAATTTATTTCATCTTA
          w67_cgd2_80   (121) TTTTCTTATATGGATAATGCTTATTATAAAGCCGATATGCCTTAAATTTATTTCATCTTA
          w70_cgd2_80   (121) TTTTCTTATATGGATAATGCTTATTATAAAGCCGATATGCCTTAAATTTATTTCATCTTA
  c.meleagridis_cgd2_80 (121) TTTTCTTATATGGATAATGCTTATTATAAAGCCGATATGCCTTAAATTTATTTCATCTTA
                               181                                                               240
          ch2_cgd2_80   (181) TAGGGAATATCAGAGGTTTTCAATATCTCTTTTTTCCTCAATATGTGGAATATTTTCTGG
          ch3_cgd2_80   (181) TAGGGAATATCAGAGGTTTTCAATATCTCTTTTTTCCTCAATATGTGGAATATTTTCTGG
          ch4_cgd2_80   (181) TAGGGAATATCAGAGGTTTTCAATATCTCTTTTTTCCTCAATATGTGGAATATTTTCTGG
        TU502_cgd2_80   (181) TAGGGAATATCAGAGGTTTTCAATATCTCTTTTTTCCTCAATATGTGGAATATTTTCTGG
      rabbit1_cgd2_80   (181) TAGGGAATATCAGAGGTTTTCAATATCTCTTTTTTCCTCAATATGTGGAATATTTTCTGG
```

```
       rabbit2_cgd2_80  (181) TAGGGAATATCAGAGGTTTTCAATATCTCTTTTTTCCTCAATATGTGGAATATTTTCTGG
       rabbit3_cgd2_80  (181) TAGGGAATATCAGAGGTTTTCAATATCTCTTTTTTCCTCAATATGTGGAATATTTTCTGG
       rabbit4_cgd2_80  (181) TAGGGAATATCAGAGGTTTTCAATATCTCTTTTTTCCTCAATATGTGGAATATTTTCTGG
           cp2_cgd2_80  (181) TAGGGAATATCAGAGGTTTTCAATATCTCTTTTTTCCTCAATATGTGGAATATTTTCTGG
           cp3_cgd2_80  (181) TAGGGAATATCAGAGGTTTTCAATATCTCTTTTTTCCTCAATATGTGGAATATTTTCTGG
           cp4_cgd2_80  (181) TAGGGAATATCAGAGGTTTTCAATATCTCTTTTTTCCTCAATATGTGGAATATTTTCTGG
          Iowa_cgd2_80  (181) TAGGGAATATCAGAGGTTTTCAATATCTCTTTTTTCCTCAATATGTGGAATATTTTCTGG
       Moredun_cgd2_80  (181) TAGGGAATATCAGAGGTTTTCAATATCTCTTTTTTCCTCAATATGTGGAATATTTTCTGG
           w65_cgd2_80  (181) TAGGGAATATCAGAGGTTTTCAATATCTCTTTTTTCCTCAATATGTGGAATATTTTCTGG
           w66_cgd2_80  (181) TAGGGAATATCAGAGGTTTTCAATATCTCTTTTTTCCTCAATATGTGGAATATTTTCTGG
           w67_cgd2_80  (181) TAGGGAATATCAGAGGTTTTCAATATCTCTTTTTTCCTCAATATGTGGAATATTTTCTGG
           w70_cgd2_80  (181) TAGGGAATATCAGAGGTTTTCAATATCTCTTTTTTCCTCAATATGTGGAATATTTTCTGG
c.meleagridis_cgd2_80  (181) TAGGGAATATCAGAGGTTTGCAATATCTCCTTTTTCCTCAATATGTGGAATATTTTCTGG
                                241
           ch2_cgd2_80  (241) AACACAGC
           ch3_cgd2_80  (241) AACACAGC
           ch4_cgd2_80  (241) AACACAGC
         TU502_cgd2_80  (241) AACACAGC
       rabbit1_cgd2_80  (241) AACACAGC
       rabbit2_cgd2_80  (241) AACACAGC
       rabbit3_cgd2_80  (241) AACACAGC
       rabbit4_cgd2_80  (241) AACACAGC
           cp2_cgd2_80  (241) AACACAGC
           cp3_cgd2_80  (241) AACACAGC
           cp4_cgd2_80  (241) AACACAGC
          Iowa_cgd2_80  (241) AACACAGC
       Moredun_cgd2_80  (241) AACACAGC
           w65_cgd2_80  (241) AACACAGC
           w66_cgd2_80  (241) AACACAGC
           w67_cgd2_80  (241) AACACAGC
           w70_cgd2_80  (241) AACACAGC
c.meleagridis_cgd2_80  (241) AACACAGC
```

Cgd6_200 gene PCR products

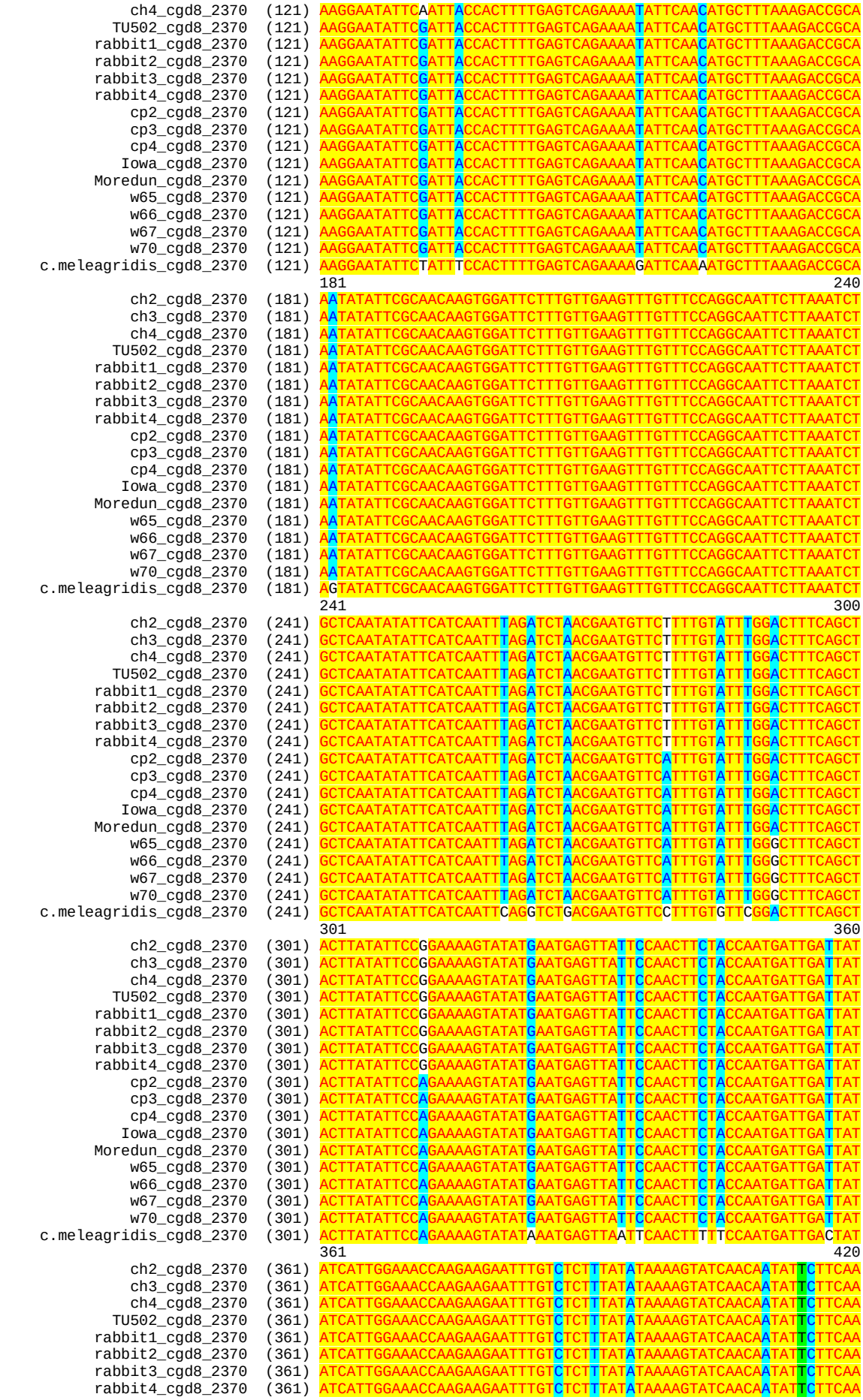
```
                                1                                                  60
           ch2_cgd6_200    (1) GGTGTGTCCAACAGGATATACTTTGGATTCAAATAGACAATGCGTTGCTAGAGAGGAGAT
           ch3_cgd6_200    (1) GGTGTGTCCAACAGGATATACTTTGGATTCAAATAGACAATGCGTTGCTAGAGAGGAGAT
           ch4_cgd6_200    (1) GGTGTGTCCAACAGGATATACTTTGGATTCAAATAGACAATGCGTTGCTAGAGAGGAGAT
         TU502_cgd6_200    (1) GGTGTGTCCAACAGGATATACTTTGGATTCAAATAGACAATGCGTTGCTAGAGAGGAGAT
       rabbit1_cgd6_200    (1) GGTGTGTCCAACAGGATATACTTTGGATTCAAATAGACAATGCGTTGCTAGAGAGGAGAT
       rabbit2_cgd6_200    (1) GGTGTGTCCAACAGGATATACTTTGGATTCAAATAGACAATGCGTTGCTAGAGAGGAGAT
       rabbit3_cgd6_200    (1) GGTGTGTCCAACAGGATATACTTTGGATTCAAATAGACAATGCGTTGCTAGAGAGGAGAT
       rabbit4_cgd6_200    (1) GGTGTGTCCAACAGGATATACTTTGGATTCAAATAGACAATGCGTTGCTAGAGAGGAGAT
           cp2_cgd6_200    (1) GGTGTGTCCAACAGGATATACTTTGGATTCAAATAGACAATGCGTTGCTAGAGAGGAGAT
           cp3_cgd6_200    (1) GGTGTGTCCAACAGGATATACTTTGGATTCAAATAGACAATGCGTTGCTAGAGAGGAGAT
           cp4_cgd6_200    (1) GGTGTGTCCAACAGGATATACTTTGGATTCAAATAGACAATGCGTTGCTAGAGAGGAGAT
          Iowa_cgd6_200    (1) GGTGTGTCCAACAGGATATACTTTGGATTCAAATAGACAATGCGTTGCTAGAGAGGAGAT
       Moredun_cgd6_200    (1) GGTGTGTCCAACAGGATATACTTTGGATTCAAATAGACAATGCGTTGCTAGAGAGGAGAT
           w65_cgd6_200    (1) GGTGTGTCCAACAGGATATACTTTGGATTCAAATAGACAATGCGTTGCTAGAGAGGAGAT
           w66_cgd6_200    (1) GGTGTGTCCAACAGGATATACTTTGGATTCAAATAGACAATGCGTTGCTAGAGAGGAGAT
           w67_cgd6_200    (1) GGTGTGTCCAACAGGATATACTTTGGATTCAAATAGACAATGCGTTGCTAGAGAGGAGAT
           w70_cgd6_200    (1) GGTGTGTCCAACAGGATATACTTTGGATTCAAATAGACAATGCGTTGCTAGAGAGGAGAT
C. meleagridis_cgd6_200    (1) GGTGTGTCCAACAGGATATACTTTGGATTCAAATAGGCAATGCGTTGCTAGAGAAGAGAT
                               61                                                 120
           ch2_cgd6_200   (61) AATGCCAGAAAGAGTCTGCTTGAATGGAGGAGAATTAATGAGTGATTTGAACTGTATGAA
           ch3_cgd6_200   (61) AATGCCAGAAAGAGTCTGCTTGAATGGAGGAGAATTAATGAGTGATTTGAACTGTATGAA
           ch4_cgd6_200   (61) AATGCCAGAAAGAGTCTGCTTGAATGGAGGAGAATTAATGAGTGATTTGAACTGTATGAA
         TU502_cgd6_200   (61) AATGCCAGAAAGAGTCTGCTTGAATGGAGGAGAATTAATGAGTGATTTGAACTGTATGAA
       rabbit1_cgd6_200   (61) AATGCCAGAAAGAGTCTGCTTGAATGGAGGAGAATTAATGAGTGATTTGAACTGTATGAA
       rabbit2_cgd6_200   (61) AATGCCAGAAAGAGTCTGCTTGAATGGAGGAGAATTAATGAGTGATTTGAACTGTATGAA
       rabbit3_cgd6_200   (61) AATGCCAGAAAGAGTCTGCTTGAATGGAGGAGAATTAATGAGTGATTTGAACTGTATGAA
       rabbit4_cgd6_200   (61) AATGCCAGAAAGAGTCTGCTTGAATGGAGGAGAATTAATGAGTGATTTGAACTGTATGAA
           cp2_cgd6_200   (61) AATGCCAGAAAGAGTCTGCTTGAATGGAGGAGAATTAATGAGCGATTTTAATTGTATGAA
           cp3_cgd6_200   (61) AATGCCAGAAAGAGTCTGCTTGAATGGAGGAGAATTAATGAGCGATTTTAATTGTATGAA
           cp4_cgd6_200   (61) AATGCCAGAAAGAGTCTGCTTGAATGGAGGAGAATTAATGAGCGATTTTAATTGTATGAA
          Iowa_cgd6_200   (61) AATGCCAGAAAGAGTCTGCTTGAATGGAGGAGAATTAATGAGCGATTTTAATTGTATGAA
       Moredun_cgd6_200   (61) AATGCCAGAAAGAGTCTGCTTGAATGGAGGAGAATTAATGAGCGATTTTAATTGTATGAA
```

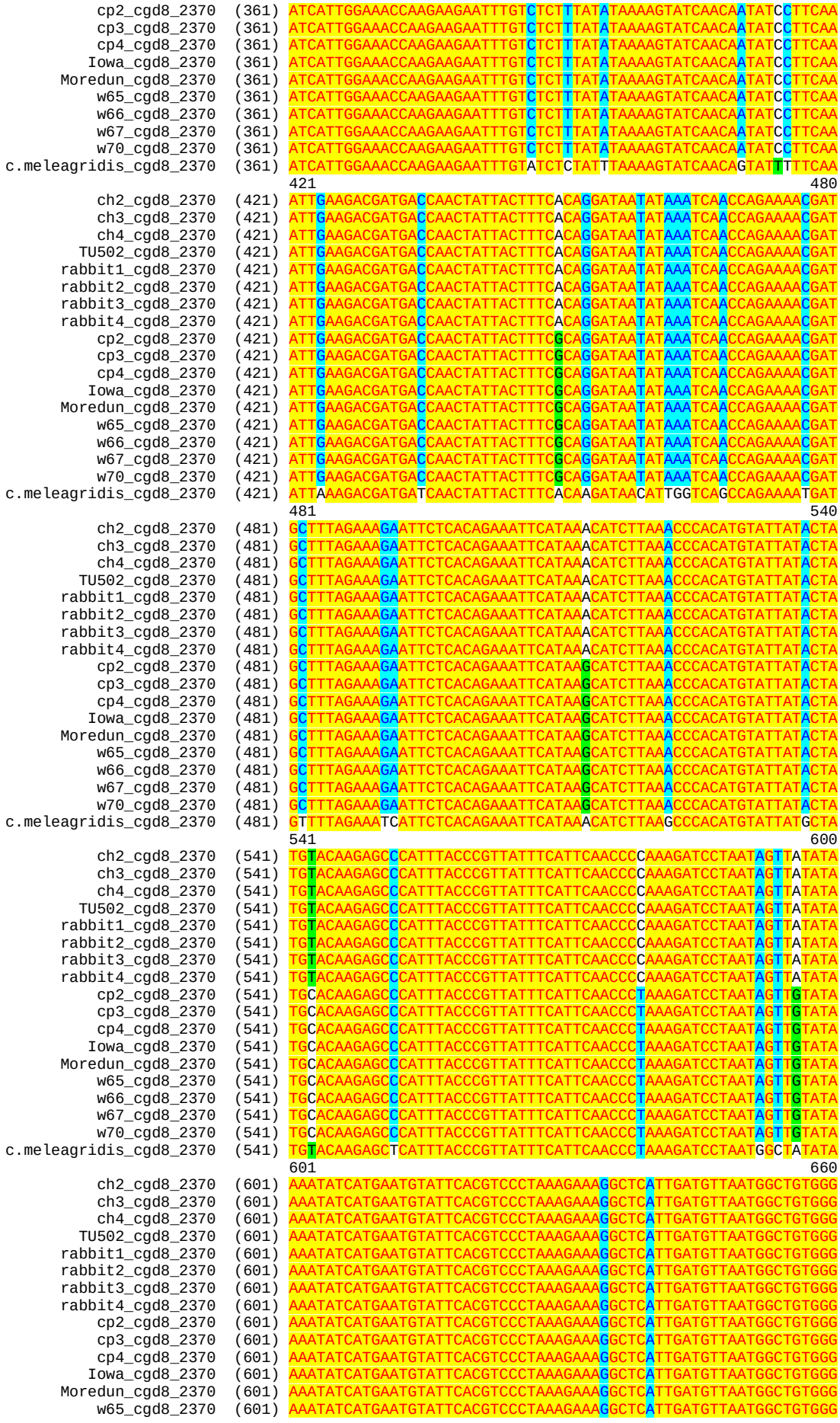
```
          w65_cgd6_200   (61) AATGCCAGAAAGAGTCTGCTTGAATGGAGGAGAATTAATGAGTGATTTTAATTGTATGAA
          w66_cgd6_200   (61) AATGCCAGAAAGAGTCTGCTTGAATGGAGGAGAATTAATGAGTGATTTTAATTGTATGAA
          w67_cgd6_200   (61) AATGCCAGAAAGAGTCTGCTTGAATGGAGGAGAATTAATGAGTGATTTTAATTGTATGAA
          w70_cgd6_200   (61) AATGCCAGAAAGAGTCTGCTTGAATGGAGGAGAATTAATGAGTGATTTTAATTGTATGAA
C. meleagridis_cgd6_200  (61) AATGCCAGAAAGAGTCTGCTTGAATGGAGGAGAATTAATGAATGATTTTAATTGTTTGAA
                              121                                                       180
          ch2_cgd6_200  (121) GACAACAAACCCAATCATGAAATGTCCAGTGGAATATACATTGGTAGGTGAAAATATGTG
          ch3_cgd6_200  (121) GACAACAAACCCAATCATGAAATGTCCAGTGGAATATACATTGGTAGGTGAAAATATGTG
          ch4_cgd6_200  (121) GACAACAAACCCAATCATGAAATGTCCAGTGGAATATACATTGGTAGGTGAAAATATGTG
        TU502_cgd6_200  (121) GACAACAAACCCAATCATGAAATGTCCAGTGGAATATACATTGGTAGGTGAAAATATGTG
      rabbit1_cgd6_200  (121) GACAACAAACCCAATCATGAAATGTCCAGTGGAATATACATTGGTAGGTGAAAATATGTG
      rabbit2_cgd6_200  (121) GACAACAAACCCAATCATGAAATGTCCAGTGGAATATACATTGGTAGGTGAAAATATGTG
      rabbit3_cgd6_200  (121) GACAACAAACCCAATCATGAAATGTCCAGTGGAATATACATTGGTAGGTGAAAATATGTG
      rabbit4_cgd6_200  (121) GACAACAAACCCAATCATGAAATGTCCAGTGGAATATACATTGGTAGGTGAAAATATGTG
          cp2_cgd6_200  (121) GACAACAAACCCAATCATGAAATGTCCAGTGGAATATACATTGGTAGGTGAAAATATGTG
          cp3_cgd6_200  (121) GACAACAAACCCAATCATGAAATGTCCAGTGGAATATACATTGGTAGGTGAAAATATGTG
          cp4_cgd6_200  (121) GACAACAAACCCAATCATGAAATGTCCAGTGGAATATACATTGGTAGGTGAAAATATGTG
         Iowa_cgd6_200  (121) GACAACAAACCCAATCATGAAATGTCCAGTGGAATATACATTGGTAGGTGAAAATATGTG
      Moredun_cgd6_200  (121) GACAACAAACCCAATCATGAAATGTCCAGTGGAATATACATTGGTAGGTGAAAATATGTG
          w65_cgd6_200  (121) GACAACAAACCCAATCATGAAATGTCCAGTGGAATATACATTGGTAGGTGAAAATATGTG
          w66_cgd6_200  (121) GACAACAAACCCAATCATGAAATGTCCAGTGGAATATACATTGGTAGGTGAAAATATGTG
          w67_cgd6_200  (121) GACAACAAACCCAATCATGAAATGTCCAGTGGAATATACATTGGTAGGTGAAAATATGTG
          w70_cgd6_200  (121) GACAACAAACCCAATCATGAAATGTCCAGTGGAATATACATTGGTAGGTGAAAATATGTG
C. meleagridis_cgd6_200  (121) GACAACAAACCCAATAATGAAATGCCCAGTAGAATATACATTGGTAGGTGAAAATATGTG
                              181                                                       240
          ch2_cgd6_200  (181) TCAAAAGAATGTTGAGATCGATCCAGTTGCTGTATGCCCATCAGGATTTACTTTAACTGA
          ch3_cgd6_200  (181) TCAAAAGAATGTTGAGATCGATCCAGTTGCTGTATGCCCATCAGGATTTACTTTAACTGA
          ch4_cgd6_200  (181) TCAAAAGAATGTTGAGATCGATCCAGTTGCTGTATGCCCATCAGGATTTACTTTAACTGA
        TU502_cgd6_200  (181) TCAAAAGAATGTTGAGATCGATCCAGTTGCTGTATGCCCATCAGGATTTACTTTAACTGA
      rabbit1_cgd6_200  (181) TCAAAAGAATGTTGAGATCGATCCAGTTGCTGTATGCCCATCAGGATTTACTTTAACTGA
      rabbit2_cgd6_200  (181) TCAAAAGAATGTTGAGATCGATCCAGTTGCTGTATGCCCATCAGGATTTACTTTAACTGA
      rabbit3_cgd6_200  (181) TCAAAAGAATGTTGAGATCGATCCAGTTGCTGTATGCCCATCAGGATTTACTTTAACTGA
      rabbit4_cgd6_200  (181) TCAAAAGAATGTTGAGATCGATCCAGTTGCTGTATGCCCATCAGGATTTACTTTAACTGA
          cp2_cgd6_200  (181) TCAAAAGAATGTTGAGATCGACCCAGTTGCTGTATGCCCATCAGGATTTACTTTAACTGA
          cp3_cgd6_200  (181) TCAAAAGAATGTTGAGATCGACCCAGTTGCTGTATGCCCATCAGGATTTACTTTAACTGA
          cp4_cgd6_200  (181) TCAAAAGAATGTTGAGATCGACCCAGTTGCTGTATGCCCATCAGGATTTACTTTAACTGA
         Iowa_cgd6_200  (181) TCAAAAGAATGTTGAGATCGACCCAGTTGCTGTATGCCCATCAGGATTTACTTTAACTGA
      Moredun_cgd6_200  (181) TCAAAAGAATGTTGAGATCGACCCAGTTGCTGTATGCCCATCAGGATTTACTTTAACTGA
          w65_cgd6_200  (181) TCAAAAGAATGTTGAGATCGACCCAGTTGCTGTATGCCCATCAGGATTTACTTTAACTGA
          w66_cgd6_200  (181) TCAAAAGAATGTTGAGATCGACCCAGTTGCTGTATGCCCATCAGGATTTACTTTAACTGA
          w67_cgd6_200  (181) TCAAAAGAATGTTGAGATCGACCCAGTTGCTGTATGCCCATCAGGATTTACTTTAACTGA
          w70_cgd6_200  (181) TCAAAAGAATGTTGAGATCGACCCAGTTGCTGTATGCCCATCAGGATTTACTTTAACTGA
C. meleagridis_cgd6_200  (181) TCAAAAGAATGTTGAGATCGATCCAGTTGCTGTTTGCCCATCAGGATTCACTTTAACTGA
                              241                                                       300
          ch2_cgd6_200  (241) CGGTCAAATGTGCAGCGGCTCAAAAACTGTTGCACCCATCAAGAAATGTATGCAAGGAGT
          ch3_cgd6_200  (241) CGGTCAAATGTGCAGCGGCTCAAAAACTGTTGCACCCATCAAGAAATGTATGCAAGGAGT
          ch4_cgd6_200  (241) CGGTCAAATGTGCAGCGGCTCAAAAACTGTTGCACCCATCAAGAAATGTATGCAAGGAGT
        TU502_cgd6_200  (241) CGGTCAAATGTGCAGCGGCTCAAAAACTGTTGCACCCATCAAGAAATGTATGCAAGGAGT
      rabbit1_cgd6_200  (241) CGGTCAAATGTGCAGTGGCTCAAAAACTGTTGCACCCATCAAGAAATGTATGCAAGGAGT
      rabbit2_cgd6_200  (241) CGGTCAAATGTGCAGTGGCTCAAAAACTGTTGCACCCATCAAGAAATGTATGCAAGGAGT
      rabbit3_cgd6_200  (241) CGGTCAAATGTGCAGTGGCTCAAAAACTGTTGCACCCATCAAGAAATGTATGCAAGGAGT
      rabbit4_cgd6_200  (241) CGGTCAAATGTGCAGTGGCTCAAAAACTGTTGCACCCATCAAGAAATGTATGCAAGGAGT
          cp2_cgd6_200  (241) CGGTCAAATGTGCAGTGGCTCAAAGACTGTTGCACCAATCAAGAAATGTATGCAAGGAGT
          cp3_cgd6_200  (241) CGGTCAAATGTGCAGTGGCTCAAAGACTGTTGCACCAATCAAGAAATGTATGCAAGGAGT
          cp4_cgd6_200  (241) CGGTCAAATGTGCAGTGGCTCAAAGACTGTTGCACCAATCAAGAAATGTATGCAAGGAGT
         Iowa_cgd6_200  (241) CGGTCAAATGTGCAGTGGCTCAAAGACTGTTGCACCAATCAAGAAATGTATGCAAGGAGT
      Moredun_cgd6_200  (241) CGGTCAAATGTGCAGTGGCTCAAAGACTGTTGCACCAATCAAGAAATGTATGCAAGGAGT
          w65_cgd6_200  (241) CGGTCAAATGTGCAGTGGCTCAAAGACTGTTGCACCAATCAAGAAATGTATGCAAGGAGT
          w66_cgd6_200  (241) CGGTCAAATGTGCAGTGGCTCAAAGACTGTTGCACCAATCAAGAAATGTATGCAAGGAGT
          w67_cgd6_200  (241) CGGTCAAATGTGCAGTGGCTCAAAGACTGTTGCACCAATCAAGAAATGTATGCAAGGAGT
          w70_cgd6_200  (241) CGGTCAAATGTGCAGTGGCTCAAAGACTGTTGCACCAATCAAGAAATGTATGCAAGGAGT
C. meleagridis_cgd6_200  (241) CGGTCAAATGTGCAGTGGTTCAAAGACTGTTGCACCAATCAAGAAATGTATGCAAGGAGT
                              301                                                       360
          ch2_cgd6_200  (301) TTTAAATGAAATGCAAACTGAATGTATCTTACAAAAGAGTGTATCCCCAATTAGTACCTG
          ch3_cgd6_200  (301) TTTAAATGAAATGCAAACTGAATGTATCTTACAAAAGAGTGTATCCCCAATTAGTACCTG
          ch4_cgd6_200  (301) TTTAAATGAAATGCAAACTGAATGTATCTTACAAAAGAGTGTATCCCCAATTAGTACCTG
        TU502_cgd6_200  (301) TTTAAATGAAATGCAAACTGAATGTATCTTACAAAAGAGTGTATCCCCAATTAGTACCTG
      rabbit1_cgd6_200  (301) TTTAAATGAAATGCAAACTGAATGTATCTTACAAAAGAGTGTATCCCCAATTAGTACCTG
      rabbit2_cgd6_200  (301) TTTAAATGAAATGCAAACTGAATGTATCTTACAAAAGAGTGTATCCCCAATTAGTACCTG
      rabbit3_cgd6_200  (301) TTTAAATGAAATGCAAACTGAATGTATCTTACAAAAGAGTGTATCCCCAATTAGTACCTG
      rabbit4_cgd6_200  (301) TTTAAATGAAATGCAAACTGAATGTATCTTACAAAAGAGTGTATCCCCAATTAGTACCTG
          cp2_cgd6_200  (301) TTTAAATGAAATGCAAAATGAATGTATCTTACAAAAGAGTGTATCCCCAATTAGTACCTG
          cp3_cgd6_200  (301) TTTAAATGAAATGCAAAATGAATGTATCTTACAAAAGAGTGTATCCCCAATTAGTACCTG
          cp4_cgd6_200  (301) TTTAAATGAAATGCAAAATGAATGTATCTTACAAAAGAGTGTATCCCCAATTAGTACCTG
         Iowa_cgd6_200  (301) TTTAAATGAAATGCAAAATGAATGTATCTTACAAAAGAGTGTATCCCCAATTAGTACCTG
      Moredun_cgd6_200  (301) TTTAAATGAAATGCAAAATGAATGTATCTTACAAAAGAGTGTATCCCCAATTAGTACCTG
          w65_cgd6_200  (301) TTTAAATGAAATGCAAAATGAATGTATCTTACAAAAGAGTGTATCCCCAATTAGTACCTG
          w66_cgd6_200  (301) TTTAAATGAAATGCAAAATGAATGTATCTTACAAAAGAGTGTATCCCCAATTAGTACCTG
          w67_cgd6_200  (301) TTTAAATGAAATGCAAAATGAATGTATCTTACAAAAGAGTGTATCCCCAATTAGTACCTG
          w70_cgd6_200  (301) TTTAAATGAAATGCAAAATGAATGTATCTTACAAAAGAGTGTATCCCCAATTAGTACCTG
C. meleagridis_cgd6_200  (301) CTTAAATGAAATGCAAACAGAATGCATCTTGCAAAAGAGTGTATCCCCAATTAGTACCTG
                              361                                                       420
```

```
            ch2_cgd6_200  (361) TCCATCACCTGATTTCACTTTAGTTAGTAATGAAAGATGTGTAAGAGAAGTTTTATATGA
            ch3_cgd6_200  (361) TCCATCACCTGATTTCACTTTAGTTAGTAATGAAAGATGTGTAAGAGAAGTTTTATATGA
            ch4_cgd6_200  (361) TCCATCACCTGATTTCACTTTAGTTAGTAATGAAAGATGTGTAAGAGAAGTTTTATATGA
          TU502_cgd6_200  (361) TCCATCACCTGATTTCACTTTAGTTAGTAATGAAAGATGTGTAAGAGAAGTTTTATATGA
        rabbit1_cgd6_200  (361) TCCATCACCTGATTTCACTTTAGTTAGTAATGAAAGATGTGTAAGAGAAGTTTTATATGA
        rabbit2_cgd6_200  (361) TCCATCACCTGATTTCACTTTAGTTAGTAATGAAAGATGTGTAAGAGAAGTTTTATATGA
        rabbit3_cgd6_200  (361) TCCATCACCTGATTTCACTTTAGTTAGTAATGAAAGATGTGTAAGAGAAGTTTTATATGA
        rabbit4_cgd6_200  (361) TCCATCACCTGATTTCACTTTAGTTAGTAATGAAAGATGTGTAAGAGAAGTTTTATATGA
            cp2_cgd6_200  (361) TCCATCACCTGATTTCACTTTAGTTAGTAATGAAAGATGTGTAAGAGAAGTTTTATATGA
            cp3_cgd6_200  (361) TCCATCACCTGATTTCACTTTAGTTAGTAATGAAAGATGTGTAAGAGAAGTTTTATATGA
            cp4_cgd6_200  (361) TCCATCACCTGATTTCACTTTAGTTAGTAATGAAAGATGTGTAAGAGAAGTTTTATATGA
           Iowa_cgd6_200  (361) TCCATCACCTGATTTCACTTTAGTTAGTAATGAAAGATGTGTAAGAGAAGTTTTATATGA
        Moredun_cgd6_200  (361) TCCATCACCTGATTTCACTTTAGTTAGTAATGAAAGATGTGTAAGAGAAGTTTTATATGA
            w65_cgd6_200  (361) TCCATCACCTGATTTCACTTTAGTTAGTAATGAAAGATGTGTAAGAGAAGTTTTATATGA
            w66_cgd6_200  (361) TCCATCACCTGATTTCACTTTAGTTAGTAATGAAAGATGTGTAAGAGAAGTTTTATATGA
            w67_cgd6_200  (361) TCCATCACCTGATTTCACTTTAGTTAGTAATGAAAGATGTGTAAGAGAAGTTTTATATGA
            w70_cgd6_200  (361) TCCATCACCTGATTTCACTTTAGTTAGTAATGAAAGATGTGTAAGAGAAGTTTTATATGA
C. meleagridis_cgd6_200  (361) TCCATCACCTGATTTCACTTTAGTTAGTAATGAAAGATGTGTAAGAGAAGTTTTATATGA
```

Cgd8_2370 gene PCR products

```
                                 1                                                  60
            ch2_cgd8_2370    (1) ACGGAAATTGGAATTGAATTTGAATTCCATATTACTAACAAGGCGAATCTGGAAACAGCA
            ch3_cgd8_2370    (1) ACGGAAATTGGAATTGAATTTGAATTCCATATTACTAACAAGGCGAATCTGGAAACAGCA
            ch4_cgd8_2370    (1) ACGGAAATTGGAATTGAATTTGAATTCCATATTACTAACAAGGCGAATCTGGAAACAGCA
          TU502_cgd8_2370    (1) ACGGAAATTGGAATTGAATTTGAATTCCATATTACTAACAAGGCGAATCTGGAAACAGCA
        rabbit1_cgd8_2370    (1) ACGGAAATTGGAATTGAATTTGAATTCCATATTACTAACAAGGCGAATCTGGAAACAGCA
        rabbit2_cgd8_2370    (1) ACGGAAATTGGAATTGAATTTGAATTCCATATTACTAACAAGGCGAATCTGGAAACAGCA
        rabbit3_cgd8_2370    (1) ACGGAAATTGGAATTGAATTTGAATTCCATATTACTAACAAGGCGAATCTGGAAACAGCA
        rabbit4_cgd8_2370    (1) ACGGAAATTGGAATTGAATTTGAATTCCATATTACTAACAAGGCGAATCTGGAAACAGCA
            cp2_cgd8_2370    (1) ACGGAAATTGGAATTGAATTTGAATTCCATATTACTAACAAGGCGAATCTGGAAACAGCA
            cp3_cgd8_2370    (1) ACGGAAATTGGAATTGAATTTGAATTCCATATTACTAACAAGGCGAATCTGGAAACAGCA
            cp4_cgd8_2370    (1) ACGGAAATTGGAATTGAATTTGAATTCCATATTACTAACAAGGCGAATCTGGAAACAGCA
           Iowa_cgd8_2370    (1) ACGGAAATTGGAATTGAATTTGAATTCCATATTACTAACAAGGCGAATCTGGAAACAGCA
        Moredun_cgd8_2370    (1) ACGGAAATTGGAATTGAATTTGAATTCCATATTACTAACAAGGCGAATCTGGAAACAGCA
            w65_cgd8_2370    (1) ACGGAAATTGGAATTGAATTTGAATTCCATATTACTAACAAGGCGAATCTGGAAACAGCA
            w66_cgd8_2370    (1) ACGGAAATTGGAATTGAATTTGAATTCCATATTACTAACAAGGCGAATCTGGAAACAGCA
            w67_cgd8_2370    (1) ACGGAAATTGGAATTGAATTTGAATTCCATATTACTAACAAGGCGAATCTGGAAACAGCA
            w70_cgd8_2370    (1) ACGGAAATTGGAATTGAATTTGAATTCCATATTACTAACAAGGCGAATCTGGAAACAGCA
c.meleagridis_cgd8_2370    (1) ACGGAAATTGGAATTGAATTTGAGTTCCATATTGCTAAAAAGGCGAATCTGGAAACAGCA
                                61                                                 120
            ch2_cgd8_2370   (61) AAATGCGTAGTTTTTGTAACTGAAGAAGAGAGAACGCTTCTAGCAGGATTAGGAGCTGCT
            ch3_cgd8_2370   (61) AAATGCGTAGTTTTTGTAACTGAAGAAGAGAGAACGCTTCTAGCAGGATTAGGAGCTGCT
            ch4_cgd8_2370   (61) AAATGCGTAGTTTTTGTAACTGAAGAAGAGAGAACGCTTCTAGCAGGATTAGGAGCTGCT
          TU502_cgd8_2370   (61) AAATGCGTAGTTTTTGTAACTGAAGAAGAGAGAACGCTTCTAGCAGGATTAGGAGCTGCT
        rabbit1_cgd8_2370   (61) AAATGCGTAGTATTTGTAACTGAAGAAGAGAGAACGCTTCTCGCAGGATTAGGAGCTGCT
        rabbit2_cgd8_2370   (61) AAATGCGTAGTATTTGTAACTGAAGAAGAGAGAACGCTTCTCGCAGGATTAGGAGCTGCT
        rabbit3_cgd8_2370   (61) AAATGCGTAGTATTTGTAACTGAAGAAGAGAGAACGCTTCTCGCAGGATTAGGAGCTGCT
        rabbit4_cgd8_2370   (61) AAATGCGTAGTATTTGTAACTGAAGAAGAGAGAACGCTTCTCGCAGGATTAGGAGCTGCT
            cp2_cgd8_2370   (61) AAATGCGTAGTTTTTGTAACTGAAGAAGAGAGAACGCTTCTCGCAGGATTAGGAGCTGCT
            cp3_cgd8_2370   (61) AAATGCGTAGTTTTTGTAACTGAAGAAGAGAGAACGCTTCTCGCAGGATTAGGAGCTGCT
            cp4_cgd8_2370   (61) AAATGCGTAGTTTTTGTAACTGAAGAAGAGAGAACGCTTCTCGCAGGATTAGGAGCTGCT
           Iowa_cgd8_2370   (61) AAATGCGTAGTTTTTGTAACTGAAGAAGAGAGAACGCTTCTCGCAGGATTAGGAGCTGCT
        Moredun_cgd8_2370   (61) AAATGCGTAGTTTTTGTAACTGAAGAAGAGAGAACGCTTCTCGCAGGATTAGGAGCTGCT
            w65_cgd8_2370   (61) AAATGCGTAGTTTTTGTAACTGAAGAAGAGAGAACGCTTCTCGCAGGATTAGGAGCTGCT
            w66_cgd8_2370   (61) AAATGCGTAGTTTTTGTAACTGAAGAAGAGAGAACGCTTCTCGCAGGATTAGGAGCTGCT
            w67_cgd8_2370   (61) AAATGCGTAGTTTTTGTAACTGAAGAAGAGAGAACGCTTCTCGCAGGATTAGGAGCTGCT
            w70_cgd8_2370   (61) AAATGCGTAGTTTTTGTAACTGAAGAAGAGAGAACGCTTCTCGCAGGATTAGGAGCTGCT
c.meleagridis_cgd8_2370   (61) AAATGCGTAGTTTTTGTAACTGAAGAAGAGAGAACACTTCTCGCGGGATTAGGAGCTGCT
                               121                                                 180
            ch2_cgd8_2370  (121) AAGGAATATTCAATTACCACTTTTGAGTCAGAAAATATTCAACATGCTTTAAAGACCGCA
            ch3_cgd8_2370  (121) AAGGAATATTCAATTACCACTTTTGAGTCAGAAAATATTCAACATGCTTTAAAGACCGCA
```


ch4_cgd8_2370 (121) AAGGAATATTCAATTACCACTTTTGAGTCAGAAAATATTCAACATGCTTTAAAGACCGCA
TU502_cgd8_2370 (121) AAGGAATATTCGATTACCACTTTTGAGTCAGAAAATATTCAACATGCTTTAAAGACCGCA
rabbit1_cgd8_2370 (121) AAGGAATATTCGATTACCACTTTTGAGTCAGAAAATATTCAACATGCTTTAAAGACCGCA
rabbit2_cgd8_2370 (121) AAGGAATATTCGATTACCACTTTTGAGTCAGAAAATATTCAACATGCTTTAAAGACCGCA
rabbit3_cgd8_2370 (121) AAGGAATATTCGATTACCACTTTTGAGTCAGAAAATATTCAACATGCTTTAAAGACCGCA
rabbit4_cgd8_2370 (121) AAGGAATATTCGATTACCACTTTTGAGTCAGAAAATATTCAACATGCTTTAAAGACCGCA
cp2_cgd8_2370 (121) AAGGAATATTCGATTACCACTTTTGAGTCAGAAAATATTCAACATGCTTTAAAGACCGCA
cp3_cgd8_2370 (121) AAGGAATATTCGATTACCACTTTTGAGTCAGAAAATATTCAACATGCTTTAAAGACCGCA
cp4_cgd8_2370 (121) AAGGAATATTCGATTACCACTTTTGAGTCAGAAAATATTCAACATGCTTTAAAGACCGCA
Iowa_cgd8_2370 (121) AAGGAATATTCGATTACCACTTTTGAGTCAGAAAATATTCAACATGCTTTAAAGACCGCA
Moredun_cgd8_2370 (121) AAGGAATATTCGATTACCACTTTTGAGTCAGAAAATATTCAACATGCTTTAAAGACCGCA
w65_cgd8_2370 (121) AAGGAATATTCGATTACCACTTTTGAGTCAGAAAATATTCAACATGCTTTAAAGACCGCA
w66_cgd8_2370 (121) AAGGAATATTCGATTACCACTTTTGAGTCAGAAAATATTCAACATGCTTTAAAGACCGCA
w67_cgd8_2370 (121) AAGGAATATTCGATTACCACTTTTGAGTCAGAAAATATTCAACATGCTTTAAAGACCGCA
w70_cgd8_2370 (121) AAGGAATATTCGATTACCACTTTTGAGTCAGAAAATATTCAACATGCTTTAAAGACCGCA
c.meleagridis_cgd8_2370 (121) AAGGAATATTCTATTTCCACTTTTGAGTCAGAAAAGATTCAAAATGCTTTAAAGACCGCA
181 240
ch2_cgd8_2370 (181) AATATATTCGCAACAAGTGGATTCTTTGTTGAAGTTTGTTTCCAGGCAATTCTTAAATCT
ch3_cgd8_2370 (181) AATATATTCGCAACAAGTGGATTCTTTGTTGAAGTTTGTTTCCAGGCAATTCTTAAATCT
ch4_cgd8_2370 (181) AATATATTCGCAACAAGTGGATTCTTTGTTGAAGTTTGTTTCCAGGCAATTCTTAAATCT
TU502_cgd8_2370 (181) AATATATTCGCAACAAGTGGATTCTTTGTTGAAGTTTGTTTCCAGGCAATTCTTAAATCT
rabbit1_cgd8_2370 (181) AATATATTCGCAACAAGTGGATTCTTTGTTGAAGTTTGTTTCCAGGCAATTCTTAAATCT
rabbit2_cgd8_2370 (181) AATATATTCGCAACAAGTGGATTCTTTGTTGAAGTTTGTTTCCAGGCAATTCTTAAATCT
rabbit3_cgd8_2370 (181) AATATATTCGCAACAAGTGGATTCTTTGTTGAAGTTTGTTTCCAGGCAATTCTTAAATCT
rabbit4_cgd8_2370 (181) AATATATTCGCAACAAGTGGATTCTTTGTTGAAGTTTGTTTCCAGGCAATTCTTAAATCT
cp2_cgd8_2370 (181) AATATATTCGCAACAAGTGGATTCTTTGTTGAAGTTTGTTTCCAGGCAATTCTTAAATCT
cp3_cgd8_2370 (181) AATATATTCGCAACAAGTGGATTCTTTGTTGAAGTTTGTTTCCAGGCAATTCTTAAATCT
cp4_cgd8_2370 (181) AATATATTCGCAACAAGTGGATTCTTTGTTGAAGTTTGTTTCCAGGCAATTCTTAAATCT
Iowa_cgd8_2370 (181) AATATATTCGCAACAAGTGGATTCTTTGTTGAAGTTTGTTTCCAGGCAATTCTTAAATCT
Moredun_cgd8_2370 (181) AATATATTCGCAACAAGTGGATTCTTTGTTGAAGTTTGTTTCCAGGCAATTCTTAAATCT
w65_cgd8_2370 (181) AATATATTCGCAACAAGTGGATTCTTTGTTGAAGTTTGTTTCCAGGCAATTCTTAAATCT
w66_cgd8_2370 (181) AATATATTCGCAACAAGTGGATTCTTTGTTGAAGTTTGTTTCCAGGCAATTCTTAAATCT
w67_cgd8_2370 (181) AATATATTCGCAACAAGTGGATTCTTTGTTGAAGTTTGTTTCCAGGCAATTCTTAAATCT
w70_cgd8_2370 (181) AATATATTCGCAACAAGTGGATTCTTTGTTGAAGTTTGTTTCCAGGCAATTCTTAAATCT
c.meleagridis_cgd8_2370 (181) AGTATATTCGCAACAAGTGGATTCTTTGTTGAAGTTTGTTTCCAGGCAATTCTTAAATCT
241 300
ch2_cgd8_2370 (241) GCTCAATATATTCATCAATTTAGATCTAACGAATGTTCTTTTGTATTTGGACTTTCAGCT
ch3_cgd8_2370 (241) GCTCAATATATTCATCAATTTAGATCTAACGAATGTTCTTTTGTATTTGGACTTTCAGCT
ch4_cgd8_2370 (241) GCTCAATATATTCATCAATTTAGATCTAACGAATGTTCTTTTGTATTTGGACTTTCAGCT
TU502_cgd8_2370 (241) GCTCAATATATTCATCAATTTAGATCTAACGAATGTTCTTTTGTATTTGGACTTTCAGCT
rabbit1_cgd8_2370 (241) GCTCAATATATTCATCAATTTAGATCTAACGAATGTTCTTTTGTATTTGGACTTTCAGCT
rabbit2_cgd8_2370 (241) GCTCAATATATTCATCAATTTAGATCTAACGAATGTTCTTTTGTATTTGGACTTTCAGCT
rabbit3_cgd8_2370 (241) GCTCAATATATTCATCAATTTAGATCTAACGAATGTTCTTTTGTATTTGGACTTTCAGCT
rabbit4_cgd8_2370 (241) GCTCAATATATTCATCAATTTAGATCTAACGAATGTTCTTTTGTATTTGGACTTTCAGCT
cp2_cgd8_2370 (241) GCTCAATATATTCATCAATTTAGATCTAACGAATGTTCATTTGTATTTGGACTTTCAGCT
cp3_cgd8_2370 (241) GCTCAATATATTCATCAATTTAGATCTAACGAATGTTCATTTGTATTTGGACTTTCAGCT
cp4_cgd8_2370 (241) GCTCAATATATTCATCAATTTAGATCTAACGAATGTTCATTTGTATTTGGACTTTCAGCT
Iowa_cgd8_2370 (241) GCTCAATATATTCATCAATTTAGATCTAACGAATGTTCATTTGTATTTGGACTTTCAGCT
Moredun_cgd8_2370 (241) GCTCAATATATTCATCAATTTAGATCTAACGAATGTTCATTTGTATTTGGACTTTCAGCT
w65_cgd8_2370 (241) GCTCAATATATTCATCAATTTAGATCTAACGAATGTTCATTTGTATTTGGGCTTTCAGCT
w66_cgd8_2370 (241) GCTCAATATATTCATCAATTTAGATCTAACGAATGTTCATTTGTATTTGGGCTTTCAGCT
w67_cgd8_2370 (241) GCTCAATATATTCATCAATTTAGATCTAACGAATGTTCATTTGTATTTGGGCTTTCAGCT
w70_cgd8_2370 (241) GCTCAATATATTCATCAATTTAGATCTAACGAATGTTCATTTGTATTTGGGCTTTCAGCT
c.meleagridis_cgd8_2370 (241) GCTCAATATATTCATCAATTCAGGTCTGACGAATGTTCCTTTGTGTTCGGACTTTCAGCT
301 360
ch2_cgd8_2370 (301) ACTTATATTCCGGAAAAGTATATGAATGAGTTATTCCAACTTCTACCAATGATTGATTAT
ch3_cgd8_2370 (301) ACTTATATTCCGGAAAAGTATATGAATGAGTTATTCCAACTTCTACCAATGATTGATTAT
ch4_cgd8_2370 (301) ACTTATATTCCGGAAAAGTATATGAATGAGTTATTCCAACTTCTACCAATGATTGATTAT
TU502_cgd8_2370 (301) ACTTATATTCCGGAAAAGTATATGAATGAGTTATTCCAACTTCTACCAATGATTGATTAT
rabbit1_cgd8_2370 (301) ACTTATATTCCGGAAAAGTATATGAATGAGTTATTCCAACTTCTACCAATGATTGATTAT
rabbit2_cgd8_2370 (301) ACTTATATTCCGGAAAAGTATATGAATGAGTTATTCCAACTTCTACCAATGATTGATTAT
rabbit3_cgd8_2370 (301) ACTTATATTCCGGAAAAGTATATGAATGAGTTATTCCAACTTCTACCAATGATTGATTAT
rabbit4_cgd8_2370 (301) ACTTATATTCCGGAAAAGTATATGAATGAGTTATTCCAACTTCTACCAATGATTGATTAT
cp2_cgd8_2370 (301) ACTTATATTCCAGAAAAGTATATGAATGAGTTATTCCAACTTCTACCAATGATTGATTAT
cp3_cgd8_2370 (301) ACTTATATTCCAGAAAAGTATATGAATGAGTTATTCCAACTTCTACCAATGATTGATTAT
cp4_cgd8_2370 (301) ACTTATATTCCAGAAAAGTATATGAATGAGTTATTCCAACTTCTACCAATGATTGATTAT
Iowa_cgd8_2370 (301) ACTTATATTCCAGAAAAGTATATGAATGAGTTATTCCAACTTCTACCAATGATTGATTAT
Moredun_cgd8_2370 (301) ACTTATATTCCAGAAAAGTATATGAATGAGTTATTCCAACTTCTACCAATGATTGATTAT
w65_cgd8_2370 (301) ACTTATATTCCAGAAAAGTATATGAATGAGTTATTCCAACTTCTACCAATGATTGATTAT
w66_cgd8_2370 (301) ACTTATATTCCAGAAAAGTATATGAATGAGTTATTCCAACTTCTACCAATGATTGATTAT
w67_cgd8_2370 (301) ACTTATATTCCAGAAAAGTATATGAATGAGTTATTCCAACTTCTACCAATGATTGATTAT
w70_cgd8_2370 (301) ACTTATATTCCAGAAAAGTATATGAATGAGTTATTCCAACTTCTACCAATGATTGATTAT
c.meleagridis_cgd8_2370 (301) ACTTATATTCCAGAAAAGTATATAAATGAGTTAATTCAACTTTTTCCAATGATTGACTAT
361 420
ch2_cgd8_2370 (361) ATCATTGGAAACCAAGAAGAATTTGTCTCTTTATATAAAAGTATCAACAATATTCTTCAA
ch3_cgd8_2370 (361) ATCATTGGAAACCAAGAAGAATTTGTCTCTTTATATAAAAGTATCAACAATATTCTTCAA
ch4_cgd8_2370 (361) ATCATTGGAAACCAAGAAGAATTTGTCTCTTTATATAAAAGTATCAACAATATTCTTCAA
TU502_cgd8_2370 (361) ATCATTGGAAACCAAGAAGAATTTGTCTCTTTATATAAAAGTATCAACAATATTCTTCAA
rabbit1_cgd8_2370 (361) ATCATTGGAAACCAAGAAGAATTTGTCTCTTTATATAAAAGTATCAACAATATTCTTCAA
rabbit2_cgd8_2370 (361) ATCATTGGAAACCAAGAAGAATTTGTCTCTTTATATAAAAGTATCAACAATATTCTTCAA
rabbit3_cgd8_2370 (361) ATCATTGGAAACCAAGAAGAATTTGTCTCTTTATATAAAAGTATCAACAATATTCTTCAA
rabbit4_cgd8_2370 (361) ATCATTGGAAACCAAGAAGAATTTGTCTCTTTATATAAAAGTATCAACAATATTCTTCAA


cp2_cgd8_2370 (361) ATCATTGGAAACCAAGAAGAATTTGTCTCTTTATATAAAAGTATCAACAATATCCTTCAA
cp3_cgd8_2370 (361) ATCATTGGAAACCAAGAAGAATTTGTCTCTTTATATAAAAGTATCAACAATATCCTTCAA
cp4_cgd8_2370 (361) ATCATTGGAAACCAAGAAGAATTTGTCTCTTTATATAAAAGTATCAACAATATCCTTCAA
Iowa_cgd8_2370 (361) ATCATTGGAAACCAAGAAGAATTTGTCTCTTTATATAAAAGTATCAACAATATCCTTCAA
Moredun_cgd8_2370 (361) ATCATTGGAAACCAAGAAGAATTTGTCTCTTTATATAAAAGTATCAACAATATCCTTCAA
w65_cgd8_2370 (361) ATCATTGGAAACCAAGAAGAATTTGTCTCTTTATATAAAAGTATCAACAATATCCTTCAA
w66_cgd8_2370 (361) ATCATTGGAAACCAAGAAGAATTTGTCTCTTTATATAAAAGTATCAACAATATCCTTCAA
w67_cgd8_2370 (361) ATCATTGGAAACCAAGAAGAATTTGTCTCTTTATATAAAAGTATCAACAATATCCTTCAA
w70_cgd8_2370 (361) ATCATTGGAAACCAAGAAGAATTTGTCTCTTTATATAAAAGTATCAACAATATCCTTCAA
c.meleagridis_cgd8_2370 (361) ATCATTGGAAACCAAGAAGAATTTGTATCTCTATTTAAAAGTATCAACAGTATTTTCAA
421 480
ch2_cgd8_2370 (421) ATTGAAGACGATGACCAACTATTACTTTCACAGGATAATATAAATCAACCAGAAAACGAT
ch3_cgd8_2370 (421) ATTGAAGACGATGACCAACTATTACTTTCACAGGATAATATAAATCAACCAGAAAACGAT
ch4_cgd8_2370 (421) ATTGAAGACGATGACCAACTATTACTTTCACAGGATAATATAAATCAACCAGAAAACGAT
TU502_cgd8_2370 (421) ATTGAAGACGATGACCAACTATTACTTTCACAGGATAATATAAATCAACCAGAAAACGAT
rabbit1_cgd8_2370 (421) ATTGAAGACGATGACCAACTATTACTTTCACAGGATAATATAAATCAACCAGAAAACGAT
rabbit2_cgd8_2370 (421) ATTGAAGACGATGACCAACTATTACTTTCACAGGATAATATAAATCAACCAGAAAACGAT
rabbit3_cgd8_2370 (421) ATTGAAGACGATGACCAACTATTACTTTCACAGGATAATATAAATCAACCAGAAAACGAT
rabbit4_cgd8_2370 (421) ATTGAAGACGATGACCAACTATTACTTTCACAGGATAATATAAATCAACCAGAAAACGAT
cp2_cgd8_2370 (421) ATTGAAGACGATGACCAACTATTACTTTCGCAGGATAATATAAATCAACCAGAAAACGAT
cp3_cgd8_2370 (421) ATTGAAGACGATGACCAACTATTACTTTCGCAGGATAATATAAATCAACCAGAAAACGAT
cp4_cgd8_2370 (421) ATTGAAGACGATGACCAACTATTACTTTCGCAGGATAATATAAATCAACCAGAAAACGAT
Iowa_cgd8_2370 (421) ATTGAAGACGATGACCAACTATTACTTTCGCAGGATAATATAAATCAACCAGAAAACGAT
Moredun_cgd8_2370 (421) ATTGAAGACGATGACCAACTATTACTTTCGCAGGATAATATAAATCAACCAGAAAACGAT
w65_cgd8_2370 (421) ATTGAAGACGATGACCAACTATTACTTTCGCAGGATAATATAAATCAACCAGAAAACGAT
w66_cgd8_2370 (421) ATTGAAGACGATGACCAACTATTACTTTCGCAGGATAATATAAATCAACCAGAAAACGAT
w67_cgd8_2370 (421) ATTGAAGACGATGACCAACTATTACTTTCGCAGGATAATATAAATCAACCAGAAAACGAT
w70_cgd8_2370 (421) ATTGAAGACGATGACCAACTATTACTTTCGCAGGATAATATAAATCAACCAGAAAACGAT
c.meleagridis_cgd8_2370 (421) ATTAAAGACGATGATCAACTATTACTTTCACAAGATAACATTGGTCAGCCAGAAAATGAT
481 540
ch2_cgd8_2370 (481) GCTTTAGAAAGAATTCTCACAGAAATTCATAAACATCTTAAACCCACATGTATTATACTA
ch3_cgd8_2370 (481) GCTTTAGAAAGAATTCTCACAGAAATTCATAAACATCTTAAACCCACATGTATTATACTA
ch4_cgd8_2370 (481) GCTTTAGAAAGAATTCTCACAGAAATTCATAAACATCTTAAACCCACATGTATTATACTA
TU502_cgd8_2370 (481) GCTTTAGAAAGAATTCTCACAGAAATTCATAAACATCTTAAACCCACATGTATTATACTA
rabbit1_cgd8_2370 (481) GCTTTAGAAAGAATTCTCACAGAAATTCATAAACATCTTAAACCCACATGTATTATACTA
rabbit2_cgd8_2370 (481) GCTTTAGAAAGAATTCTCACAGAAATTCATAAACATCTTAAACCCACATGTATTATACTA
rabbit3_cgd8_2370 (481) GCTTTAGAAAGAATTCTCACAGAAATTCATAAACATCTTAAACCCACATGTATTATACTA
rabbit4_cgd8_2370 (481) GCTTTAGAAAGAATTCTCACAGAAATTCATAAACATCTTAAACCCACATGTATTATACTA
cp2_cgd8_2370 (481) GCTTTAGAAAGAATTCTCACAGAAATTCATAAGCATCTTAAACCCACATGTATTATACTA
cp3_cgd8_2370 (481) GCTTTAGAAAGAATTCTCACAGAAATTCATAAGCATCTTAAACCCACATGTATTATACTA
cp4_cgd8_2370 (481) GCTTTAGAAAGAATTCTCACAGAAATTCATAAGCATCTTAAACCCACATGTATTATACTA
Iowa_cgd8_2370 (481) GCTTTAGAAAGAATTCTCACAGAAATTCATAAGCATCTTAAACCCACATGTATTATACTA
Moredun_cgd8_2370 (481) GCTTTAGAAAGAATTCTCACAGAAATTCATAAGCATCTTAAACCCACATGTATTATACTA
w65_cgd8_2370 (481) GCTTTAGAAAGAATTCTCACAGAAATTCATAAGCATCTTAAACCCACATGTATTATACTA
w66_cgd8_2370 (481) GCTTTAGAAAGAATTCTCACAGAAATTCATAAGCATCTTAAACCCACATGTATTATACTA
w67_cgd8_2370 (481) GCTTTAGAAAGAATTCTCACAGAAATTCATAAGCATCTTAAACCCACATGTATTATACTA
w70_cgd8_2370 (481) GCTTTAGAAAGAATTCTCACAGAAATTCATAAGCATCTTAAACCCACATGTATTATACTA
c.meleagridis_cgd8_2370 (481) GTTTTAGAAATCATTCTCACAGAAATTCATAAACATCTTAAGCCCACATGTATTATGCTA
541 600
ch2_cgd8_2370 (541) TGTACAAGAGCCCATTTACCCGTTATTTCATTCAACCCCAAAGATCCTAATAGTTATATA
ch3_cgd8_2370 (541) TGTACAAGAGCCCATTTACCCGTTATTTCATTCAACCCCAAAGATCCTAATAGTTATATA
ch4_cgd8_2370 (541) TGTACAAGAGCCCATTTACCCGTTATTTCATTCAACCCCAAAGATCCTAATAGTTATATA
TU502_cgd8_2370 (541) TGTACAAGAGCCCATTTACCCGTTATTTCATTCAACCCCAAAGATCCTAATAGTTATATA
rabbit1_cgd8_2370 (541) TGTACAAGAGCCCATTTACCCGTTATTTCATTCAACCCCAAAGATCCTAATAGTTATATA
rabbit2_cgd8_2370 (541) TGTACAAGAGCCCATTTACCCGTTATTTCATTCAACCCCAAAGATCCTAATAGTTATATA
rabbit3_cgd8_2370 (541) TGTACAAGAGCCCATTTACCCGTTATTTCATTCAACCCCAAAGATCCTAATAGTTATATA
rabbit4_cgd8_2370 (541) TGTACAAGAGCCCATTTACCCGTTATTTCATTCAACCCCAAAGATCCTAATAGTTATATA
cp2_cgd8_2370 (541) TGCACAAGAGCCCATTTACCCGTTATTTCATTCAACCCTAAAGATCCTAATAGTTGTATA
cp3_cgd8_2370 (541) TGCACAAGAGCCCATTTACCCGTTATTTCATTCAACCCTAAAGATCCTAATAGTTGTATA
cp4_cgd8_2370 (541) TGCACAAGAGCCCATTTACCCGTTATTTCATTCAACCCTAAAGATCCTAATAGTTGTATA
Iowa_cgd8_2370 (541) TGCACAAGAGCCCATTTACCCGTTATTTCATTCAACCCTAAAGATCCTAATAGTTGTATA
Moredun_cgd8_2370 (541) TGCACAAGAGCCCATTTACCCGTTATTTCATTCAACCCTAAAGATCCTAATAGTTGTATA
w65_cgd8_2370 (541) TGCACAAGAGCCCATTTACCCGTTATTTCATTCAACCCTAAAGATCCTAATAGTTGTATA
w66_cgd8_2370 (541) TGCACAAGAGCCCATTTACCCGTTATTTCATTCAACCCTAAAGATCCTAATAGTTGTATA
w67_cgd8_2370 (541) TGCACAAGAGCCCATTTACCCGTTATTTCATTCAACCCTAAAGATCCTAATAGTTGTATA
w70_cgd8_2370 (541) TGCACAAGAGCCCATTTACCCGTTATTTCATTCAACCCTAAAGATCCTAATAGTTGTATA
c.meleagridis_cgd8_2370 (541) TGTACAAGAGCTCATTTACCCGTTATTTCATTCAACCCTAAAGATCCTAATGGCTATATA
601 660
ch2_cgd8_2370 (601) AAATATCATGAATGTATTCACGTCCCTAAAGAAAGGCTCATTGATGTTAATGGCTGTGGG
ch3_cgd8_2370 (601) AAATATCATGAATGTATTCACGTCCCTAAAGAAAGGCTCATTGATGTTAATGGCTGTGGG
ch4_cgd8_2370 (601) AAATATCATGAATGTATTCACGTCCCTAAAGAAAGGCTCATTGATGTTAATGGCTGTGGG
TU502_cgd8_2370 (601) AAATATCATGAATGTATTCACGTCCCTAAAGAAAGGCTCATTGATGTTAATGGCTGTGGG
rabbit1_cgd8_2370 (601) AAATATCATGAATGTATTCACGTCCCTAAAGAAAGGCTCATTGATGTTAATGGCTGTGGG
rabbit2_cgd8_2370 (601) AAATATCATGAATGTATTCACGTCCCTAAAGAAAGGCTCATTGATGTTAATGGCTGTGGG
rabbit3_cgd8_2370 (601) AAATATCATGAATGTATTCACGTCCCTAAAGAAAGGCTCATTGATGTTAATGGCTGTGGG
rabbit4_cgd8_2370 (601) AAATATCATGAATGTATTCACGTCCCTAAAGAAAGGCTCATTGATGTTAATGGCTGTGGG
cp2_cgd8_2370 (601) AAATATCATGAATGTATTCACGTCCCTAAAGAAAGGCTCATTGATGTTAATGGCTGTGGG
cp3_cgd8_2370 (601) AAATATCATGAATGTATTCACGTCCCTAAAGAAAGGCTCATTGATGTTAATGGCTGTGGG
cp4_cgd8_2370 (601) AAATATCATGAATGTATTCACGTCCCTAAAGAAAGGCTCATTGATGTTAATGGCTGTGGG
Iowa_cgd8_2370 (601) AAATATCATGAATGTATTCACGTCCCTAAAGAAAGGCTCATTGATGTTAATGGCTGTGGG
Moredun_cgd8_2370 (601) AAATATCATGAATGTATTCACGTCCCTAAAGAAAGGCTCATTGATGTTAATGGCTGTGGG
w65_cgd8_2370 (601) AAATATCATGAATGTATTCACGTCCCTAAAGAAAGGCTCATTGATGTTAATGGCTGTGGG

```
          w66_cgd8_2370  (601) AAATATCATGAATGTATTCACGTCCCTAAAGAAAGGCTCATTGATGTTAATGGCTGTGGG
          w67_cgd8_2370  (601) AAATATCATGAATGTATTCACGTCCCTAAAGAAAGGCTCATTGATGTTAATGGCTGTGGG
          w70_cgd8_2370  (601) AAATATCATGAATGTATTCACGTCCCTAAAGAAAGGCTCATTGATGTTAATGGCTGTGGG
c.meleagridis_cgd8_2370  (601) AAATATCATGAATGTATTCACGTCCCTAAAGAAAAGCTCGTTGATGTTAATGGCTGTGGG
                               661
          ch2_cgd8_2370  (661) G
          ch3_cgd8_2370  (661) G
          ch4_cgd8_2370  (661) G
        TU502_cgd8_2370  (661) G
      rabbit1_cgd8_2370  (661) G
      rabbit2_cgd8_2370  (661) G
      rabbit3_cgd8_2370  (661) G
      rabbit4_cgd8_2370  (661) G
          cp2_cgd8_2370  (661) G
          cp3_cgd8_2370  (661) G
          cp4_cgd8_2370  (661) G
         Iowa_cgd8_2370  (661) G
      Moredun_cgd8_2370  (661) G
          w65_cgd8_2370  (661) G
          w66_cgd8_2370  (661) G
          w67_cgd8_2370  (661) G
          w70_cgd8_2370  (661) G
c.meleagridis_cgd8_2370  (661) G
```

Cgd2_2430 gene PCR products

```
                              1                                                  60
    ch2_cgd2_2430     (1) ACCTATAATTGATCGAAAAAGTGTTTTCCAAGCTCATGCATGTAAGGTAGAAACAGTTGA
    ch3_cgd2_2430     (1) ACCTATAATTGATCGAAAAAGTGTTTTCCAAGCTCATGCATGTAAGGTAGAAACAGTTGA
    ch4_cgd2_2430     (1) ACCTATAATTGATCGAAAAAGTGTTTTCCAAGCTCATGCATGTAAGGTAGAAACAGTTGA
  TU502_cgd2_2430     (1) ACCTATAATTGATCGAAAAAGTGTTTTCCAAGCTCATGCATGTAAGGTAGAAACAGTTGA
rabbit1_cgd2_2430     (1) ACCTATAATTGATCGAAAAAGTGTTTTCCAAGCTCATGCGTGTAAGGTAGAAACAGTTGA
rabbit2_cgd2_2430     (1) ACCTATAATTGATCGAAAAAGTGTTTTCCAAGCTCATGCGTGTAAGGTAGAAACAGTTGA
rabbit3_cgd2_2430     (1) ACCTATAATTGATCGAAAAAGTGTTTTCCAAGCTCATGCGTGTAAGGTAGAAACAGTTGA
rabbit4_cgd2_2430     (1) ACCTATAATTGATCGAAAAAGTGTTTTCCAAGCTCATGCGTGTAAGGTAGAAACAGTTGA
    cp2_cgd2_2430     (1) ACCTATAATTGATCGAAAAAGTGTTTTCCAAGCTCATGCGTGTAAGGTAGAAACAGTTGA
    cp3_cgd2_2430     (1) ACCTATAATTGATCGAAAAAGTGTTTTCCAAGCTCATGCGTGTAAGGTAGAAACAGTTGA
    cp4_cgd2_2430     (1) ACCTATAATTGATCGAAAAAGTGTTTTCCAAGCTCATGCGTGTAAGGTAGAAACAGTTGA
   Iowa_cgd2_2430     (1) ACCTATAATTGATCGAAAAAGTGTTTTCCAAGCTCATGCGTGTAAGGTAGAAACAGTTGA
Moredun_cgd2_2430     (1) ACCTATAATTGATCGAAAAAGTGTTTTCCAAGCTCATGCGTGTAAGGTAGAAACAGTTGA
    w65_cgd2_2430     (1) ACCTATAATTGATCGAAAAAGTGTTTTCCAAGCTCATGCGTGTAAGGTAGAAACAGTTGA
    w66_cgd2_2430     (1) ACCTATAATTGATCGAAAAAGTGTTTTCCAAGCTCATGCGTGTAAGGTAGAAACAGTTGA
    w67_cgd2_2430     (1) ACCTATAATTGATCGAAAAAGTGTTTTCCAAGCTCATGCGTGTAAGGTAGAAACAGTTGA
    w70_cgd2_2430     (1) ACCTATAATTGATCGAAAAAGTGTTTTCCAAGCTCATGCGTGTAAGGTAGAAACAGTTGA
                              61                                                 120
    ch2_cgd2_2430    (61) ACAAGTTAAAAAGATTATTAAATGGTTGCTTTCAAACCCAAAGATTGCTAAAGCAACGCA
    ch3_cgd2_2430    (61) ACAAGTTAAAAAGATTATTAAATGGTTGCTTTCAAACCCAAAGATTGCTAAAGCAACGCA
    ch4_cgd2_2430    (61) ACAAGTTAAAAAGATTATTAAATGGTTGCTTTCAAACCCAAAGATTGCTAAAGCAACGCA
  TU502_cgd2_2430    (61) ACAAGTTAAAAAGATTATTAAATGGTTGCTTTCAAACCCAAAGATTGCTAAAGCAACGCA
rabbit1_cgd2_2430    (61) ACAAGTTAAAAAGATTATTAAATGGTTGCTTTCAAACCCAAAGATTGCTAAAGCAACGCA
rabbit2_cgd2_2430    (61) ACAAGTTAAAAAGATTATTAAATGGTTGCTTTCAAACCCAAAGATTGCTAAAGCAACGCA
rabbit3_cgd2_2430    (61) ACAAGTTAAAAAGATTATTAAATGGTTGCTTTCAAACCCAAAGATTGCTAAAGCAACGCA
rabbit4_cgd2_2430    (61) ACAAGTTAAAAAGATTATTAAATGGTTGCTTTCAAACCCAAAGATTGCTAAAGCAACGCA
    cp2_cgd2_2430    (61) ACAAGTTAAAAAAGTTATCAAATGGTTGCTTTCGAACCCAAAGATTGCTAAAGCAACGCA
    cp3_cgd2_2430    (61) ACAAGTTAAAAAAGTTATCAAATGGTTGCTTTCGAACCCAAAGATTGCTAAAGCAACGCA
    cp4_cgd2_2430    (61) ACAAGTTAAAAAAGTTATCAAATGGTTGCTTTCGAACCCAAAGATTGCTAAAGCAACGCA
   Iowa_cgd2_2430    (61) ACAAGTTAAAAAAGTTATCAAATGGTTGCTTTCGAACCCAAAGATTGCTAAAGCAACGCA
Moredun_cgd2_2430    (61) ACAAGTTAAAAAAGTTATCAAATGGTTGCTTTCGAACCCAAAGATTGCTAAAGCAACGCA
    w65_cgd2_2430    (61) ACAAGTTAAAAAAGTTATCAAATGGTTGCTTTCGAACCCAAAGATTGCTAAAGCAACGCA
    w66_cgd2_2430    (61) ACAAGTTAAAAAAGTTATCAAATGGTTGCTTTCGAACCCAAAGATTGCTAAAGCAACGCA
    w67_cgd2_2430    (61) ACAAGTTAAAAAAGTTATCAAATGGTTGCTTTCGAACCCAAAGATTGCTAAAGCAACGCA
    w70_cgd2_2430    (61) ACAAGTTAAAAAAGTTATCAAATGGTTGCTTTCGAACCCAAAGATTGCTAAAGCAACGCA
                             121                                                 180
    ch2_cgd2_2430   (121) CAATATTTGGTCATATAGAATATTCAAAGAGAAAAATTTCGCAGAAGTATCTGAAGGATC
    ch3_cgd2_2430   (121) CAATATTTGGTCATATAGAATATTCAAAGAGAAAAATTTCGCAGAAGTATCTGAAGGATC
    ch4_cgd2_2430   (121) CAATATTTGGTCATATAGAATATTCAAAGAGAAAAATTTCGCAGAAGTATCTGAAGGATC
  TU502_cgd2_2430   (121) CAATATTTGGTCATATAGAATATTCAAAGAGAAAAATTTCGCAGAAGTATCTGAAGGATC
```

```
rabbit1_cgd2_2430  (121) CAATATTTGGTCATATAGAATATTCAAGGAGAAAAATTTCGCAGAAGTATCTGAAGGATC
rabbit2_cgd2_2430  (121) CAATATTTGGTCATATAGAATATTCAAGGAGAAAAATTTCGCAGAAGTATCTGAAGGATC
rabbit3_cgd2_2430  (121) CAATATTTGGTCATATAGAATATTCAAGGAGAAAAATTTCGCAGAAGTATCTGAAGGATC
rabbit4_cgd2_2430  (121) CAATATTTGGTCATATAGAATATTCAAGGAGAAAAATTTCGCAGAAGTATCTGAAGGATC
    cp2_cgd2_2430  (121) CAATATTTGGTCATATAGAATATTCAAGGAGAAAAATTTCGCAGAAGTAACTGAAGGATC
    cp3_cgd2_2430  (121) CAATATTTGGTCATATAGAATATTCAAGGAGAAAAATTTCGCAGAAGTAACTGAAGGATC
    cp4_cgd2_2430  (121) CAATATTTGGTCATATAGAATATTCAAGGAGAAAAATTTCGCAGAAGTAACTGAAGGATC
   Iowa_cgd2_2430  (121) CAATATTTGGTCATATAGAATATTCAAGGAGAAAAATTTCGCAGAAGTAACTGAAGGATC
Moredun_cgd2_2430  (121) CAATATTTGGTCATATAGAATATTCAAGGAGAAAAATTTCGCAGAAGTAACTGAAGGATC
    w65_cgd2_2430  (121) CAATATTTGGTCATATAGAATATTCAAGGAGAAAAATTTCGCAGAAGTAACTGAAGGATC
    w66_cgd2_2430  (121) CAATATTTGGTCATATAGAATATTCAAGGAGAAAAATTTCGCAGAAGTAACTGAAGGATC
    w67_cgd2_2430  (121) CAATATTTGGTCATATAGAATATTCAAGGAGAAAAATTTCGCAGAAGTAACTGAAGGATC
    w70_cgd2_2430  (121) CAATATTTGGTCATATAGAATATTCAAGGAGAAAAATTTCGCAGAAGTAACTGAAGGATC
                         181                                                   240
    ch2_cgd2_2430  (181) TTTCCCCATTGGGTACGATATCATATCACAAGATCATGACTCGGATGGCGAGAATGCAGC
    ch3_cgd2_2430  (181) TTTCCCCATTGGGTACGATATCATATCACAAGATCATGACTCGGATGGCGAGAATGCAGC
    ch4_cgd2_2430  (181) TTTCCCCATTGGGTACGATATCATATCACAAGATCATGACTCGGATGGCGAGAATGCAGC
  TU502_cgd2_2430  (181) TTTCCCCATTGGGTACGATATCATATCACAAGATCATGACTCGGATGGCGAGAATGCAGC
rabbit1_cgd2_2430  (181) TTTCCCCATCGGGTACGATATCATATCACAAGATCATGACTCGGATGGCGAGAATGCAGC
rabbit2_cgd2_2430  (181) TTTCCCCATCGGGTACGATATCATATCACAAGATCATGACTCGGATGGCGAGAATGCAGC
rabbit3_cgd2_2430  (181) TTTCCCCATCGGGTACGATATCATATCACAAGATCATGACTCGGATGGCGAGAATGCAGC
rabbit4_cgd2_2430  (181) TTTCCCCATCGGGTACGATATCATATCACAAGATCATGACTCGGATGGCGAGAATGCAGC
    cp2_cgd2_2430  (181) TTTCCCCATTGGGTACGATATCATATCACAAGATCACGACTCGGATGGCGAGAATGCAGC
    cp3_cgd2_2430  (181) TTTCCCCATTGGGTACGATATCATATCACAAGATCACGACTCGGATGGCGAGAATGCAGC
    cp4_cgd2_2430  (181) TTTCCCCATTGGGTACGATATCATATCACAAGATCACGACTCGGATGGCGAGAATGCAGC
   Iowa_cgd2_2430  (181) TTTCCCCATTGGGTACGATATCATATCACAAGATCACGACTCGGATGGCGAGAATGCAGC
Moredun_cgd2_2430  (181) TTTCCCCATTGGGTACGATATCATATCACAAGATCACGACTCGGATGGCGAGAATGCAGC
    w65_cgd2_2430  (181) TTTCCCCATTGGGTACGATATCATATCACAAGATCACGACTCGGATGGCGAGAATGCAGC
    w66_cgd2_2430  (181) TTTCCCCATTGGGTACGATATCATATCACAAGATCACGACTCGGATGGCGAGAATGCAGC
    w67_cgd2_2430  (181) TTTCCCCATTGGGTACGATATCATATCACAAGATCACGACTCGGATGGCGAGAATGCAGC
    w70_cgd2_2430  (181) TTTCCCCATTGGGTACGATATCATATCACAAGATCACGACTCGGATGGCGAGAATGCAGC
                         241                                                   300
    ch2_cgd2_2430  (241) AGGAGGCAGACTTCAACATCTTCTTGAAATTACTAATGCAAAGAATGTATTCGTCATGGT
    ch3_cgd2_2430  (241) AGGAGGCAGACTTCAACATCTTCTTGAAATTACTAATGCAAAGAATGTATTCGTCATGGT
    ch4_cgd2_2430  (241) AGGAGGCAGACTTCAACATCTTCTTGAAATTACTAATGCAAAGAATGTATTCGTCATGGT
  TU502_cgd2_2430  (241) AGGAGGCAGACTTCAACATCTTCTTGAAATTACTAATGCAAAGAATGTATTCGTCATGGT
rabbit1_cgd2_2430  (241) AGGAGGTAGACTTCAACATCTTCTTGAAATTACTAATGCAAAGAATGTATTCGTCATGGT
rabbit2_cgd2_2430  (241) AGGAGGTAGACTTCAACATCTTCTTGAAATTACTAATGCAAAGAATGTATTCGTCATGGT
rabbit3_cgd2_2430  (241) AGGAGGTAGACTTCAACATCTTCTTGAAATTACTAATGCAAAGAATGTATTCGTCATGGT
rabbit4_cgd2_2430  (241) AGGAGGTAGACTTCAACATCTTCTTGAAATTACTAATGCAAAGAATGTATTCGTCATGGT
    cp2_cgd2_2430  (241) AGGAGGCAGACTTCAACATCTTCTTGAAATTACTAATGCAAAGAATGTATTCGTTATGGT
    cp3_cgd2_2430  (241) AGGAGGCAGACTTCAACATCTTCTTGAAATTACTAATGCAAAGAATGTATTCGTTATGGT
    cp4_cgd2_2430  (241) AGGAGGCAGACTTCAACATCTTCTTGAAATTACTAATGCAAAGAATGTATTCGTTATGGT
   Iowa_cgd2_2430  (241) AGGAGGCAGACTTCAACATCTTCTTGAAATTACTAATGCAAAGAATGTATTCGTTATGGT
Moredun_cgd2_2430  (241) AGGAGGCAGACTTCAACATCTTCTTGAAATTACTAATGCAAAGAATGTATTCGTTATGGT
    w65_cgd2_2430  (241) AGGAGGCAGACTTCAACATCTTCTTGAAATTACTAATGCAAAGAATGTATTCGTTATGGT
    w66_cgd2_2430  (241) AGGAGGCAGACTTCAACATCTTCTTGAAATTACTAATGCAAAGAATGTATTCGTTATGGT
    w67_cgd2_2430  (241) AGGAGGCAGACTTCAACATCTTCTTGAAATTACTAATGCAAAGAATGTATTCGTTATGGT
    w70_cgd2_2430  (241) AGGAGGCAGACTTCAACATCTTCTTGAAATTACTAATGCAAAGAATGTATTCGTTATGGT
                         301                                                   360
    ch2_cgd2_2430  (301) GTCTAGATGGTACGGGGGTGTTCAACTAGGGCCAGACAGGTTTAAGCATATTAATAATGC
    ch3_cgd2_2430  (301) GTCTAGATGGTACGGGGGTGTTCAACTAGGGCCAGACAGGTTTAAGCATATTAATAATGC
    ch4_cgd2_2430  (301) GTCTAGATGGTACGGGGGTGTTCAACTAGGGCCAGACAGGTTTAAGCATATTAATAATGC
  TU502_cgd2_2430  (301) GTCTAGATGGTACGGGGGTGTTCAACTAGGGCCAGACAGGTTTAAGCATATTAATAATGC
rabbit1_cgd2_2430  (301) GTCTAGATGGTACGGGGGTGTTCAACTAGGGCCAGACAGGTTTAAGCATATTAATAATGC
rabbit2_cgd2_2430  (301) GTCTAGATGGTACGGGGGTGTTCAACTAGGGCCAGACAGGTTTAAGCATATTAATAATGC
rabbit3_cgd2_2430  (301) GTCTAGATGGTACGGGGGTGTTCAACTAGGGCCAGACAGGTTTAAGCATATTAATAATGC
rabbit4_cgd2_2430  (301) GTCTAGATGGTACGGGGGTGTTCAACTAGGGCCAGACAGGTTTAAGCATATTAATAATGC
    cp2_cgd2_2430  (301) GTCTAGATGGTACGGGGGTATTCAACTAGGGCCAGACAGGTTTAAACATATTAATAATGC
    cp3_cgd2_2430  (301) GTCTAGATGGTACGGGGGTATTCAACTAGGGCCAGACAGGTTTAAACATATTAATAATGC
    cp4_cgd2_2430  (301) GTCTAGATGGTACGGGGGTATTCAACTAGGGCCAGACAGGTTTAAACATATTAATAATGC
   Iowa_cgd2_2430  (301) GTCTAGATGGTACGGGGGTATTCAACTAGGGCCAGACAGGTTTAAACATATTAATAATGC
Moredun_cgd2_2430  (301) GTCTAGATGGTACGGGGGTATTCAACTAGGGCCAGACAGGTTTAAACATATTAATAATGC
    w65_cgd2_2430  (301) GTCTAGATGGTACGGGGGTATTCAACTAGGGCCAGACAGGTTTAAACATATTAATAATGC
    w66_cgd2_2430  (301) GTCTAGATGGTACGGGGGTATTCAACTAGGGCCAGACAGGTTTAAACATATTAATAATGC
    w67_cgd2_2430  (301) GTCTAGATGGTACGGGGGTATTCAACTAGGGCCAGACAGGTTTAAACATATTAATAATGC
    w70_cgd2_2430  (301) GTCTAGATGGTACGGGGGTATTCAACTAGGGCCAGACAGGTTTAAACATATTAATAATGC
                         361
    ch2_cgd2_2430  (361) T
    ch3_cgd2_2430  (361) T
    ch4_cgd2_2430  (361) T
  TU502_cgd2_2430  (361) T
rabbit1_cgd2_2430  (361) T
rabbit2_cgd2_2430  (361) T
rabbit3_cgd2_2430  (361) T
rabbit4_cgd2_2430  (361) T
    cp2_cgd2_2430  (361) T
    cp3_cgd2_2430  (361) T
    cp4_cgd2_2430  (361) T
   Iowa_cgd2_2430  (361) T
Moredun_cgd2_2430  (361) T
    w65_cgd2_2430  (361) T
```

```
w66_cgd2_2430  (361) T
w67_cgd2_2430  (361) T
w70_cgd2_2430  (361) T
```

Chro.30149 gene PCR products

```
                                  1                                                  60
          ch2_chro.30149    (1) CAGGTGGTGTAACTAGAGAATGGTATAATATTCTAGCTAGAGAAATGTTCAATCCTGATT
          ch3_chro.30149    (1) CAGGTGGTGTAACTAGAGAATGGTATAATATTCTAGCTAGAGAAATGTTCAATCCTGATT
          ch4_chro.30149    (1) CAGGTGGTGTAACTAGAGAATGGTATAATATTCTAGCTAGAGAAATGTTCAATCCTGATT
        TU502_chro.30149    (1) CAGGTGGTGTAACTAGAGAATGGTATAATATTCTAGCTAGAGAAATGTTCAATCCTGATT
      rabbit1_chro.30149    (1) CAGGTGGTGTAACTAGAGAATGGTATAATATTCTAGCTAGAGAAATGTTCAATCCTGATT
      rabbit2_chro.30149    (1) CAGGTGGTGTAACTAGAGAATGGTATAATATTCTAGCTAGAGAAATGTTCAATCCTGATT
      rabbit3_chro.30149    (1) CAGGTGGTGTAACTAGAGAATGGTATAATATTCTAGCTAGAGAAATGTTCAATCCTGATT
      rabbit4_chro.30149    (1) CAGGTGGTGTAACTAGAGAATGGTATAATATTCTAGCTAGAGAAATGTTCAATCCTGATT
         Iowa_chro.30149    (1) CAGGTGGTGTAACTAGAGAATGGTATAATATTCTAGCTAGAGAAATGTTCAATCCTGATT
      moredun_chro.30149    (1) CAGGTGGTGTAACTAGAGAATGGTATAATATTCTAGCTAGAGAAATGTTCAATCCTGATT
          cp2_chro.30149    (1) CAGGTGGTGTAACTAGAGAATGGTATAATATTCTAGCTAGAGAAATGTTCAATCCTGATT
          cp3_chro.30149    (1) CAGGTGGTGTAACTAGAGAATGGTATAATATTCTAGCTAGAGAAATGTTCAATCCTGATT
  c.meleagridis_chro.30149  (1) CAGGTGGTGTAACAAGGGAATGGTATAATATTCTAGCTAGGGAGATGTTTAATCCCGACT
                                 61                                                 120
          ch2_chro.30149   (61) ATGCTTTATTTAGAAGAGAAGGATCAAAGAGTGAATTTAACCATCCAAATCCATTAAGTT
          ch3_chro.30149   (61) ATGCTTTATTTAGAAGAGAAGGATCAAAGAGTGAATTTAACCATCCAAATCCATTAAGTT
          ch4_chro.30149   (61) ATGCTTTATTTAGAAGAGAAGGATCAAAGAGTGAATTTAACCATCCAAATCCATTAAGTT
        TU502_chro.30149   (61) ATGCTTTATTTAGAAGAGAAGGATCAAAGAGTGAATTTAACCATCCAAATCCATTAAGTT
      rabbit1_chro.30149   (61) ATGCTTTATTTAGAAGAGAAGGATCAAAGAGTGAATTTAACCATCCAAATCCATTAAGTT
      rabbit2_chro.30149   (61) ATGCTTTATTTAGAAGAGAAGGATCAAAGAGTGAATTTAACCATCCAAATCCATTAAGTT
      rabbit3_chro.30149   (61) ATGCTTTATTTAGAAGAGAAGGATCAAAGAGTGAATTTAACCATCCAAATCCATTAAGTT
      rabbit4_chro.30149   (61) ATGCTTTATTTAGAAGAGAAGGATCAAAGAGTGAATTTAACCATCCAAATCCATTAAGTT
         Iowa_chro.30149   (61) ATGCTTTATTTAGAAGAGAAGGATCAAAGAGTGAATTTAACCATCCAAATCCATTAAGTT
      moredun_chro.30149   (61) ATGCTTTATTTAGAAGAGAAGGATCAAAGAGTGAATTTAACCATCCAAATCCATTAAGTT
          cp2_chro.30149   (61) ATGCTTTATTTAGAAGAGAAGGATCAAAGAGTGAATTTAACCATCCAAATCCATTAAGTT
          cp3_chro.30149   (61) ATGCTTTATTTAGAAGAGAAGGATCAAAGAGTGAATTTAACCATCCAAATCCATTAAGTT
  c.meleagridis_chro.30149   (61) ACGCTTTGTTCAGAAGAGAAGGATCAAAGAGTGAATTTAACCATCCAAACCCATTAAGTT
                                121                                                 180
          ch2_chro.30149  (121) ATATAAATGCTGATCATCTTCATTTCTTTAAATTTATTGGACGTATAATAGGGAAATGTA
          ch3_chro.30149  (121) ATATAAATGCTGATCATCTTCATTTCTTTAAATTTATTGGACGTATAATAGGGAAATGTA
          ch4_chro.30149  (121) ATATAAATGCTGATCATCTTCATTTCTTTAAATTTATTGGACGTATAATAGGGAAATGTA
        TU502_chro.30149  (121) ATATAAATGCTGATCATCTTCATTTCTTTAAATTTATTGGACGTATAATAGGGAAATGTA
      rabbit1_chro.30149  (121) ATATAAATGCTGATCATCTTCATTTCTTTAAATTTATTGGACGTATAATAGGGAAATGTA
      rabbit2_chro.30149  (121) ATATAAATGCTGATCATCTTCATTTCTTTAAATTTATTGGACGTATAATAGGGAAATGTA
      rabbit3_chro.30149  (121) ATATAAATGCTGATCATCTTCATTTCTTTAAATTTATTGGACGTATAATAGGGAAATGTA
      rabbit4_chro.30149  (121) ATATAAATGCTGATCATCTTCATTTCTTTAAATTTATTGGACGTATAATAGGGAAATGTA
         Iowa_chro.30149  (121) ATATAAATGCTGATCATCTTCATTTCTTTAAATTTATTGGACGTATAATAGGGAAATGTA
      moredun_chro.30149  (121) ATATAAATGCTGATCATCTTCATTTCTTTAAATTTATTGGACGTATAATAGGGAAATGTA
          cp2_chro.30149  (121) ATATAAATGCTGATCATCTTCATTTCTTTAAATTTATTGGACGTATAATAGGGAAATGTA
          cp3_chro.30149  (121) ATATAAATGCTGATCATCTTCATTTCTTTAAATTTATTGGACGTATAATAGGGAAATGTA
  c.meleagridis_chro.30149  (121) ATATAAATGCTGATCATCTTCATTTCTTTAAGTTTATCGGACGTATCATAGGTAAATGCA
                                181                                                 240
          ch2_chro.30149  (181) TTTATGATGGTCAACATTTGGATGCATGGTTTACTCGATCATTTTATAAGAATATGTTAG
          ch3_chro.30149  (181) TTTATGATGGTCAACATTTGGATGCATGGTTTACTCGATCATTTTATAAGAATATGTTAG
          ch4_chro.30149  (181) TTTATGATGGTCAACATTTGGATGCATGGTTTACTCGATCATTTTATAAGAATATGTTAG
        TU502_chro.30149  (181) TTTATGATGGTCAACATTTGGATGCATGGTTTACTCGATCATTTTATAAGAATATGTTAG
      rabbit1_chro.30149  (181) TTTATGATGGTCAACATTTGGATGCATGGTTTACTCGATCATTTTATAAGAATATGTTAG
      rabbit2_chro.30149  (181) TTTATGATGGTCAACATTTGGATGCATGGTTTACTCGATCATTTTATAAGAATATGTTAG
      rabbit3_chro.30149  (181) TTTATGATGGTCAACATTTGGATGCATGGTTTACTCGATCATTTTATAAGAATATGTTAG
      rabbit4_chro.30149  (181) TTTATGATGGTCAACATTTGGATGCATGGTTTACTCGATCATTTTATAAGAATATGTTAG
         Iowa_chro.30149  (181) TTTATGATGGTCAACATTTGGATGCATGGTTTACTCGATCATTTTATAAGAATATGTTAG
      moredun_chro.30149  (181) TTTATGATGGTCAACATTTGGATGCATGGTTTACTCGATCATTTTATAAGAATATGTTAG
          cp2_chro.30149  (181) TTTATGATGGTCAACATTTGGATGCATGGTTTACTCGATCATTTTATAAGAATATGTTAG
          cp3_chro.30149  (181) TTTATGATGGTCAACATTTGGATGCATGGTTTACTCGATCATTTTATAAGAATATGTTAG
  c.meleagridis_chro.30149  (181) TTTATGATGGTCAACATTTGGATGCATGGTTTACTCGATCATTTTATAAGAACATGTTAG
                                241                                                 300
          ch2_chro.30149  (241) GACAACCAATAACACCATCAGATGCAGAATCAATAGATCCTGAACTTTATAAGAATTTGA
          ch3_chro.30149  (241) GACAACCAATAACACCATCAGATGCAGAATCAATAGATCCTGAACTTTATAAGAATTTGA
          ch4_chro.30149  (241) GACAACCAATAACACCATCAGATGCAGAATCAATAGATCCTGAACTTTATAAGAATTTGA
        TU502_chro.30149  (241) GACAACCAATAACACCATCAGATGCAGAATCAATAGATCCTGAACTTTATAAGAATTTGA
      rabbit1_chro.30149  (241) GACAACCAATAACACCATCAGATGCAGAATCAATAGATCCTGAACTTTATAAGAATTTGA
      rabbit2_chro.30149  (241) GACAACCAATAACACCATCAGATGCAGAATCAATAGATCCTGAACTTTATAAGAATTTGA
      rabbit3_chro.30149  (241) GACAACCAATAACACCATCAGATGCAGAATCAATAGATCCTGAACTTTATAAGAATTTGA
      rabbit4_chro.30149  (241) GACAACCAATAACACCATCAGATGCAGAATCAATAGATCCTGAACTTTATAAGAATTTGA
         Iowa_chro.30149  (241) GACAACCAATAACACCATCAGATGCAGAATCAATAGATCCTGAACTTTATAAGAATTTGA
      moredun_chro.30149  (241) GACAACCAATAACACCATCAGATGCAGAATCAATAGATCCTGAACTTTATAAGAATTTGA
          cp2_chro.30149  (241) GACAACCAATAACACCATCAGATGCAGAATCAATAGATCCTGAACTTTATAAGAATTTGA
          cp3_chro.30149  (241) GACAACCAATAACACCATCAGATGCAGAATCAATAGATCCTGAACTTTATAAGAATTTGA
  c.meleagridis_chro.30149  (241) GACAGCCAATAACACCGTCAGACGCAGAATCAATAGATCCTGAACTTTATAAGAATTTGA
                                301
          ch2_chro.30149  (301) ATGTAATG
```

```
            ch3_chro.30149  (301) ATGTAATG
            ch4_chro.30149  (301) ATGTAATG
          TU502_chro.30149  (301) ATGTAATG
        rabbit1_chro.30149  (301) ATGTAATG
        rabbit2_chro.30149  (301) ATGTAATG
        rabbit3_chro.30149  (301) ATGTAATG
        rabbit4_chro.30149  (301) ATGTAATG
           Iowa_chro.30149  (301) ATGTAATG
        moredun_chro.30149  (301) ATGTAATG
            cp2_chro.30149  (301) ATGTAATG
            cp3_chro.30149  (301) ATGTAATG
c.meleagridis_chro.30149  (301) ATGTAATG
```

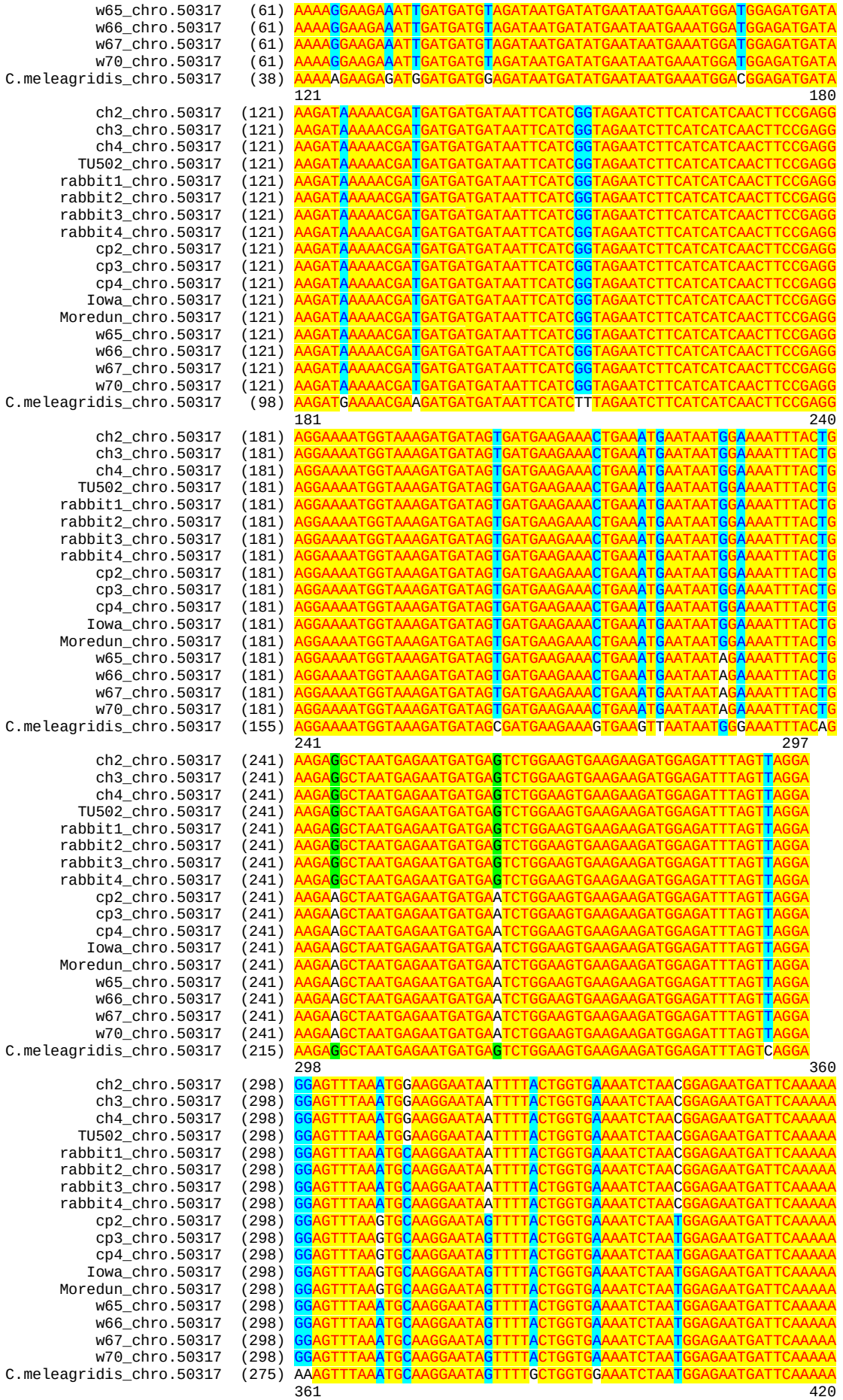
Chro.20156 gene PCR products

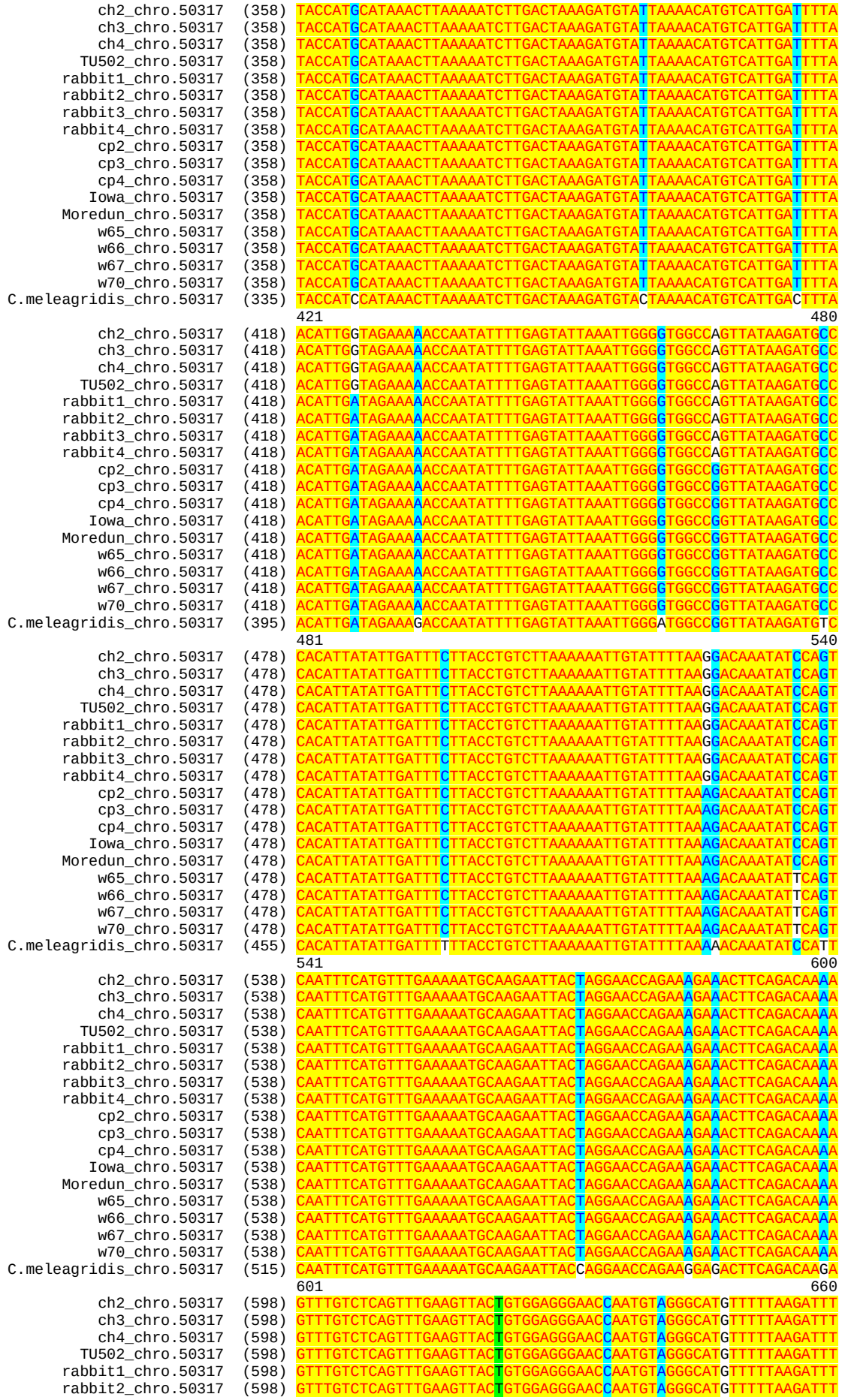
```
                          1                                                  60
      ch2_chro.20156    (1) TTTCGCTTGAAGCCGTAAACTTTAGGATGTTACCAAGGGCAGAGTCTCAAAAAGAATGTA
      ch3_chro.20156    (1) TTTCGCTTGAAGCCGTAAACTTTAGGATGTTACCAAGGGCAGAGTCTCAAAAAGAATGTA
      ch4_chro.20156    (1) TTTCGCTTGAAGCCGTAAACTTTAGGATGTTACCAAGGGCAGAGTCTCAAAAAGAATGTA
    TU502_chro.20156    (1) TTTCGCTTGAAGCCGTAAACTTTAGGATGTTACCAAGGGCAGAGTCTCAAAAAGAATGTA
  rabbit1_chro.20156    (1) TTTCGCTTGAAGCCGTAAACTTTAGGATGTTACCAAGGGCAGAGTCTCAAAAAGAATGTA
  rabbit2_chro.20156    (1) TTTCGCTTGAAGCCGTAAACTTTAGGATGTTACCAAGGGCAGAGTCTCAAAAAGAATGTA
  rabbit3_chro.20156    (1) TTTCGCTTGAAGCCGTAAACTTTAGGATGTTACCAAGGGCAGAGTCTCAAAAAGAATGTA
  rabbit4_chro.20156    (1) TTTCGCTTGAAGCCGTAAACTTTAGGATGTTACCAAGGGCAGAGTCTCAAAAAGAATGTA
      cp2_chro.20156    (1) TTTCGCTTGAAGCCGTAAACTTTAGGATGTTACCAAGGGCAGAGTCTCAAAAAGAATGTA
      cp3_chro.20156    (1) TTTCGCTTGAAGCCGTAAACTTTAGGATGTTACCAAGGGCAGAGTCTCAAAAAGAATGTA
      cp4_chro.20156    (1) TTTCGCTTGAAGCCGTAAACTTTAGGATGTTACCAAGGGCAGAGTCTCAAAAAGAATGTA
     Iowa_chro.20156    (1) TTTCGCTTGAAGCCGTAAACTTTAGGATGTTACCAAGGGCAGAGTCTCAAAAAGAATGTA
  Moredun_chro.20156    (1) TTTCGCTTGAAGCCGTAAACTTTAGGATGTTACCAAGGGCAGAGTCTCAAAAAGAATGTA
      w65_chro.20156    (1) TTTCGCTTGAAGCCGTAAACTTTAGGATGTTACCAAGGGCAGAGTCTCAAAAAGAATGTA
      w66_chro.20156    (1) TTTCGCTTGAAGCCGTAAACTTTAGGATGTTACCAAGGGCAGAGTCTCAAAAAGAATGTA
      w67_chro.20156    (1) TTTCGCTTGAAGCCGTAAACTTTAGGATGTTACCAAGGGCAGAGTCTCAAAAAGAATGTA
      w70_chro.20156    (1) TTTCGCTTGAAGCCGTAAACTTTAGGATGTTACCAAGGGCAGAGTCTCAAAAAGAATGTA
                         61                                                 120
      ch2_chro.20156   (61) AACTGCCAAGCGGAAAAAAGAGTAGTAGAATTTTATGTTTTACAACGGCAGGAAGACGGC
      ch3_chro.20156   (61) AACTGCCAAGCGGAAAAAAGAGTAGTAGAATTTTATGTTTTACAACGGCAGGAAGACGGC
      ch4_chro.20156   (61) AACTGCCAAGCGGAAAAAAGAGTAGTAGAATTTTATGTTTTACAACGGCAGGAAGACGGC
    TU502_chro.20156   (61) AACTGCCAAGCGGAAAAAAGAGTAGTAGAATTTTATGTTTTACAACGGCAGGAAGACGGC
  rabbit1_chro.20156   (61) AACTGCCAAGCGGAAAAAAGAGTAGTAGAATTTTATGTTTTACAACGGCAGGAAGACGGC
  rabbit2_chro.20156   (61) AACTGCCAAGCGGAAAAAAGAGTAGTAGAATTTTATGTTTTACAACGGCAGGAAGACGGC
  rabbit3_chro.20156   (61) AACTGCCAAGCGGAAAAAAGAGTAGTAGAATTTTATGTTTTACAACGGCAGGAAGACGGC
  rabbit4_chro.20156   (61) AACTGCCAAGCGGAAAAAAGAGTAGTAGAATTTTATGTTTTACAACGGCAGGAAGACGGC
      cp2_chro.20156   (61) AACTGCCAAGCGGAAAAAAGAGTAGTAGAATTTTATGTTTTACAACAGCAGGAAGACGGC
      cp3_chro.20156   (61) AACTGCCAAGCGGAAAAAAGAGTAGTAGAATTTTATGTTTTACAACAGCAGGAAGACGGC
      cp4_chro.20156   (61) AACTGCCAAGCGGAAAAAAGAGTAGTAGAATTTTATGTTTTACAACAGCAGGAAGACGGC
     Iowa_chro.20156   (61) AACTGCCAAGCGGAAAAAAGAGTAGTAGAATTTTATGTTTTACAACAGCAGGAAGACGGC
  Moredun_chro.20156   (61) AACTGCCAAGCGGAAAAAAGAGTAGTAGAATTTTATGTTTTACAACAGCAGGAAGACGGC
      w65_chro.20156   (61) AACTGCCAAGCGGAAAAAAGAGTAGTAGAATTTTATGTTTTACAACAGCAGGAAGACGGC
      w66_chro.20156   (61) AACTGCCAAGCGGAAAAAAGAGTAGTAGAATTTTATGTTTTACAACAGCAGGAAGACGGC
      w67_chro.20156   (61) AACTGCCAAGCGGAAAAAAGAGTAGTAGAATTTTATGTTTTACAACAGCAGGAAGACGGC
      w70_chro.20156   (61) AACTGCCAAGCGGAAAAAAGAGTAGTAGAATTTTATGTTTTACAACAGCAGGAAGACGGC
                        121                                                 180
      ch2_chro.20156  (121) ATATTGATCCAACCGAACAAATGATTGAGACAATGTCAATATCAAATCCTGAAGATATAG
      ch3_chro.20156  (121) ATATTGATCCAACCGAACAAATGATTGAGACAATGTCAATATCAAATCCTGAAGATATAG
      ch4_chro.20156  (121) ATATTGATCCAACCGAACAAATGATTGAGACAATGTCAATATCAAATCCTGAAGATATAG
    TU502_chro.20156  (121) ATATTGATCCAACCGAACAAATGATTGAGACAATGTCAATATCAAATCCTGAAGATATAG
  rabbit1_chro.20156  (121) ATATTGATCCAACCGAACAAATGATTGAGACAATGTCAATATCAAATCCTGAAGATATAG
  rabbit2_chro.20156  (121) ATATTGATCCAACCGAACAAATGATTGAGACAATGTCAATATCAAATCCTGAAGATATAG
  rabbit3_chro.20156  (121) ATATTGATCCAACCGAACAAATGATTGAGACAATGTCAATATCAAATCCTGAAGATATAG
  rabbit4_chro.20156  (121) ATATTGATCCAACCGAACAAATGATTGAGACAATGTCAATATCAAATCCTGAAGATATAG
      cp2_chro.20156  (121) ATATTGATCCAACCGAACAAATGATTGAGACAATGTCAATATCAAATCCTGAAGATATAG
      cp3_chro.20156  (121) ATATTGATCCAACCGAACAAATGATTGAGACAATGTCAATATCAAATCCTGAAGATATAG
      cp4_chro.20156  (121) ATATTGATCCAACCGAACAAATGATTGAGACAATGTCAATATCAAATCCTGAAGATATAG
     Iowa_chro.20156  (121) ATATTGATCCAACCGAACAAATGATTGAGACAATGTCAATATCAAATCCTGAAGATATAG
  Moredun_chro.20156  (121) ATATTGATCCAACCGAACAAATGATTGAGACAATGTCAATATCAAATCCTGAAGATATAG
      w65_chro.20156  (121) ATATTGATCCAACCGAACAAATGATTGAGACAATGTCAATATCAAATCCTGAAGATATAG
      w66_chro.20156  (121) ATATTGATCCAACCGAACAAATGATTGAGACAATGTCAATATCAAATCCTGAAGATATAG
      w67_chro.20156  (121) ATATTGATCCAACCGAACAAATGATTGAGACAATGTCAATATCAAATCCTGAAGATATAG
      w70_chro.20156  (121) ATATTGATCCAACCGAACAAATGATTGAGACAATGTCAATATCAAATCCTGAAGATATAG
                        181                                                 240
      ch2_chro.20156  (181) AAGAAAGTCCAGCTCAAGTAAATGAAGTCGAAAACATGGTCCCAGAAAACTTGCCTGGTA
      ch3_chro.20156  (181) AAGAAAGTCCAGCTCAAGTAAATGAAGTCGAAAACATGGTCCCAGAAAACTTGCCTGGTA
      ch4_chro.20156  (181) AAGAAAGTCCAGCTCAAGTAAATGAAGTCGAAAACATGGTCCCAGAAAACTTGCCTGGTA
```

```
  TU502_chro.20156  (181) AAGAAAGTCCAGCTCAAGTAAATGAAGTCGAAAAACATGGTCCCAGAAAACTTGCCTGGTA
rabbit1_chro.20156  (181) AAGAAAGTCCAGCTCTAGTAAATGAAGTCGAAAAACATGGTCCCAGAAAACTTGCCTGGTA
rabbit2_chro.20156  (181) AAGAAAGTCCAGCTCTAGTAAATGAAGTCGAAAAACATGGTCCCAGAAAACTTGCCTGGTA
rabbit3_chro.20156  (181) AAGAAAGTCCAGCTCTAGTAAATGAAGTCGAAAAACATGGTCCCAGAAAACTTGCCTGGTA
rabbit4_chro.20156  (181) AAGAAAGTCCAGCTCTAGTAAATGAAGTCGAAAAACATGGTCCCAGAAAACTTGCCTGGTA
    cp2_chro.20156  (181) AAGAAGGTCCAGCTCAAGTAAATGAAGCCGAAAAACATGGTCCCAGACCACTTGCCTGGTA
    cp3_chro.20156  (181) AAGAAGGTCCAGCTCAAGTAAATGAAGCCGAAAAACATGGTCCCAGACCACTTGCCTGGTA
    cp4_chro.20156  (181) AAGAAGGTCCAGCTCAAGTAAATGAAGCCGAAAAACATGGTCCCAGACCACTTGCCTGGTA
   Iowa_chro.20156  (181) AAGAAGGTCCAGCTCAAGTAAATGAAGCCGAAAAACATGGTCCCAGACCACTTGCCTGGTA
Moredun_chro.20156  (181) AAGAAGGTCCAGCTCAAGTAAATGAAGCCGAAAAACATGGTCCCAGACCACTTGCCTGGTA
    w65_chro.20156  (181) AAGAAGGTCCAGCTCAAGTAAATGAAGCCGAAAAACATGGTCCCAGACCACTTGCCTGGTA
    w66_chro.20156  (181) AAGAAGGTCCAGCTCAAGTAAATGAAGCCGAAAAACATGGTCCCAGACCACTTGCCTGGTA
    w67_chro.20156  (181) AAGAAGGTCCAGCTCAAGTAAATGAAGCCGAAAAACATGGTCCCAGACCACTTGCCTGGTA
    w70_chro.20156  (181) AAGAAGGTCCAGCTCAAGTAAATGAAGCCGAAAAACATGGTCCCAGACCACTTGCCTGGTA
                          241
    ch2_chro.20156  (241) TCAATGCC
    ch3_chro.20156  (241) TCAATGCC
    ch4_chro.20156  (241) TCAATGCC
  TU502_chro.20156  (241) TCAATGCC
rabbit1_chro.20156  (241) TCAATGCC
rabbit2_chro.20156  (241) TCAATGCC
rabbit3_chro.20156  (241) TCAATGCC
rabbit4_chro.20156  (241) TCAATGCC
    cp2_chro.20156  (241) TCAATGCC
    cp3_chro.20156  (241) TCAATGCC
    cp4_chro.20156  (241) TCAATGCC
   Iowa_chro.20156  (241) TCAATGCC
Moredun_chro.20156  (241) TCAATGCC
    w65_chro.20156  (241) TCAATGCC
    w66_chro.20156  (241) TCAATGCC
    w67_chro.20156  (241) TCAATGCC
    w70_chro.20156  (241) TCAATGCC
```

Chro.50317 gene PCR products

```
                                1                                                  60
        ch2_chro.50317    (1) GGTCTCGGTCTGATCGCAAAAAAGAACCAAAGAGTGAGGATGACAGTGGAAATAATACTA
        ch3_chro.50317    (1) GGTCTCGGTCTGATCGCAAAAAAGAACCAAAGAGTGAGGATGACAGTGGAAATAATACTA
        ch4_chro.50317    (1) GGTCTCGGTCTGATCGCAAAAAAGAACCAAAGAGTGAGGATGACAGTGGAAATAATACTA
      TU502_chro.50317    (1) GGTCTCGGTCTGATCGCAAAAAAGAACCAAAGAGTGAGGATGACAGTGGAAATAATACTA
    rabbit1_chro.50317    (1) GGTCTCGGTCTGATCGCAAAAAAGAACCAAAGAGTGAGGATGACAGTGGAAATAATACTA
    rabbit2_chro.50317    (1) GGTCTCGGTCTGATCGCAAAAAAGAACCAAAGAGTGAGGATGACAGTGGAAATAATACTA
    rabbit3_chro.50317    (1) GGTCTCGGTCTGATCGCAAAAAAGAACCAAAGAGTGAGGATGACAGTGGAAATAATACTA
    rabbit4_chro.50317    (1) GGTCTCGGTCTGATCGCAAAAAAGAACCAAAGAGTGAGGATGACAGTGGAAATAATACTA
        cp2_chro.50317    (1) GGTCTCGGCCTGATCGCAAAAAAGAACCAAAGAGTGAGGATGACAGTGGAAATAATACTA
        cp3_chro.50317    (1) GGTCTCGGCCTGATCGCAAAAAAGAACCAAAGAGTGAGGATGACAGTGGAAATAATACTA
        cp4_chro.50317    (1) GGTCTCGGCCTGATCGCAAAAAAGAACCAAAGAGTGAGGATGACAGTGGAAATAATACTA
       Iowa_chro.50317    (1) GGTCTCGGCCTGATCGCAAAAAAGAACCAAAGAGTGAGGATGACAGTGGAAATAATACTA
    Moredun_chro.50317    (1) GGTCTCGGCCTGATCGCAAAAAAGAACCAAAGAGTGAGGATGACAGTGGAAATAATACTA
        w65_chro.50317    (1) GGTCTCGGCCTGATCGCAAAAAAGAACCAAAGAGTGAGGATGACAGTGGAAATAATACTA
        w66_chro.50317    (1) GGTCTCGGCCTGATCGCAAAAAAGAACCAAAGAGTGAGGATGACAGTGGAAATAATACTA
        w67_chro.50317    (1) GGTCTCGGCCTGATCGCAAAAAAGAACCAAAGAGTGAGGATGACAGTGGAAATAATACTA
        w70_chro.50317    (1) GGTCTCGGCCTGATCGCAAAAAAGAACCAAAGAGTGAGGATGACAGTGGAAATAATACTA
C.meleagridis_chro.50317  (1) ----------------------GAACCAAAGAGCGAGGATGACAATGGAAATAATACTA
                               61                                                 120
        ch2_chro.50317   (61) AAAAGGAAGAAATTGATGATGCAGATAATGATATGAATAATGAAATGGATGGAGATGATA
        ch3_chro.50317   (61) AAAAGGAAGAAATTGATGATGCAGATAATGATATGAATAATGAAATGGATGGAGATGATA
        ch4_chro.50317   (61) AAAAGGAAGAAATTGATGATGCAGATAATGATATGAATAATGAAATGGATGGAGATGATA
      TU502_chro.50317   (61) AAAAGGAAGAAATTGATGATGCAGATAATGATATGAATAATGAAATGGATGGAGATGATA
    rabbit1_chro.50317   (61) AAAAGGAAGAAATTGATGATGCAGATAATGATATGAATAATGAAATGGATGGAGATGATA
    rabbit2_chro.50317   (61) AAAAGGAAGAAATTGATGATGCAGATAATGATATGAATAATGAAATGGATGGAGATGATA
    rabbit3_chro.50317   (61) AAAAGGAAGAAATTGATGATGCAGATAATGATATGAATAATGAAATGGATGGAGATGATA
    rabbit4_chro.50317   (61) AAAAGGAAGAAATTGATGATGCAGATAATGATATGAATAATGAAATGGATGGAGATGATA
        cp2_chro.50317   (61) AAAAGGAAGAAATTGATGATGTAGATAATGATATGAATAATGAAATGGATGGAGATGATA
        cp3_chro.50317   (61) AAAAGGAAGAAATTGATGATGTAGATAATGATATGAATAATGAAATGGATGGAGATGATA
        cp4_chro.50317   (61) AAAAGGAAGAAATTGATGATGTAGATAATGATATGAATAATGAAATGGATGGAGATGATA
       Iowa_chro.50317   (61) AAAAGGAAGAAATTGATGATGTAGATAATGATATGAATAATGAAATGGATGGAGATGATA
    Moredun_chro.50317   (61) AAAAGGAAGAAATTGATGATGTAGATAATGATATGAATAATGAAATGGATGGAGATGATA
```


w65_chro.50317 (61) AAAAGGAAGAAATTGATGATGTAGATAATGATATGAATAATGAAATGGATGGAGATGATA
w66_chro.50317 (61) AAAAGGAAGAAATTGATGATGTAGATAATGATATGAATAATGAAATGGATGGAGATGATA
w67_chro.50317 (61) AAAAGGAAGAAATTGATGATGTAGATAATGATATGAATAATGAAATGGATGGAGATGATA
w70_chro.50317 (61) AAAAGGAAGAAATTGATGATGTAGATAATGATATGAATAATGAAATGGATGGAGATGATA
C.meleagridis_chro.50317 (38) AAAAAGAAGAGATGGATGATGGAGATAATGATATGAATAATGAAATGGACGGAGATGATA
121 180
ch2_chro.50317 (121) AAGATAAAAACGATGATGATGATAATTCATCGGTAGAATCTTCATCATCAACTTCCGAGG
ch3_chro.50317 (121) AAGATAAAAACGATGATGATGATAATTCATCGGTAGAATCTTCATCATCAACTTCCGAGG
ch4_chro.50317 (121) AAGATAAAAACGATGATGATGATAATTCATCGGTAGAATCTTCATCATCAACTTCCGAGG
TU502_chro.50317 (121) AAGATAAAAACGATGATGATGATAATTCATCGGTAGAATCTTCATCATCAACTTCCGAGG
rabbit1_chro.50317 (121) AAGATAAAAACGATGATGATGATAATTCATCGGTAGAATCTTCATCATCAACTTCCGAGG
rabbit2_chro.50317 (121) AAGATAAAAACGATGATGATGATAATTCATCGGTAGAATCTTCATCATCAACTTCCGAGG
rabbit3_chro.50317 (121) AAGATAAAAACGATGATGATGATAATTCATCGGTAGAATCTTCATCATCAACTTCCGAGG
rabbit4_chro.50317 (121) AAGATAAAAACGATGATGATGATAATTCATCGGTAGAATCTTCATCATCAACTTCCGAGG
cp2_chro.50317 (121) AAGATAAAAACGATGATGATGATAATTCATCGGTAGAATCTTCATCATCAACTTCCGAGG
cp3_chro.50317 (121) AAGATAAAAACGATGATGATGATAATTCATCGGTAGAATCTTCATCATCAACTTCCGAGG
cp4_chro.50317 (121) AAGATAAAAACGATGATGATGATAATTCATCGGTAGAATCTTCATCATCAACTTCCGAGG
Iowa_chro.50317 (121) AAGATAAAAACGATGATGATGATAATTCATCGGTAGAATCTTCATCATCAACTTCCGAGG
Moredun_chro.50317 (121) AAGATAAAAACGATGATGATGATAATTCATCGGTAGAATCTTCATCATCAACTTCCGAGG
w65_chro.50317 (121) AAGATAAAAACGATGATGATGATAATTCATCGGTAGAATCTTCATCATCAACTTCCGAGG
w66_chro.50317 (121) AAGATAAAAACGATGATGATGATAATTCATCGGTAGAATCTTCATCATCAACTTCCGAGG
w67_chro.50317 (121) AAGATAAAAACGATGATGATGATAATTCATCGGTAGAATCTTCATCATCAACTTCCGAGG
w70_chro.50317 (121) AAGATAAAAACGATGATGATGATAATTCATCGGTAGAATCTTCATCATCAACTTCCGAGG
C.meleagridis_chro.50317 (98) AAGATGAAAACGAAGATGATGATAATTCATCTTAGAATCTTCATCATCAACTTCCGAGG
181 240
ch2_chro.50317 (181) AGGAAAATGGTAAAGATGATAGTGATGAAGAAACTGAAATGAATAATGGAAAATTTACTG
ch3_chro.50317 (181) AGGAAAATGGTAAAGATGATAGTGATGAAGAAACTGAAATGAATAATGGAAAATTTACTG
ch4_chro.50317 (181) AGGAAAATGGTAAAGATGATAGTGATGAAGAAACTGAAATGAATAATGGAAAATTTACTG
TU502_chro.50317 (181) AGGAAAATGGTAAAGATGATAGTGATGAAGAAACTGAAATGAATAATGGAAAATTTACTG
rabbit1_chro.50317 (181) AGGAAAATGGTAAAGATGATAGTGATGAAGAAACTGAAATGAATAATGGAAAATTTACTG
rabbit2_chro.50317 (181) AGGAAAATGGTAAAGATGATAGTGATGAAGAAACTGAAATGAATAATGGAAAATTTACTG
rabbit3_chro.50317 (181) AGGAAAATGGTAAAGATGATAGTGATGAAGAAACTGAAATGAATAATGGAAAATTTACTG
rabbit4_chro.50317 (181) AGGAAAATGGTAAAGATGATAGTGATGAAGAAACTGAAATGAATAATGGAAAATTTACTG
cp2_chro.50317 (181) AGGAAAATGGTAAAGATGATAGTGATGAAGAAACTGAAATGAATAATGGAAAATTTACTG
cp3_chro.50317 (181) AGGAAAATGGTAAAGATGATAGTGATGAAGAAACTGAAATGAATAATGGAAAATTTACTG
cp4_chro.50317 (181) AGGAAAATGGTAAAGATGATAGTGATGAAGAAACTGAAATGAATAATGGAAAATTTACTG
Iowa_chro.50317 (181) AGGAAAATGGTAAAGATGATAGTGATGAAGAAACTGAAATGAATAATGGAAAATTTACTG
Moredun_chro.50317 (181) AGGAAAATGGTAAAGATGATAGTGATGAAGAAACTGAAATGAATAATGGAAAATTTACTG
w65_chro.50317 (181) AGGAAAATGGTAAAGATGATAGTGATGAAGAAACTGAAATGAATAATAGAAAATTTACTG
w66_chro.50317 (181) AGGAAAATGGTAAAGATGATAGTGATGAAGAAACTGAAATGAATAATAGAAAATTTACTG
w67_chro.50317 (181) AGGAAAATGGTAAAGATGATAGTGATGAAGAAACTGAAATGAATAATAGAAAATTTACTG
w70_chro.50317 (181) AGGAAAATGGTAAAGATGATAGTGATGAAGAAACTGAAATGAATAATAGAAAATTTACTG
C.meleagridis_chro.50317 (155) AGGAAAATGGTAAAGATGATAGCGATGAAGAAAGTGAAGTTAATAATGGGAAATTTACAG
241 297
ch2_chro.50317 (241) AAGAGGCTAATGAGAATGATGAGTCTGGAAGTGAAGAAGATGGAGATTTAGTTAGGA
ch3_chro.50317 (241) AAGAGGCTAATGAGAATGATGAGTCTGGAAGTGAAGAAGATGGAGATTTAGTTAGGA
ch4_chro.50317 (241) AAGAGGCTAATGAGAATGATGAGTCTGGAAGTGAAGAAGATGGAGATTTAGTTAGGA
TU502_chro.50317 (241) AAGAGGCTAATGAGAATGATGAGTCTGGAAGTGAAGAAGATGGAGATTTAGTTAGGA
rabbit1_chro.50317 (241) AAGAGGCTAATGAGAATGATGAGTCTGGAAGTGAAGAAGATGGAGATTTAGTTAGGA
rabbit2_chro.50317 (241) AAGAGGCTAATGAGAATGATGAGTCTGGAAGTGAAGAAGATGGAGATTTAGTTAGGA
rabbit3_chro.50317 (241) AAGAGGCTAATGAGAATGATGAGTCTGGAAGTGAAGAAGATGGAGATTTAGTTAGGA
rabbit4_chro.50317 (241) AAGAGGCTAATGAGAATGATGAGTCTGGAAGTGAAGAAGATGGAGATTTAGTTAGGA
cp2_chro.50317 (241) AAGAAGCTAATGAGAATGATGAATCTGGAAGTGAAGAAGATGGAGATTTAGTTAGGA
cp3_chro.50317 (241) AAGAAGCTAATGAGAATGATGAATCTGGAAGTGAAGAAGATGGAGATTTAGTTAGGA
cp4_chro.50317 (241) AAGAAGCTAATGAGAATGATGAATCTGGAAGTGAAGAAGATGGAGATTTAGTTAGGA
Iowa_chro.50317 (241) AAGAAGCTAATGAGAATGATGAATCTGGAAGTGAAGAAGATGGAGATTTAGTTAGGA
Moredun_chro.50317 (241) AAGAAGCTAATGAGAATGATGAATCTGGAAGTGAAGAAGATGGAGATTTAGTTAGGA
w65_chro.50317 (241) AAGAAGCTAATGAGAATGATGAATCTGGAAGTGAAGAAGATGGAGATTTAGTTAGGA
w66_chro.50317 (241) AAGAAGCTAATGAGAATGATGAATCTGGAAGTGAAGAAGATGGAGATTTAGTTAGGA
w67_chro.50317 (241) AAGAAGCTAATGAGAATGATGAATCTGGAAGTGAAGAAGATGGAGATTTAGTTAGGA
w70_chro.50317 (241) AAGAAGCTAATGAGAATGATGAATCTGGAAGTGAAGAAGATGGAGATTTAGTTAGGA
C.meleagridis_chro.50317 (215) AAGAGGCTAATGAGAATGATGAGTCTGGAAGTGAAGAAGATGGAGATTTAGTCAGGA
298 360
ch2_chro.50317 (298) GGAGTTTAAATGGAAGGAATAATTTTACTGGTGAAAATCTAACGGAGAATGATTCAAAAA
ch3_chro.50317 (298) GGAGTTTAAATGGAAGGAATAATTTTACTGGTGAAAATCTAACGGAGAATGATTCAAAAA
ch4_chro.50317 (298) GGAGTTTAAATGGAAGGAATAATTTTACTGGTGAAAATCTAACGGAGAATGATTCAAAAA
TU502_chro.50317 (298) GGAGTTTAAATGGAAGGAATAATTTTACTGGTGAAAATCTAACGGAGAATGATTCAAAAA
rabbit1_chro.50317 (298) GGAGTTTAAATGCAAGGAATAATTTTACTGGTGAAAATCTAACGGAGAATGATTCAAAAA
rabbit2_chro.50317 (298) GGAGTTTAAATGCAAGGAATAATTTTACTGGTGAAAATCTAACGGAGAATGATTCAAAAA
rabbit3_chro.50317 (298) GGAGTTTAAATGCAAGGAATAATTTTACTGGTGAAAATCTAACGGAGAATGATTCAAAAA
rabbit4_chro.50317 (298) GGAGTTTAAATGCAAGGAATAATTTTACTGGTGAAAATCTAACGGAGAATGATTCAAAAA
cp2_chro.50317 (298) GGAGTTTAAGTGCAAGGAATAGTTTTACTGGTGAAAATCTAATGGAGAATGATTCAAAAA
cp3_chro.50317 (298) GGAGTTTAAGTGCAAGGAATAGTTTTACTGGTGAAAATCTAATGGAGAATGATTCAAAAA
cp4_chro.50317 (298) GGAGTTTAAGTGCAAGGAATAGTTTTACTGGTGAAAATCTAATGGAGAATGATTCAAAAA
Iowa_chro.50317 (298) GGAGTTTAAGTGCAAGGAATAGTTTTACTGGTGAAAATCTAATGGAGAATGATTCAAAAA
Moredun_chro.50317 (298) GGAGTTTAAGTGCAAGGAATAGTTTTACTGGTGAAAATCTAATGGAGAATGATTCAAAAA
w65_chro.50317 (298) GGAGTTTAAATGCAAGGAATAGTTTTACTGGTGAAAATCTAATGGAGAATGATTCAAAAA
w66_chro.50317 (298) GGAGTTTAAATGCAAGGAATAGTTTTACTGGTGAAAATCTAATGGAGAATGATTCAAAAA
w67_chro.50317 (298) GGAGTTTAAATGCAAGGAATAGTTTTACTGGTGAAAATCTAATGGAGAATGATTCAAAAA
w70_chro.50317 (298) GGAGTTTAAATGCAAGGAATAGTTTTACTGGTGAAAATCTAATGGAGAATGATTCAAAAA
C.meleagridis_chro.50317 (275) AAAGTTTAAATGCAAGGAATAGTTTTGCTGGTGGAAATCTAATGGAGAATGATTCAAAAA
361 420


ch2_chro.50317 (358) TACCATGCATAAACTTAAAAATCTTGACTAAAGATGTATTAAAACATGTCATTGATTTTA
ch3_chro.50317 (358) TACCATGCATAAACTTAAAAATCTTGACTAAAGATGTATTAAAACATGTCATTGATTTTA
ch4_chro.50317 (358) TACCATGCATAAACTTAAAAATCTTGACTAAAGATGTATTAAAACATGTCATTGATTTTA
TU502_chro.50317 (358) TACCATGCATAAACTTAAAAATCTTGACTAAAGATGTATTAAAACATGTCATTGATTTTA
rabbit1_chro.50317 (358) TACCATGCATAAACTTAAAAATCTTGACTAAAGATGTATTAAAACATGTCATTGATTTTA
rabbit2_chro.50317 (358) TACCATGCATAAACTTAAAAATCTTGACTAAAGATGTATTAAAACATGTCATTGATTTTA
rabbit3_chro.50317 (358) TACCATGCATAAACTTAAAAATCTTGACTAAAGATGTATTAAAACATGTCATTGATTTTA
rabbit4_chro.50317 (358) TACCATGCATAAACTTAAAAATCTTGACTAAAGATGTATTAAAACATGTCATTGATTTTA
cp2_chro.50317 (358) TACCATGCATAAACTTAAAAATCTTGACTAAAGATGTATTAAAACATGTCATTGATTTTA
cp3_chro.50317 (358) TACCATGCATAAACTTAAAAATCTTGACTAAAGATGTATTAAAACATGTCATTGATTTTA
cp4_chro.50317 (358) TACCATGCATAAACTTAAAAATCTTGACTAAAGATGTATTAAAACATGTCATTGATTTTA
Iowa_chro.50317 (358) TACCATGCATAAACTTAAAAATCTTGACTAAAGATGTATTAAAACATGTCATTGATTTTA
Moredun_chro.50317 (358) TACCATGCATAAACTTAAAAATCTTGACTAAAGATGTATTAAAACATGTCATTGATTTTA
w65_chro.50317 (358) TACCATGCATAAACTTAAAAATCTTGACTAAAGATGTATTAAAACATGTCATTGATTTTA
w66_chro.50317 (358) TACCATGCATAAACTTAAAAATCTTGACTAAAGATGTATTAAAACATGTCATTGATTTTA
w67_chro.50317 (358) TACCATGCATAAACTTAAAAATCTTGACTAAAGATGTATTAAAACATGTCATTGATTTTA
w70_chro.50317 (358) TACCATGCATAAACTTAAAAATCTTGACTAAAGATGTATTAAAACATGTCATTGATTTTA
C.meleagridis_chro.50317 (335) TACCATCCATAAACTTAAAAATCTTGACTAAAGATGTACTAAAACATGTCATTGACTTTA
421 480
ch2_chro.50317 (418) ACATTGGTAGAAAAACCAATATTTTGAGTATTAAATTGGGGTGGCCAGTTATAAGATGCC
ch3_chro.50317 (418) ACATTGGTAGAAAAACCAATATTTTGAGTATTAAATTGGGGTGGCCAGTTATAAGATGCC
ch4_chro.50317 (418) ACATTGGTAGAAAAACCAATATTTTGAGTATTAAATTGGGGTGGCCAGTTATAAGATGCC
TU502_chro.50317 (418) ACATTGGTAGAAAAACCAATATTTTGAGTATTAAATTGGGGTGGCCAGTTATAAGATGCC
rabbit1_chro.50317 (418) ACATTGATAGAAAAACCAATATTTTGAGTATTAAATTGGGGTGGCCAGTTATAAGATGCC
rabbit2_chro.50317 (418) ACATTGATAGAAAAACCAATATTTTGAGTATTAAATTGGGGTGGCCAGTTATAAGATGCC
rabbit3_chro.50317 (418) ACATTGATAGAAAAACCAATATTTTGAGTATTAAATTGGGGTGGCCAGTTATAAGATGCC
rabbit4_chro.50317 (418) ACATTGATAGAAAAACCAATATTTTGAGTATTAAATTGGGGTGGCCAGTTATAAGATGCC
cp2_chro.50317 (418) ACATTGATAGAAAAACCAATATTTTGAGTATTAAATTGGGGTGGCCGGTTATAAGATGCC
cp3_chro.50317 (418) ACATTGATAGAAAAACCAATATTTTGAGTATTAAATTGGGGTGGCCGGTTATAAGATGCC
cp4_chro.50317 (418) ACATTGATAGAAAAACCAATATTTTGAGTATTAAATTGGGGTGGCCGGTTATAAGATGCC
Iowa_chro.50317 (418) ACATTGATAGAAAAACCAATATTTTGAGTATTAAATTGGGGTGGCCGGTTATAAGATGCC
Moredun_chro.50317 (418) ACATTGATAGAAAAACCAATATTTTGAGTATTAAATTGGGGTGGCCGGTTATAAGATGCC
w65_chro.50317 (418) ACATTGATAGAAAAACCAATATTTTGAGTATTAAATTGGGGTGGCCGGTTATAAGATGCC
w66_chro.50317 (418) ACATTGATAGAAAAACCAATATTTTGAGTATTAAATTGGGGTGGCCGGTTATAAGATGCC
w67_chro.50317 (418) ACATTGATAGAAAAACCAATATTTTGAGTATTAAATTGGGGTGGCCGGTTATAAGATGCC
w70_chro.50317 (418) ACATTGATAGAAAAACCAATATTTTGAGTATTAAATTGGGGTGGCCGGTTATAAGATGCC
C.meleagridis_chro.50317 (395) ACATTGATAGAAAGACCAATATTTTGAGTATTAAATTGGGATGGCCGGTTATAAGATGTC
481 540
ch2_chro.50317 (478) CACATTATATTGATTTCTTACCTGTCTTAAAAAATTGTATTTTAAGGACAAATATCCAGT
ch3_chro.50317 (478) CACATTATATTGATTTCTTACCTGTCTTAAAAAATTGTATTTTAAGGACAAATATCCAGT
ch4_chro.50317 (478) CACATTATATTGATTTCTTACCTGTCTTAAAAAATTGTATTTTAAGGACAAATATCCAGT
TU502_chro.50317 (478) CACATTATATTGATTTCTTACCTGTCTTAAAAAATTGTATTTTAAGGACAAATATCCAGT
rabbit1_chro.50317 (478) CACATTATATTGATTTCTTACCTGTCTTAAAAAATTGTATTTTAAGGACAAATATCCAGT
rabbit2_chro.50317 (478) CACATTATATTGATTTCTTACCTGTCTTAAAAAATTGTATTTTAAGGACAAATATCCAGT
rabbit3_chro.50317 (478) CACATTATATTGATTTCTTACCTGTCTTAAAAAATTGTATTTTAAGGACAAATATCCAGT
rabbit4_chro.50317 (478) CACATTATATTGATTTCTTACCTGTCTTAAAAAATTGTATTTTAAGGACAAATATCCAGT
cp2_chro.50317 (478) CACATTATATTGATTTCTTACCTGTCTTAAAAAATTGTATTTTAAAGACAAATATCCAGT
cp3_chro.50317 (478) CACATTATATTGATTTCTTACCTGTCTTAAAAAATTGTATTTTAAAGACAAATATCCAGT
cp4_chro.50317 (478) CACATTATATTGATTTCTTACCTGTCTTAAAAAATTGTATTTTAAAGACAAATATCCAGT
Iowa_chro.50317 (478) CACATTATATTGATTTCTTACCTGTCTTAAAAAATTGTATTTTAAAGACAAATATCCAGT
Moredun_chro.50317 (478) CACATTATATTGATTTCTTACCTGTCTTAAAAAATTGTATTTTAAAGACAAATATCCAGT
w65_chro.50317 (478) CACATTATATTGATTTCTTACCTGTCTTAAAAAATTGTATTTTAAAGACAAATATTCAGT
w66_chro.50317 (478) CACATTATATTGATTTCTTACCTGTCTTAAAAAATTGTATTTTAAAGACAAATATTCAGT
w67_chro.50317 (478) CACATTATATTGATTTCTTACCTGTCTTAAAAAATTGTATTTTAAAGACAAATATTCAGT
w70_chro.50317 (478) CACATTATATTGATTTCTTACCTGTCTTAAAAAATTGTATTTTAAAGACAAATATTCAGT
C.meleagridis_chro.50317 (455) CACATTATATTGATTTTTACCTGTCTTAAAAAATTGTATTTTAAAACAAATATCCATT
541 600
ch2_chro.50317 (538) CAATTTCATGTTTGAAAAATGCAAGAATTACTAGGAACCAGAAAGAAAACTTCAGACAAAA
ch3_chro.50317 (538) CAATTTCATGTTTGAAAAATGCAAGAATTACTAGGAACCAGAAAGAAAACTTCAGACAAAA
ch4_chro.50317 (538) CAATTTCATGTTTGAAAAATGCAAGAATTACTAGGAACCAGAAAGAAAACTTCAGACAAAA
TU502_chro.50317 (538) CAATTTCATGTTTGAAAAATGCAAGAATTACTAGGAACCAGAAAGAAAACTTCAGACAAAA
rabbit1_chro.50317 (538) CAATTTCATGTTTGAAAAATGCAAGAATTACTAGGAACCAGAAAGAAAACTTCAGACAAAA
rabbit2_chro.50317 (538) CAATTTCATGTTTGAAAAATGCAAGAATTACTAGGAACCAGAAAGAAAACTTCAGACAAAA
rabbit3_chro.50317 (538) CAATTTCATGTTTGAAAAATGCAAGAATTACTAGGAACCAGAAAGAAAACTTCAGACAAAA
rabbit4_chro.50317 (538) CAATTTCATGTTTGAAAAATGCAAGAATTACTAGGAACCAGAAAGAAAACTTCAGACAAAA
cp2_chro.50317 (538) CAATTTCATGTTTGAAAAATGCAAGAATTACTAGGAACCAGAAAGAAAACTTCAGACAAAA
cp3_chro.50317 (538) CAATTTCATGTTTGAAAAATGCAAGAATTACTAGGAACCAGAAAGAAAACTTCAGACAAAA
cp4_chro.50317 (538) CAATTTCATGTTTGAAAAATGCAAGAATTACTAGGAACCAGAAAGAAAACTTCAGACAAAA
Iowa_chro.50317 (538) CAATTTCATGTTTGAAAAATGCAAGAATTACTAGGAACCAGAAAGAAAACTTCAGACAAAA
Moredun_chro.50317 (538) CAATTTCATGTTTGAAAAATGCAAGAATTACTAGGAACCAGAAAGAAAACTTCAGACAAAA
w65_chro.50317 (538) CAATTTCATGTTTGAAAAATGCAAGAATTACTAGGAACCAGAAAGAAAACTTCAGACAAAA
w66_chro.50317 (538) CAATTTCATGTTTGAAAAATGCAAGAATTACTAGGAACCAGAAAGAAAACTTCAGACAAAA
w67_chro.50317 (538) CAATTTCATGTTTGAAAAATGCAAGAATTACTAGGAACCAGAAAGAAAACTTCAGACAAAA
w70_chro.50317 (538) CAATTTCATGTTTGAAAAATGCAAGAATTACTAGGAACCAGAAAGAAAACTTCAGACAAAA
C.meleagridis_chro.50317 (515) CAATTTCATGTTTGAAAAATGCAAGAATTACCAGGAACCAGAAGGAGACTTCAGACAAGA
601 660
ch2_chro.50317 (598) GTTTGTCTCAGTTTGAAGTTACTGTGGAGGGAACCAATGTAGGGCATGTTTTTAAGATTT
ch3_chro.50317 (598) GTTTGTCTCAGTTTGAAGTTACTGTGGAGGGAACCAATGTAGGGCATGTTTTTAAGATTT
ch4_chro.50317 (598) GTTTGTCTCAGTTTGAAGTTACTGTGGAGGGAACCAATGTAGGGCATGTTTTTAAGATTT
TU502_chro.50317 (598) GTTTGTCTCAGTTTGAAGTTACTGTGGAGGGAACCAATGTAGGGCATGTTTTTAAGATTT
rabbit1_chro.50317 (598) GTTTGTCTCAGTTTGAAGTTACTGTGGAGGGAACCAATGTAGGGCATGTTTTTAAGATTT
rabbit2_chro.50317 (598) GTTTGTCTCAGTTTGAAGTTACTGTGGAGGGAACCAATGTAGGGCATGTTTTTAAGATTT

```
       rabbit3_chro.50317  (598) GTTTGTCTCAGTTTGAAGTTACTGTGGAGGGAACCAATGTAGGGCATGTTTTTAAGATTT
       rabbit4_chro.50317  (598) GTTTGTCTCAGTTTGAAGTTACTGTGGAGGGAACCAATGTAGGGCATGTTTTTAAGATTT
           cp2_chro.50317  (598) GTTTGTCTCAGTTTGAAGTTACCGTGGAGGGAACCAATGTAGGGCATATTTTTAAGATTT
           cp3_chro.50317  (598) GTTTGTCTCAGTTTGAAGTTACCGTGGAGGGAACCAATGTAGGGCATATTTTTAAGATTT
           cp4_chro.50317  (598) GTTTGTCTCAGTTTGAAGTTACCGTGGAGGGAACCAATGTAGGGCATATTTTTAAGATTT
          Iowa_chro.50317  (598) GTTTGTCTCAGTTTGAAGTTACCGTGGAGGGAACCAATGTAGGGCATATTTTTAAGATTT
       Moredun_chro.50317  (598) GTTTGTCTCAGTTTGAAGTTACCGTGGAGGGAACCAATGTAGGGCATATTTTTAAGATTT
           w65_chro.50317  (598) GTTTGTCTCAGTTTGAAGTTACCGTGGAGGGAACCAATGTAGGGCATATTTTTAAGATTT
           w66_chro.50317  (598) GTTTGTCTCAGTTTGAAGTTACCGTGGAGGGAACCAATGTAGGGCATATTTTTAAGATTT
           w67_chro.50317  (598) GTTTGTCTCAGTTTGAAGTTACCGTGGAGGGAACCAATGTAGGGCATATTTTTAAGATTT
           w70_chro.50317  (598) GTTTGTCTCAGTTTGAAGTTACCGTGGAGGGAACCAATGTAGGGCATATTTTTAAGATTT
C.meleagridis_chro.50317  (575) GTTTGTCTCAGTTTGAAGTTACTGTGGAGGGAACTAATGTGGGGCATATTTTTAAGATTT
                                661                                                 720
           ch2_chro.50317  (658) CGCCACGTTACATTAATCACGATAAAATTAGATTCAATGATATTCAAACAGTATATAAAT
           ch3_chro.50317  (658) CGCCACGTTACATTAATCACGATAAAATTAGATTCAATGATATTCAAACAGTATATAAAT
           ch4_chro.50317  (658) CGCCACGTTACATTAATCACGATAAAATTAGATTCAATGATATTCAAACAGTATATAAAT
         TU502_chro.50317  (658) CGCCACGTTACATTAATCACGATAAAATTAGATTCAATGATATTCAAACAGTATATAAAT
       rabbit1_chro.50317  (658) CGCCACGTTACATTAATCACGATAAAATTAGATTCAATGATATTCAAACAGTATATAAAT
       rabbit2_chro.50317  (658) CGCCACGTTACATTAATCACGATAAAATTAGATTCAATGATATTCAAACAGTATATAAAT
       rabbit3_chro.50317  (658) CGCCACGTTACATTAATCACGATAAAATTAGATTCAATGATATTCAAACAGTATATAAAT
       rabbit4_chro.50317  (658) CGCCACGTTACATTAATCACGATAAAATTAGATTCAATGATATTCAAACAGTATATAAAT
           cp2_chro.50317  (658) CGCCACGTTACATTAATCACGATAAAATTAGATTCAATGATATTCAAACAGTATATAAAT
           cp3_chro.50317  (658) CGCCACGTTACATTAATCACGATAAAATTAGATTCAATGATATTCAAACAGTATATAAAT
           cp4_chro.50317  (658) CGCCACGTTACATTAATCACGATAAAATTAGATTCAATGATATTCAAACAGTATATAAAT
          Iowa_chro.50317  (658) CGCCACGTTACATTAATCACGATAAAATTAGATTCAATGATATTCAAACAGTATATAAAT
       Moredun_chro.50317  (658) CGCCACGTTACATTAATCACGATAAAATTAGATTCAATGATATTCAAACAGTATATAAAT
           w65_chro.50317  (658) CGCCACGTTACATTAATCACGATAAAATTAGATTCAATGATATTCAAACAGTATATAAAT
           w66_chro.50317  (658) CGCCACGTTACATTAATCACGATAAAATTAGATTCAATGATATTCAAACAGTATATAAAT
           w67_chro.50317  (658) CGCCACGTTACATTAATCACGATAAAATTAGATTCAATGATATTCAAACAGTATATAAAT
           w70_chro.50317  (658) CGCCACGTTACATTAATCACGATAAAATTAGATTCAATGATATTCAAACAGTATATAAAT
C.meleagridis_chro.50317  (635) CGCCACGTTACATTAATCACGATAAAATTAGATTCAATGTT----------------
                                721
           ch2_chro.50317  (718) ATTATGGTGT
           ch3_chro.50317  (718) ATTATGGTGT
           ch4_chro.50317  (718) ATTATGGTGT
         TU502_chro.50317  (718) ATTATGGTGT
       rabbit1_chro.50317  (718) ATTATGGTGT
       rabbit2_chro.50317  (718) ATTATGGTGT
       rabbit3_chro.50317  (718) ATTATGGTGT
       rabbit4_chro.50317  (718) ATTATGGTGT
           cp2_chro.50317  (718) ATTATGGTGT
           cp3_chro.50317  (718) ATTATGGTGT
           cp4_chro.50317  (718) ATTATGGTGT
          Iowa_chro.50317  (718) ATTATGGTGT
       Moredun_chro.50317  (718) ATTATGGTGT
           w65_chro.50317  (718) ATTATGGTGT
           w66_chro.50317  (718) ATTATGGTGT
           w67_chro.50317  (718) ATTATGGTGT
           w70_chro.50317  (718) ATTATGGTGT
C.meleagridis_chro.50317  (678) ----------
```

Chro.50330 gene PCR products

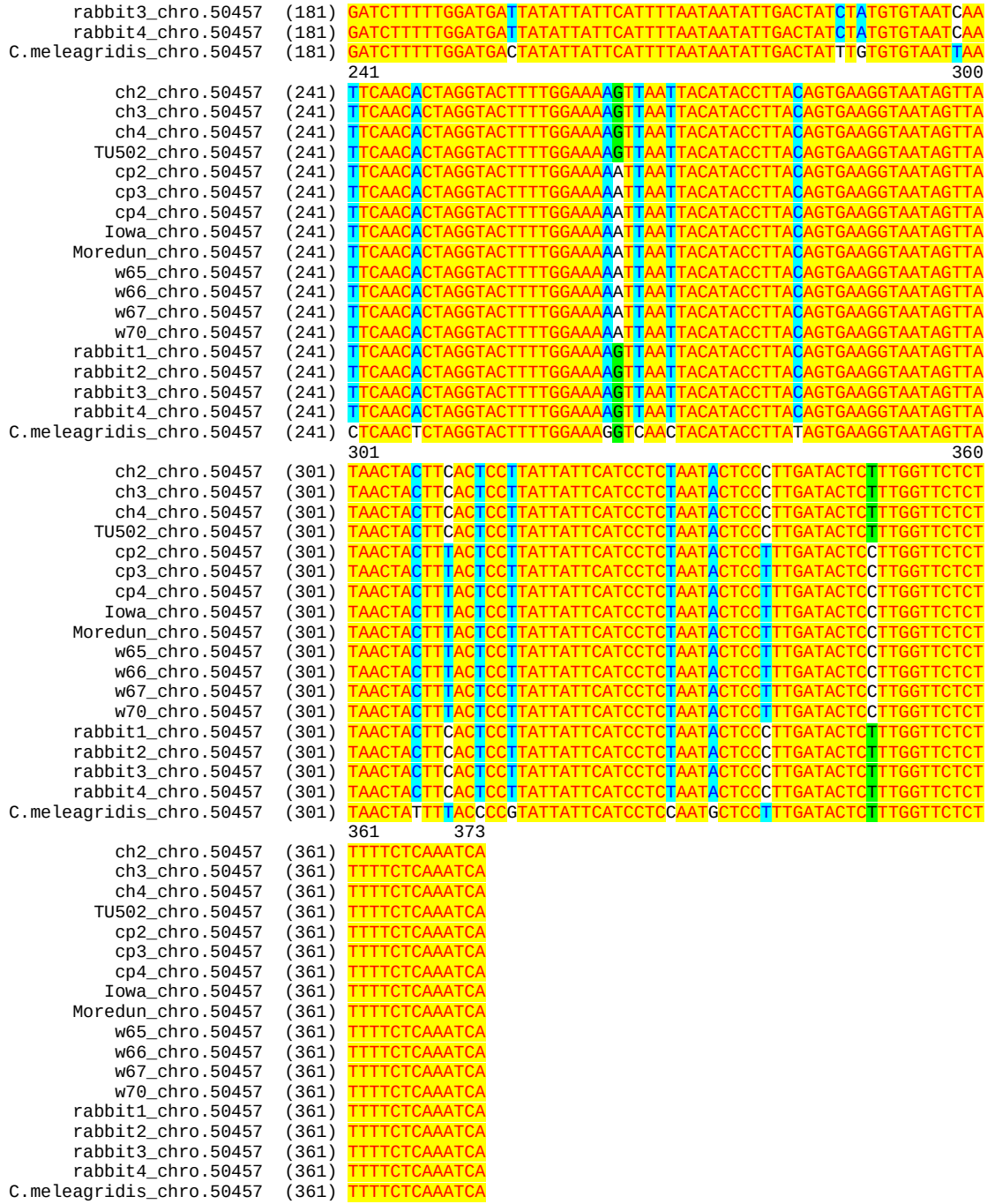
```
                              1                                                   60
    ch2_chro.50330    (1) CAGGTTCAGAAACTCTGGGAAGAGAGTCAAATATATGAGGCAGATGTAGATCCATCTCGT
    ch3_chro.50330    (1) CAGGTTCAGAAACTCTGGGAAGAGAGTCAAATATATGAGGCAGATGTAGATCCATCTCGT
    ch4_chro.50330    (1) CAGGTTCAGAAACTCTGGGAAGAGAGTCAAATATATGAGGCAGATGTAGATCCATCTCGT
  TU502_chro.50330    (1) CAGGTTCAGAAACTCTGGGAAGAGAGTCAAATATATGAGGCAGATGTAGATCCATCTCGT
    cp2_chro.50330    (1) CAGGTTCAAAAACTCTGGGAAGAGAGTCAAATATATGAGGCAGATGTAGATCCATCTCGT
    cp3_chro.50330    (1) CAGGTTCAAAAACTCTGGGAAGAGAGTCAAATATATGAGGCAGATGTAGATCCATCTCGT
    cp4_chro.50330    (1) CAGGTTCAAAAACTCTGGGAAGAGAGTCAAATATATGAGGCAGATGTAGATCCATCTCGT
   Iowa_chro.50330    (1) CAGGTTCAAAAACTCTGGGAAGAGAGTCAAATATATGAGGCAGATGTAGATCCATCTCGT
Moredun_chro.50330    (1) CAGGTTCAAAAACTCTGGGAAGAGAGTCAAATATATGAGGCAGATGTAGATCCATCTCGT
    w65_chro.50330    (1) CAGGTTCAAAAACTCTGGGAAGAGAGTCAAATATATGAGGCAGATGTAGATCCATCTCGT
    w66_chro.50330    (1) CAGGTTCAAAAACTCTGGGAAGAGAGTCAAATATATGAGGCAGATGTAGATCCATCTCGT
    w67_chro.50330    (1) CAGGTTCAAAAACTCTGGGAAGAGAGTCAAATATATGAGGCAGATGTAGATCCATCTCGT
    w70_chro.50330    (1) CAGGTTCAAAAACTCTGGGAAGAGAGTCAAATATATGAGGCAGATGTAGATCCATCTCGT
rabbit1_chro.50330    (1) CAGGTTCAGAAACTCTGGGAAGAGAGTCAAATATATGAGGCAGATGTAGATCCATCTCGT
rabbit2_chro.50330    (1) CAGGTTCAGAAACTCTGGGAAGAGAGTCAAATATATGAGGCAGATGTAGATCCATCTCGT
rabbit3_chro.50330    (1) CAGGTTCAGAAACTCTGGGAAGAGAGTCAAATATATGAGGCAGATGTAGATCCATCTCGT
rabbit4_chro.50330    (1) CAGGTTCAGAAACTCTGGGAAGAGAGTCAAATATATGAGGCAGATGTAGATCCATCTCGT
                              61                                                  120
    ch2_chro.50330   (61) GAAAAATATATGATTACATTTCCATATCCATATATGAATGGAAGGTTACATTTGGGGCAT
    ch3_chro.50330   (61) GAAAAATATATGATTACATTTCCATATCCATATATGAATGGAAGGTTACATTTGGGGCAT
    ch4_chro.50330   (61) GAAAAATATATGATTACATTTCCATATCCATATATGAATGGAAGGTTACATTTGGGGCAT
  TU502_chro.50330   (61) GAAAAATATATGATTACATTTCCATATCCATATATGAATGGAAGGTTACATTTGGGGCAT
    cp2_chro.50330   (61) GAAAAATACATGATTACATTTCCATATCCATATATGAATGGAAGGTTACATTTGGGGCAT
```

```
    cp3_chro.50330   (61) GAAAAATACATGATTACATTTCCATATCCATATATGAATGGAAGGTTACATTTGGGGCAT
    cp4_chro.50330   (61) GAAAAATACATGATTACATTTCCATATCCATATATGAATGGAAGGTTACATTTGGGGCAT
   Iowa_chro.50330   (61) GAAAAATACATGATTACATTTCCATATCCATATATGAATGGAAGGTTACATTTGGGGCAT
Moredun_chro.50330   (61) GAAAAATACATGATTACATTTCCATATCCATATATGAATGGAAGGTTACATTTGGGGCAT
    w65_chro.50330   (61) GAAAAATACATGATTACATTTCCATATCCATATATGAATGGAAGGTTACATTTGGGGCAT
    w66_chro.50330   (61) GAAAAATACATGATTACATTTCCATATCCATATATGAATGGAAGGTTACATTTGGGGCAT
    w67_chro.50330   (61) GAAAAATACATGATTACATTTCCATATCCATATATGAATGGAAGGTTACATTTGGGGCAT
    w70_chro.50330   (61) GAAAAATACATGATTACATTTCCATATCCATATATGAATGGAAGGTTACATTTGGGGCAT
rabbit1_chro.50330   (61) GAAAAATATATGATTACATTTCCATATCCATATATGAATGGAAGGTTACATTTGGGGCAT
rabbit2_chro.50330   (61) GAAAAATATATGATTACATTTCCATATCCATATATGAATGGAAGGTTACATTTGGGGCAT
rabbit3_chro.50330   (61) GAAAAATATATGATTACATTTCCATATCCATATATGAATGGAAGGTTACATTTGGGGCAT
rabbit4_chro.50330   (61) GAAAAATATATGATTACATTTCCATATCCATATATGAATGGAAGGTTACATTTGGGGCAT
                         121                                                     180
    ch2_chro.50330  (121) GCCTTTACGTTGACAAAAGCAGATTTTCAAGCAAGATTCCAAAGAATGAATAATAAAAAT
    ch3_chro.50330  (121) GCCTTTACGTTGACAAAAGCAGATTTTCAAGCAAGATTCCAAAGAATGAATAATAAAAAT
    ch4_chro.50330  (121) GCCTTTACGTTGACAAAAGCAGATTTTCAAGCAAGATTCCAAAGAATGAATAATAAAAAT
  TU502_chro.50330  (121) GCCTTTACGTTGACAAAAGCAGATTTTCAAGCAAGATTCCAAAGAATGAATAATAAAAAT
    cp2_chro.50330  (121) GCCTTTACGTTGACAAAAGCAGATTTTCAAGCAAGATTCCAAAGAATGAATAATAAAAAT
    cp3_chro.50330  (121) GCCTTTACGTTGACAAAAGCAGATTTTCAAGCAAGATTCCAAAGAATGAATAATAAAAAT
    cp4_chro.50330  (121) GCCTTTACGTTGACAAAAGCAGATTTTCAAGCAAGATTCCAAAGAATGAATAATAAAAAT
   Iowa_chro.50330  (121) GCCTTTACGTTGACAAAAGCAGATTTTCAAGCAAGATTCCAAAGAATGAATAATAAAAAT
Moredun_chro.50330  (121) GCCTTTACGTTGACAAAAGCAGATTTTCAAGCAAGATTCCAAAGAATGAATAATAAAAAT
    w65_chro.50330  (121) GCCTTTACGTTGACAAAAGCAGATTTTCAAGCAAGATTCCAAAGAATGAATAATAAAAAT
    w66_chro.50330  (121) GCCTTTACGTTGACAAAAGCAGATTTTCAAGCAAGATTCCAAAGAATGAATAATAAAAAT
    w67_chro.50330  (121) GCCTTTACGTTGACAAAAGCAGATTTTCAAGCAAGATTCCAAAGAATGAATAATAAAAAT
    w70_chro.50330  (121) GCCTTTACGTTGACAAAAGCAGATTTTCAAGCAAGATTCCAAAGAATGAATAATAAAAAT
rabbit1_chro.50330  (121) GCCTTTACGTTGACAAAAGCAGATTTTCAAGCAAGATTCCAAAGAATGAATAATAAAAAT
rabbit2_chro.50330  (121) GCCTTTACGTTGACAAAAGCAGATTTTCAAGCAAGATTCCAAAGAATGAATAATAAAAAT
rabbit3_chro.50330  (121) GCCTTTACGTTGACAAAAGCAGATTTTCAAGCAAGATTCCAAAGAATGAATAATAAAAAT
rabbit4_chro.50330  (121) GCCTTTACGTTGACAAAAGCAGATTTTCAAGCAAGATTCCAAAGAATGAATAATAAAAAT
                         181                                                     240
    ch2_chro.50330  (181) GTATTATTTCCATTTGGTTTTCATTGTACAGGAATGCCAATTTGTGCAAGCGCAGATAAG
    ch3_chro.50330  (181) GTATTATTTCCATTTGGTTTTCATTGTACAGGAATGCCAATTTGTGCAAGCGCAGATAAG
    ch4_chro.50330  (181) GTATTATTTCCATTTGGTTTTCATTGTACAGGAATGCCAATTTGTGCAAGCGCAGATAAG
  TU502_chro.50330  (181) GTATTATTTCCATTTGGTTTTCATTGTACAGGAATGCCAATTTGTGCAAGCGCAGATAAG
    cp2_chro.50330  (181) GTATTATTTCCATTTGGTTTTCATTGTACGGGAATGCCAATTTGTGCAAGCGCAGATAAG
    cp3_chro.50330  (181) GTATTATTTCCATTTGGTTTTCATTGTACGGGAATGCCAATTTGTGCAAGCGCAGATAAG
    cp4_chro.50330  (181) GTATTATTTCCATTTGGTTTTCATTGTACGGGAATGCCAATTTGTGCAAGCGCAGATAAG
   Iowa_chro.50330  (181) GTATTATTTCCATTTGGTTTTCATTGTACGGGAATGCCAATTTGTGCAAGCGCAGATAAG
Moredun_chro.50330  (181) GTATTATTTCCATTTGGTTTTCATTGTACGGGAATGCCAATTTGTGCAAGCGCAGATAAG
    w65_chro.50330  (181) GTATTATTTCCATTTGGTTTTCATTGTACGGGAATGCCAATTTGTGCAAGCGCAGATAAG
    w66_chro.50330  (181) GTATTATTTCCATTTGGTTTTCATTGTACGGGAATGCCAATTTGTGCAAGCGCAGATAAG
    w67_chro.50330  (181) GTATTATTTCCATTTGGTTTTCATTGTACGGGAATGCCAATTTGTGCAAGCGCAGATAAG
    w70_chro.50330  (181) GTATTATTTCCATTTGGTTTTCATTGTACGGGAATGCCAATTTGTGCAAGCGCAGATAAG
rabbit1_chro.50330  (181) GTATTATTTCCATTTGGTTTTCATTGTACAGGAATGCCAATTTGTGCAAGCGCAGATAAG
rabbit2_chro.50330  (181) GTATTATTTCCATTTGGTTTTCATTGTACAGGAATGCCAATTTGTGCAAGCGCAGATAAG
rabbit3_chro.50330  (181) GTATTATTTCCATTTGGTTTTCATTGTACAGGAATGCCAATTTGTGCAAGCGCAGATAAG
rabbit4_chro.50330  (181) GTATTATTTCCATTTGGTTTTCATTGTACAGGAATGCCAATTTGTGCAAGCGCAGATAAG
                         241                                                     300
    ch2_chro.50330  (241) TTAAAGATGGAGTTGAATTCACCAAAGTTAGCTGAAACTGATGATGATAAAGAACAAGAG
    ch3_chro.50330  (241) TTAAAGATGGAGTTGAATTCACCAAAGTTAGCTGAAACTGATGATGATAAAGAACAAGAG
    ch4_chro.50330  (241) TTAAAGATGGAGTTGAATTCACCAAAGTTAGCTGAAACTGATGATGATAAAGAACAAGAG
  TU502_chro.50330  (241) TTAAAGATGGAGTTGAATTCACCAAAGTTAGCTGAAACTGATGATGATAAAGAACAAGAG
    cp2_chro.50330  (241) TTAAAGATGGAGTTGAATTCACCAAAGTTAGCTGAAACTGATGATGATAAAGAACAAGAG
    cp3_chro.50330  (241) TTAAAGATGGAGTTGAATTCACCAAAGTTAGCTGAAACTGATGATGATAAAGAACAAGAG
    cp4_chro.50330  (241) TTAAAGATGGAGTTGAATTCACCAAAGTTAGCTGAAACTGATGATGATAAAGAACAAGAG
   Iowa_chro.50330  (241) TTAAAGATGGAGTTGAATTCACCAAAGTTAGCTGAAACTGATGATGATAAAGAACAAGAG
Moredun_chro.50330  (241) TTAAAGATGGAGTTGAATTCACCAAAGTTAGCTGAAACTGATGATGATAAAGAACAAGAG
    w65_chro.50330  (241) TTAAAGATGGAGTTGAATTCACCAAAGTTAGCTGAAACTGATGATGATAAAGAACAAGAG
    w66_chro.50330  (241) TTAAAGATGGAGTTGAATTCACCAAAGTTAGCTGAAACTGATGATGATAAAGAACAAGAG
    w67_chro.50330  (241) TTAAAGATGGAGTTGAATTCACCAAAGTTAGCTGAAACTGATGATGATAAAGAACAAGAG
    w70_chro.50330  (241) TTAAAGATGGAGTTGAATTCACCAAAGTTAGCTGAAACTGATGATGATAAAGAACAAGAG
rabbit1_chro.50330  (241) TTAAAGATGGAGTTGAATTCACCAAAGTTAGCTGAAACTGATGATGATAAAGAACAAGAG
rabbit2_chro.50330  (241) TTAAAGATGGAGTTGAATTCACCAAAGTTAGCTGAAACTGATGATGATAAAGAACAAGAG
rabbit3_chro.50330  (241) TTAAAGATGGAGTTGAATTCACCAAAGTTAGCTGAAACTGATGATGATAAAGAACAAGAG
rabbit4_chro.50330  (241) TTAAAGATGGAGTTGAATTCACCAAAGTTAGCTGAAACTGATGATGATAAAGAACAAGAG
                         301                                             346
    ch2_chro.50330  (301) ACAAGCCAAGTACAACTTAAAAGTAAGGTTGCAGCTAAAACTGGGG
    ch3_chro.50330  (301) ACAAGCCAAGTACAACTTAAAAGTAAGGTTGCAGCTAAAACTGGGG
    ch4_chro.50330  (301) ACAAGCCAAGTACAACTTAAAAGTAAGGTTGCAGCTAAAACTGGGG
  TU502_chro.50330  (301) ACAAGCCAAGTACAACTTAAAAGTAAGGTTGCAGCTAAAACTGGGG
    cp2_chro.50330  (301) ACAAGCCAAGTACAACTTAAAAGTAAGGTTGCAGCTAAAACTGGGG
    cp3_chro.50330  (301) ACAAGCCAAGTACAACTTAAAAGTAAGGTTGCAGCTAAAACTGGGG
    cp4_chro.50330  (301) ACAAGCCAAGTACAACTTAAAAGTAAGGTTGCAGCTAAAACTGGGG
   Iowa_chro.50330  (301) ACAAGCCAAGTACAACTTAAAAGTAAGGTTGCAGCTAAAACTGGGG
Moredun_chro.50330  (301) ACAAGCCAAGTACAACTTAAAAGTAAGGTTGCAGCTAAAACTGGGG
    w65_chro.50330  (301) ACAAGCCAAGTACAACTTAAAAGTAAGGTTGCAGCTAAAACTGGGG
    w66_chro.50330  (301) ACAAGCCAAGTACAACTTAAAAGTAAGGTTGCAGCTAAAACTGGGG
    w67_chro.50330  (301) ACAAGCCAAGTACAACTTAAAAGTAAGGTTGCAGCTAAAACTGGGG
    w70_chro.50330  (301) ACAAGCCAAGTACAACTTAAAAGTAAGGTTGCAGCTAAAACTGGGG
rabbit1_chro.50330  (301) ACAAGCCAAGTACAACTTAAAAGTAAGGTTGCAGCTAAAACTGGGG
rabbit2_chro.50330  (301) ACAAGCCAAGTACAACTTAAAAGTAAGGTTGCAGCTAAAACTGGGG
```

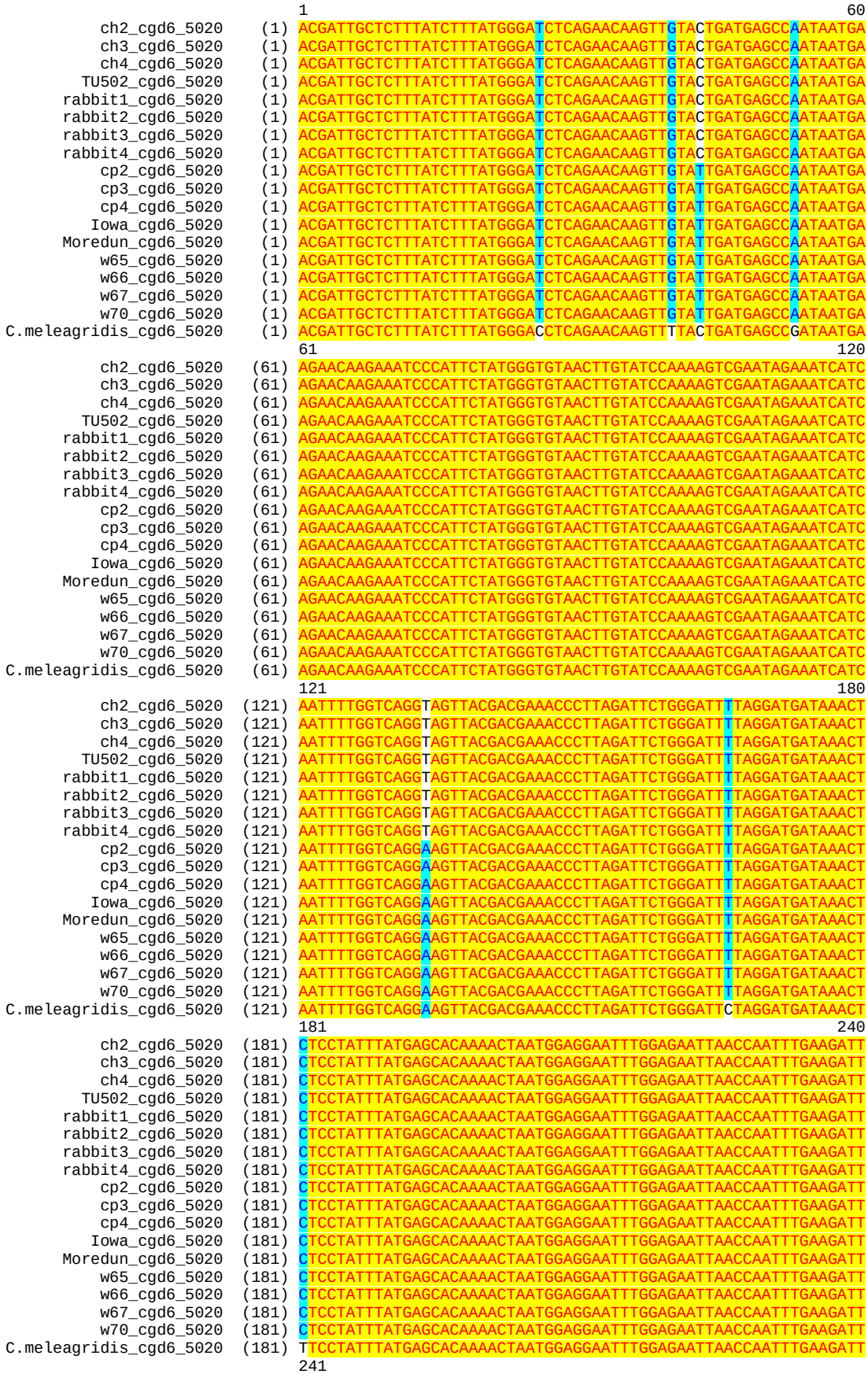
```
rabbit3_chro.50330  (301) ACAAGCCAAGTACAACTTAAAAGTAAGGTTGCAGCTAAAACTGGGG
rabbit4_chro.50330  (301) ACAAGCCAAGTACAACTTAAAAGTAAGGTTGCAGCTAAAACTGGGG
```

Chro.50457 gene PCR products

```
                                1                                                          60
      ch2_chro.50457      (1) TCCCGAATATAGTATTAATAATGATAAACTCGAGCTTGATAAATTAAAAGAAAATTATTC
      ch3_chro.50457      (1) TCCCGAATATAGTATTAATAATGATAAACTCGAGCTTGATAAATTAAAAGAAAATTATTC
      ch4_chro.50457      (1) TCCCGAATATAGTATTAATAATGATAAACTCGAGCTTGATAAATTAAAAGAAAATTATTC
    TU502_chro.50457      (1) TCCCGAATATAGTATTAATAATGATAAACTCGAGCTTGATAAATTAAAAGAAAATTATTC
      cp2_chro.50457      (1) TCCCGAATATAGTATTAATAATGATAAACTCGAGCTTGATAAATCAAAGGAAAATTATTC
      cp3_chro.50457      (1) TCCCGAATATAGTATTAATAATGATAAACTCGAGCTTGATAAATCAAAGGAAAATTATTC
      cp4_chro.50457      (1) TCCCGAATATAGTATTAATAATGATAAACTCGAGCTTGATAAATCAAAGGAAAATTATTC
     Iowa_chro.50457      (1) TCCCGAATATAGTATTAATAATGATAAACTCGAGCTTGATAAATCAAAGGAAAATTATTC
  Moredun_chro.50457      (1) TCCCGAATATAGTATTAATAATGATAAACTCGAGCTTGATAAATCAAAGGAAAATTATTC
      w65_chro.50457      (1) TCCCGAATATAGTATTAATAATGATAAACTCGAGCTTGATAAATCAAAGGAAAATTATTC
      w66_chro.50457      (1) TCCCGAATATAGTATTAATAATGATAAACTCGAGCTTGATAAATCAAAGGAAAATTATTC
      w67_chro.50457      (1) TCCCGAATATAGTATTAATAATGATAAACTCGAGCTTGATAAATCAAAGGAAAATTATTC
      w70_chro.50457      (1) TCCCGAATATAGTATTAATAATGATAAACTCGAGCTTGATAAATCAAAGGAAAATTATTC
  rabbit1_chro.50457      (1) TCCCGAATATAGTATTAATAATGATAAACTCGAGCTTGATAAATTAAAAGAAAATTATTC
  rabbit2_chro.50457      (1) TCCCGAATATAGTATTAATAATGATAAACTCGAGCTTGATAAATTAAAAGAAAATTATTC
  rabbit3_chro.50457      (1) TCCCGAATATAGTATTAATAATGATAAACTCGAGCTTGATAAATTAAAAGAAAATTATTC
  rabbit4_chro.50457      (1) TCCCGAATATAGTATTAATAATGATAAACTCGAGCTTGATAAATTAAAAGAAAATTATTC
C.meleagridis_chro.50457  (1) TCCCGAATATAACATTAATAATGATAAGCTTGAGCTTAATAAGACAAAAATAAATTATTC
                              61                                                        120
      ch2_chro.50457     (61) ACAAAATAATGAATTACTTCCATATTCATCTTTATGCGTTGTTAAAAAGAAGCCTCTCCC
      ch3_chro.50457     (61) ACAAAATAATGAATTACTTCCATATTCATCTTTATGCGTTGTTAAAAAGAAGCCTCTCCC
      ch4_chro.50457     (61) ACAAAATAATGAATTACTTCCATATTCATCTTTATGCGTTGTTAAAAAGAAGCCTCTCCC
    TU502_chro.50457     (61) ACAAAATAATGAATTACTTCCATATTCATCTTTATGCGTTGTTAAAAAGAAGCCTCTCCC
      cp2_chro.50457     (61) ACAAAATAATGAATTACTTCCATATTCATCTTTATGCATTGTTAAAAAGAAGCCTCTCCC
      cp3_chro.50457     (61) ACAAAATAATGAATTACTTCCATATTCATCTTTATGCATTGTTAAAAAGAAGCCTCTCCC
      cp4_chro.50457     (61) ACAAAATAATGAATTACTTCCATATTCATCTTTATGCATTGTTAAAAAGAAGCCTCTCCC
     Iowa_chro.50457     (61) ACAAAATAATGAATTACTTCCATATTCATCTTTATGCATTGTTAAAAAGAAGCCTCTCCC
  Moredun_chro.50457     (61) ACAAAATAATGAATTACTTCCATATTCATCTTTATGCATTGTTAAAAAGAAGCCTCTCCC
      w65_chro.50457     (61) ACAAAATAATGAATTACTTCCATATTCATCTTTATGCATTGTTAAAAAGAAGCCTCTCCC
      w66_chro.50457     (61) ACAAAATAATGAATTACTTCCATATTCATCTTTATGCATTGTTAAAAAGAAGCCTCTCCC
      w67_chro.50457     (61) ACAAAATAATGAATTACTTCCATATTCATCTTTATGCATTGTTAAAAAGAAGCCTCTCCC
      w70_chro.50457     (61) ACAAAATAATGAATTACTTCCATATTCATCTTTATGCATTGTTAAAAAGAAGCCTCTCCC
  rabbit1_chro.50457     (61) ACAAAATAATGAATTACTTCCATATTCATCTTTATGCGTTGTTAAAAAGAAGCCTCTCCC
  rabbit2_chro.50457     (61) ACAAAATAATGAATTACTTCCATATTCATCTTTATGCGTTGTTAAAAAGAAGCCTCTCCC
  rabbit3_chro.50457     (61) ACAAAATAATGAATTACTTCCATATTCATCTTTATGCGTTGTTAAAAAGAAGCCTCTCCC
  rabbit4_chro.50457     (61) ACAAAATAATGAATTACTTCCATATTCATCTTTATGCGTTGTTAAAAAGAAGCCTCTCCC
C.meleagridis_chro.50457 (61) ACAAAATAATGAGTTACTTCCATACTCATCCTTGTGTGTTGTTAAAAGGAAGCCTCTCCC
                              121                                                       180
      ch2_chro.50457    (121) TGTTATATCTTTTTGGTTAAGTCAAACTGATTCGAATGATAATAAAGTCGAGCCTATAAA
      ch3_chro.50457    (121) TGTTATATCTTTTTGGTTAAGTCAAACTGATTCGAATGATAATAAAGTCGAGCCTATAAA
      ch4_chro.50457    (121) TGTTATATCTTTTTGGTTAAGTCAAACTGATTCGAATGATAATAAAGTCGAGCCTATAAA
    TU502_chro.50457    (121) TGTTATATCTTTTTGGTTAAGTCAAACTGATTCGAATGATAATAAAGTCGAGCCTATAAA
      cp2_chro.50457    (121) TGTTATATCTTTTTGGTTAAGTCAAACTGATTCAAATGATGGTAAAGTTGAGCCTATAAA
      cp3_chro.50457    (121) TGTTATATCTTTTTGGTTAAGTCAAACTGATTCAAATGATGGTAAAGTTGAGCCTATAAA
      cp4_chro.50457    (121) TGTTATATCTTTTTGGTTAAGTCAAACTGATTCAAATGATGGTAAAGTTGAGCCTATAAA
     Iowa_chro.50457    (121) TGTTATATCTTTTTGGTTAAGTCAAACTGATTCAAATGATGGTAAAGTTGAGCCTATAAA
  Moredun_chro.50457    (121) TGTTATATCTTTTTGGTTAAGTCAAACTGATTCAAATGATGGTAAAGTTGAGCCTATAAA
      w65_chro.50457    (121) TGTTATATCTTTTTGGTTAAGTCAAACTGATTCAAATGATGGTAAAGTTGAGCCTATAAA
      w66_chro.50457    (121) TGTTATATCTTTTTGGTTAAGTCAAACTGATTCAAATGATGGTAAAGTTGAGCCTATAAA
      w67_chro.50457    (121) TGTTATATCTTTTTGGTTAAGTCAAACTGATTCAAATGATGGTAAAGTTGAGCCTATAAA
      w70_chro.50457    (121) TGTTATATCTTTTTGGTTAAGTCAAACTGATTCAAATGATGGTAAAGTTGAGCCTATAAA
  rabbit1_chro.50457    (121) TGTTATATCTTTTTGGTTAAGTCAAACTGATTCGAATGATAATAAAGTCGAGCCTATAAA
  rabbit2_chro.50457    (121) TGTTATATCTTTTTGGTTAAGTCAAACTGATTCGAATGATAATAAAGTCGAGCCTATAAA
  rabbit3_chro.50457    (121) TGTTATATCTTTTTGGTTAAGTCAAACTGATTCGAATGATAATAAAGTCGAGCCTATAAA
  rabbit4_chro.50457    (121) TGTTATATCTTTTTGGTTAAGTCAAACTGATTCGAATGATAATAAAGTCGAGCCTATAAA
C.meleagridis_chro.50457(121) TGTTATATCTTTTTGGTTAAGCCAAGCTGATTCAAATAATGATAAAGTGGAGCCTATGAA
                              181                                                       240
      ch2_chro.50457    (181) GATCTTTTTGGATGATTATATTATTCATTTTAATAATATTGACTATCTATGTGTAATCAA
      ch3_chro.50457    (181) GATCTTTTTGGATGATTATATTATTCATTTTAATAATATTGACTATCTATGTGTAATCAA
      ch4_chro.50457    (181) GATCTTTTTGGATGATTATATTATTCATTTTAATAATATTGACTATCTATGTGTAATCAA
    TU502_chro.50457    (181) GATCTTTTTGGATGATTATATTATTCATTTTAATAATATTGACTATCTATGTGTAATCAA
      cp2_chro.50457    (181) GATCTTTTTGGATGATTATATTATTCATTTTAATAATATTGACTATCTATGTGTAATTAA
      cp3_chro.50457    (181) GATCTTTTTGGATGATTATATTATTCATTTTAATAATATTGACTATCTATGTGTAATTAA
      cp4_chro.50457    (181) GATCTTTTTGGATGATTATATTATTCATTTTAATAATATTGACTATCTATGTGTAATTAA
     Iowa_chro.50457    (181) GATCTTTTTGGATGATTATATTATTCATTTTAATAATATTGACTATCTATGTGTAATTAA
  Moredun_chro.50457    (181) GATCTTTTTGGATGATTATATTATTCATTTTAATAATATTGACTATCTATGTGTAATTAA
      w65_chro.50457    (181) GATCTTTTTGGATGATTATATTATTCATTTTAATAATATTGACTATCTATGTGTAATTAA
      w66_chro.50457    (181) GATCTTTTTGGATGATTATATTATTCATTTTAATAATATTGACTATCTATGTGTAATTAA
      w67_chro.50457    (181) GATCTTTTTGGATGATTATATTATTCATTTTAATAATATTGACTATCTATGTGTAATTAA
      w70_chro.50457    (181) GATCTTTTTGGATGATTATATTATTCATTTTAATAATATTGACTATCTATGTGTAATTAA
  rabbit1_chro.50457    (181) GATCTTTTTGGATGATTATATTATTCATTTTAATAATATTGACTATCTATGTGTAATCAA
  rabbit2_chro.50457    (181) GATCTTTTTGGATGATTATATTATTCATTTTAATAATATTGACTATCTATGTGTAATCAA
```


rabbit3_chro.50457 (181) GATCTTTTTGGATGATTATATTATTCATTTTAATAATATTGACTATCTATGTGTAATCAA
rabbit4_chro.50457 (181) GATCTTTTTGGATGATTATATTATTCATTTTAATAATATTGACTATCTATGTGTAATCAA
C.meleagridis_chro.50457 (181) GATCTTTTTGGATGACTATATTATTCATTTTAATAATATTGACTATTTGTGTGTAATTAA
241 300
ch2_chro.50457 (241) TTCAACACTAGGTACTTTTGGAAAAGTTAATTACATACCTTACAGTGAAGGTAATAGTTA
ch3_chro.50457 (241) TTCAACACTAGGTACTTTTGGAAAAGTTAATTACATACCTTACAGTGAAGGTAATAGTTA
ch4_chro.50457 (241) TTCAACACTAGGTACTTTTGGAAAAGTTAATTACATACCTTACAGTGAAGGTAATAGTTA
TU502_chro.50457 (241) TTCAACACTAGGTACTTTTGGAAAAGTTAATTACATACCTTACAGTGAAGGTAATAGTTA
cp2_chro.50457 (241) TTCAACACTAGGTACTTTTGGAAAAATTAATTACATACCTTACAGTGAAGGTAATAGTTA
cp3_chro.50457 (241) TTCAACACTAGGTACTTTTGGAAAAATTAATTACATACCTTACAGTGAAGGTAATAGTTA
cp4_chro.50457 (241) TTCAACACTAGGTACTTTTGGAAAAATTAATTACATACCTTACAGTGAAGGTAATAGTTA
Iowa_chro.50457 (241) TTCAACACTAGGTACTTTTGGAAAAATTAATTACATACCTTACAGTGAAGGTAATAGTTA
Moredun_chro.50457 (241) TTCAACACTAGGTACTTTTGGAAAAATTAATTACATACCTTACAGTGAAGGTAATAGTTA
w65_chro.50457 (241) TTCAACACTAGGTACTTTTGGAAAAATTAATTACATACCTTACAGTGAAGGTAATAGTTA
w66_chro.50457 (241) TTCAACACTAGGTACTTTTGGAAAAATTAATTACATACCTTACAGTGAAGGTAATAGTTA
w67_chro.50457 (241) TTCAACACTAGGTACTTTTGGAAAAATTAATTACATACCTTACAGTGAAGGTAATAGTTA
w70_chro.50457 (241) TTCAACACTAGGTACTTTTGGAAAAATTAATTACATACCTTACAGTGAAGGTAATAGTTA
rabbit1_chro.50457 (241) TTCAACACTAGGTACTTTTGGAAAAGTTAATTACATACCTTACAGTGAAGGTAATAGTTA
rabbit2_chro.50457 (241) TTCAACACTAGGTACTTTTGGAAAAGTTAATTACATACCTTACAGTGAAGGTAATAGTTA
rabbit3_chro.50457 (241) TTCAACACTAGGTACTTTTGGAAAAGTTAATTACATACCTTACAGTGAAGGTAATAGTTA
rabbit4_chro.50457 (241) TTCAACACTAGGTACTTTTGGAAAAGTTAATTACATACCTTACAGTGAAGGTAATAGTTA
C.meleagridis_chro.50457 (241) CTCAACTCTAGGTACTTTTGGAAAGGTCAACTACATACCTTATAGTGAAGGTAATAGTTA
301 360
ch2_chro.50457 (301) TAACTACTTCACTCCTTATTATTCATCCTCTAATACTCCCTTGATACTCTTTGGTTCTCT
ch3_chro.50457 (301) TAACTACTTCACTCCTTATTATTCATCCTCTAATACTCCCTTGATACTCTTTGGTTCTCT
ch4_chro.50457 (301) TAACTACTTCACTCCTTATTATTCATCCTCTAATACTCCCTTGATACTCTTTGGTTCTCT
TU502_chro.50457 (301) TAACTACTTCACTCCTTATTATTCATCCTCTAATACTCCCTTGATACTCTTTGGTTCTCT
cp2_chro.50457 (301) TAACTACTTTACTCCTTATTATTCATCCTCTAATACTCCTTTGATACTCCTTGGTTCTCT
cp3_chro.50457 (301) TAACTACTTTACTCCTTATTATTCATCCTCTAATACTCCTTTGATACTCCTTGGTTCTCT
cp4_chro.50457 (301) TAACTACTTTACTCCTTATTATTCATCCTCTAATACTCCTTTGATACTCCTTGGTTCTCT
Iowa_chro.50457 (301) TAACTACTTTACTCCTTATTATTCATCCTCTAATACTCCTTTGATACTCCTTGGTTCTCT
Moredun_chro.50457 (301) TAACTACTTTACTCCTTATTATTCATCCTCTAATACTCCTTTGATACTCCTTGGTTCTCT
w65_chro.50457 (301) TAACTACTTTACTCCTTATTATTCATCCTCTAATACTCCTTTGATACTCCTTGGTTCTCT
w66_chro.50457 (301) TAACTACTTTACTCCTTATTATTCATCCTCTAATACTCCTTTGATACTCCTTGGTTCTCT
w67_chro.50457 (301) TAACTACTTTACTCCTTATTATTCATCCTCTAATACTCCTTTGATACTCCTTGGTTCTCT
w70_chro.50457 (301) TAACTACTTTACTCCTTATTATTCATCCTCTAATACTCCTTTGATACTCCTTGGTTCTCT
rabbit1_chro.50457 (301) TAACTACTTCACTCCTTATTATTCATCCTCTAATACTCCCTTGATACTCTTTGGTTCTCT
rabbit2_chro.50457 (301) TAACTACTTCACTCCTTATTATTCATCCTCTAATACTCCCTTGATACTCTTTGGTTCTCT
rabbit3_chro.50457 (301) TAACTACTTCACTCCTTATTATTCATCCTCTAATACTCCCTTGATACTCTTTGGTTCTCT
rabbit4_chro.50457 (301) TAACTACTTCACTCCTTATTATTCATCCTCTAATACTCCCTTGATACTCTTTGGTTCTCT
C.meleagridis_chro.50457 (301) TAACTATTTTACCCCGTATTATTCATCCTCCAATGCTCCTTTGATACTCTTTGGTTCTCT
361 373
ch2_chro.50457 (361) TTTTCTCAAATCA
ch3_chro.50457 (361) TTTTCTCAAATCA
ch4_chro.50457 (361) TTTTCTCAAATCA
TU502_chro.50457 (361) TTTTCTCAAATCA
cp2_chro.50457 (361) TTTTCTCAAATCA
cp3_chro.50457 (361) TTTTCTCAAATCA
cp4_chro.50457 (361) TTTTCTCAAATCA
Iowa_chro.50457 (361) TTTTCTCAAATCA
Moredun_chro.50457 (361) TTTTCTCAAATCA
w65_chro.50457 (361) TTTTCTCAAATCA
w66_chro.50457 (361) TTTTCTCAAATCA
w67_chro.50457 (361) TTTTCTCAAATCA
w70_chro.50457 (361) TTTTCTCAAATCA
rabbit1_chro.50457 (361) TTTTCTCAAATCA
rabbit2_chro.50457 (361) TTTTCTCAAATCA
rabbit3_chro.50457 (361) TTTTCTCAAATCA
rabbit4_chro.50457 (361) TTTTCTCAAATCA
C.meleagridis_chro.50457 (361) TTTTCTCAAATCA

Cgd6_5020 gene PCR products



```
           ch2_cgd6_5020  (241) ACCTTGGA
           ch3_cgd6_5020  (241) ACCTTGGA
           ch4_cgd6_5020  (241) ACCTTGGA
         TU502_cgd6_5020  (241) ACCTTGGA
       rabbit1_cgd6_5020  (241) ACCTTGGA
       rabbit2_cgd6_5020  (241) ACCTTGGA
       rabbit3_cgd6_5020  (241) ACCTTGGA
       rabbit4_cgd6_5020  (241) ACCTTGGA
           cp2_cgd6_5020  (241) ACCTTGGA
           cp3_cgd6_5020  (241) ACCTTGGA
           cp4_cgd6_5020  (241) ACCTTGGA
          Iowa_cgd6_5020  (241) ACCTTGGA
       Moredun_cgd6_5020  (241) ACCTTGGA
           w65_cgd6_5020  (241) ACCTTGGA
           w66_cgd6_5020  (241) ACCTTGGA
           w67_cgd6_5020  (241) ACCTTGGA
           w70_cgd6_5020  (241) ACCTTGGA
C.meleagridis_cgd6_5020  (241) ACCTTGGA
```

COWP gene PCR products

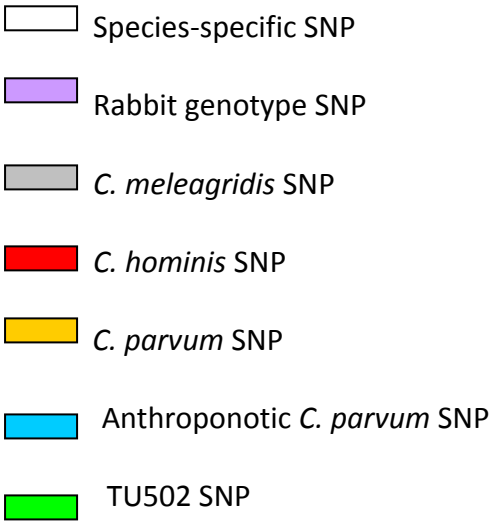
```
                          1                                                  60
           ch2_COWP    (1) CAGGCATTATCTTGTAATACTGTACCTGGAGGGCAGACAGGTTGAGTTGGAGCAGAACTA
           ch3_COWP    (1) CAGGCATTATCTTGTAATACTGTACCTGGAGGGCAGACAGGTTGAGTTGGAGCAGAACTA
           ch4_COWP    (1) CAGGCATTATCTTGTAATACTGTACCTGGAGGGCAGACAGGTTGAGTTGGAGCAGAACTA
         TU502_COWP    (1) CAGGCATTATCTTGTAATACTGTACCTGGAGGGCAGACAGGTTGAGTTGGAGCAGAACTA
       rabbit1_COWP    (1) CAGGCATTATCTTGTAATACTGTACCTGGAGGGCAGACAGGTTGAGTTGGAGCAGAACTA
       rabbit2_COWP    (1) CAGGCATTATCTTGTAATACTGTACCTGGAGGGCAGACAGGTTGAGTTGGAGCAGAACTA
       rabbit3_COWP    (1) CAGGCATTATCTTGTAATACTGTACCTGGAGGGCAGACAGGTTGAGTTGGAGCAGAACTA
       rabbit4_COWP    (1) CAGGCATTATCTTGTAATACTGTACCTGGAGGGCAGACAGGTTGAGTTGGAGCAGAACTA
           cp2_COWP    (1) CAGGCATTATCTTGTAATACTGTACCTGGAGGGCAGACAGGTTGAGTTGGAGCAGAACTA
           cp3_COWP    (1) CAGGCATTATCTTGTAATACTGTACCTGGAGGGCAGACAGGTTGAGTTGGAGCAGAACTA
           cp4_COWP    (1) CAGGCATTATCTTGTAATACTGTACCTGGAGGGCAGACAGGTTGAGTTGGAGCAGAACTA
          Iowa_COWP    (1) CAGGCATTATCTTGTAATACTGTACCTGGAGGGCAGACAGGTTGAGTTGGAGCAGAACTA
       Moredun_COWP    (1) CAGGCATTATCTTGTAATACTGTACCTGGAGGGCAGACAGGTTGAGTTGGAGCAGAACTA
           w65_COWP    (1) CAGGCATTATCTTGTAATACTGTACCTGGAGGGCAGACAGGTTGAGTTGGAGCAGAACTA
           w66_COWP    (1) CAGGCATTATCTTGTAATACTGTACCTGGAGGGCAGACAGGTTGAGTTGGAGCAGAACTA
           w67_COWP    (1) CAGGCATTATCTTGTAATACTGTACCTGGAGGGCAGACAGGTTGAGTTGGAGCAGAACTA
           w70_COWP    (1) CAGGCATTATCTTGTAATACTGTACCTGGAGGGCAGACAGGTTGAGTTGGAGCAGAACTA
C.meleagridis_COWP     (1) CAGGCATTATCTTGTAGCACTGTACCTGTAGGGCATACAGGTTGTGTTGGAGCGGAACTA
                          61                                                 120
           ch2_COWP   (61) GTTTGAATACACTGGAGGCCTTGTAAAATGAAATTTGGTGGGCATTCCTTTGCAGGAGCT
           ch3_COWP   (61) GTTTGAATACACTGGAGGCCTTGTAAAATGAAATTTGGTGGGCATTCCTTTGCAGGAGCT
           ch4_COWP   (61) GTTTGAATACACTGGAGGCCTTGTAAAATGAAATTTGGTGGGCATTCCTTTGCAGGAGCT
         TU502_COWP   (61) GTTTGAATACACTGGAGGCCTTGTAAAATGAAATTTGGTGGGCATTCCTTTGCAGGAGCT
       rabbit1_COWP   (61) GTTTGAATACACTGGAGGCCTTGTAAAATGAAATTTGGTGGGCATTCCTTTGCAGGAGCT
       rabbit2_COWP   (61) GTTTGAATACACTGGAGGCCTTGTAAAATGAAATTTGGTGGGCATTCCTTTGCAGGAGCT
       rabbit3_COWP   (61) GTTTGAATACACTGGAGGCCTTGTAAAATGAAATTTGGTGGGCATTCCTTTGCAGGAGCT
       rabbit4_COWP   (61) GTTTGAATACACTGGAGGCCTTGTAAAATGAAATTTGGTGGGCATTCCTTTGCAGGAGCT
           cp2_COWP   (61) GTTTGTATACATTGGAGGCCTTGTAAAATGAAATTTGGTGGGCATTCCTTTGCAGGAGCT
           cp3_COWP   (61) GTTTGTATACATTGGAGGCCTTGTAAAATGAAATTTGGTGGGCATTCCTTTGCAGGAGCT
           cp4_COWP   (61) GTTTGTATACATTGGAGGCCTTGTAAAATGAAATTTGGTGGGCATTCCTTTGCAGGAGCT
          Iowa_COWP   (61) GTTTGTATACATTGGAGGCCTTGTAAAATGAAATTTGGTGGGCATTCCTTTGCAGGAGCT
       Moredun_COWP   (61) GTTTGTATACATTGGAGGCCTTGTAAAATGAAATTTGGTGGGCATTCCTTTGCAGGAGCT
           w65_COWP   (61) GTTTGTATACATTGGAGGCCTTGTAAAATGAAATTTGGTGGGCATTCCTTTGCAGGAGCT
           w66_COWP   (61) GTTTGTATACATTGGAGGCCTTGTAAAATGAAATTTGGTGGGCATTCCTTTGCAGGAGCT
           w67_COWP   (61) GTTTGTATACATTGGAGGCCTTGTAAAATGAAATTTGGTGGGCATTCCTTTGCAGGAGCT
           w70_COWP   (61) GTTTGTATACATTGGAGGCCTTGTAAAATGAAATTTGGTGGGCATTCCTTTGCAGGAGCT
C.meleagridis_COWP    (61) GTTTGAATACACTGGAGGCCTTGTAATATGAAATTTGGTGGGCATTCCTTTGCAGGAGCT
                         121                                                 180
           ch2_COWP  (121) ACATATAGTACACAATCATCTCCTGAATCTGTATATCCTGGTGGGCAGACCATATCAATA
           ch3_COWP  (121) ACATATAGTACACAATCATCTCCTGAATCTGTATATCCTGGTGGGCAGACCATATCAATA
           ch4_COWP  (121) ACATATAGTACACAATCATCTCCTGAATCTGTATATCCTGGTGGGCAGACCATATCAATA
         TU502_COWP  (121) ACATATAGTACACAATCATCTCCTGAATCTGTATATCCTGGTGGGCAGACCATATCAATA
```

```
       rabbit1_COWP  (121) ACATATAGTACACAATCATCTCCTGAATCTGTATATCCTGGTGGGCAGACCATATCAATA
       rabbit2_COWP  (121) ACATATAGTACACAATCATCTCCTGAATCTGTATATCCTGGTGGGCAGACCATATCAATA
       rabbit3_COWP  (121) ACATATAGTACACAATCATCTCCTGAATCTGTATATCCTGGTGGGCAGACCATATCAATA
       rabbit4_COWP  (121) ACATATAGTACACAATCATCTCCTGAATCTGTATATCCTGGTGGGCAGACCATATCAATA
           cp2_COWP  (121) ACATATAGTACACAATCATCTCCTGAGTCTGTATATCCTGGTGGGCAGACCATATCAATA
           cp3_COWP  (121) ACATATAGTACACAATCATCTCCTGAGTCTGTATATCCTGGTGGGCAGACCATATCAATA
           cp4_COWP  (121) ACATATAGTACACAATCATCTCCTGAGTCTGTATATCCTGGTGGGCAGACCATATCAATA
          Iowa_COWP  (121) ACATATAGTACACAATCATCTCCTGAGTCTGTATATCCTGGTGGGCAGACCATATCAATA
       Moredun_COWP  (121) ACATATAGTACACAATCATCTCCTGAGTCTGTATATCCTGGTGGGCAGACCATATCAATA
           w65_COWP  (121) ACATATAGTACACAATCATCTCCTGAGTCTGTATATCCTGGTGGGCAGACCATATCAATA
           w66_COWP  (121) ACATATAGTACACAATCATCTCCTGAGTCTGTATATCCTGGTGGGCAGACCATATCAATA
           w67_COWP  (121) ACATATAGTACACAATCATCTCCTGAGTCTGTATATCCTGGTGGGCAGACCATATCAATA
           w70_COWP  (121) ACATATAGTACACAATCATCTCCTGAGTCTGTATATCCTGGTGGGCAGACCATATCAATA
 C.meleagridis_COWP  (121) ACATATAGAACACAATCATCTCCTGAGTCTGTATATCCTGGTGGGCAGACCATATCAATA
                           181                                                       240
           ch2_COWP  (181) TTTTTAATCACTTTACATTTTCCATTTTCAAATATTGAATTAGGTGGGCATGTCGATTCT
           ch3_COWP  (181) TTTTTAATCACTTTACATTTTCCATTTTCAAATATTGAATTAGGTGGGCATGTCGATTCT
           ch4_COWP  (181) TTTTTAATCACTTTACATTTTCCATTTTCAAATATTGAATTAGGTGGGCATGTCGATTCT
         TU502_COWP  (181) TTTTTAATCACTTTACATTTTCCATTTTCAAATATTGAATTAGGTGGGCATGTCGATTCT
       rabbit1_COWP  (181) TTTTTAATCACTTTACATTTTCCATTTTCAAATATTGAATTAGGTGGGCATGTCGATTCT
       rabbit2_COWP  (181) TTTTTAATCACTTTACATTTTCCATTTTCAAATATTGAATTAGGTGGGCATGTCGATTCT
       rabbit3_COWP  (181) TTTTTAATCACTTTACATTTTCCATTTTCAAATATTGAATTAGGTGGGCATGTCGATTCT
       rabbit4_COWP  (181) TTTTTAATCACTTTACATTTTCCATTTTCAAATATTGAATTAGGTGGGCATGTCGATTCT
           cp2_COWP  (181) TTTTTAATCACTTTACATTTTCCATTTTCAAATATTGAATTAGGTGGGCATGTCGATTCT
           cp3_COWP  (181) TTTTTAATCACTTTACATTTTCCATTTTCAAATATTGAATTAGGTGGGCATGTCGATTCT
           cp4_COWP  (181) TTTTTAATCACTTTACATTTTCCATTTTCAAATATTGAATTAGGTGGGCATGTCGATTCT
          Iowa_COWP  (181) TTTTTAATCACTTTACATTTTCCATTTTCAAATATTGAATTAGGTGGGCATGTCGATTCT
       Moredun_COWP  (181) TTTTTAATCACTTTACATTTTCCATTTTCAAATATTGAATTAGGTGGGCATGTCGATTCT
           w65_COWP  (181) TTTTTAATCACTTTACATTTTCCATTTTCAAATATTGAATTAGGTGGGCATGTCGATTCT
           w66_COWP  (181) TTTTTAATCACTTTACATTTTCCATTTTCAAATATTGAATTAGGTGGGCATGTCGATTCT
           w67_COWP  (181) TTTTTAATCACTTTACATTTTCCATTTTCAAATATTGAATTAGGTGGGCATGTCGATTCT
           w70_COWP  (181) TTTTTAATCACTTTACATTTTCCATTTTCAAATATTGAATTAGGTGGGCATGTCGATTCT
 C.meleagridis_COWP  (181) TTTTTAATCACTTTGCATTTTCCATTTTCAAATATTGAATTAGGTGGGCATGTAGATTCT
                           241                                                       300
           ch2_COWP  (241) AATTCAGCTGATTCTGGTGCCATACATTGTTGTCCTGACAAATTGAATCCAGGAGGACAG
           ch3_COWP  (241) AATTCAGCTGATTCTGGTGCCATACATTGTTGTCCTGACAAATTGAATCCAGGAGGACAG
           ch4_COWP  (241) AATTCAGCTGATTCTGGTGCCATACATTGTTGTCCTGACAAATTGAATCCAGGAGGACAG
         TU502_COWP  (241) AATTCAGCTGATTCTGGTGCCATACATTGTTGTCCTGACAAATTGAATCCAGGAGGACAG
       rabbit1_COWP  (241) AATTCAGCTGATTCTGGTGCCATACATTGTTGTCCTGACAAATTGAATCCAGGAGGACAG
       rabbit2_COWP  (241) AATTCAGCTGATTCTGGTGCCATACATTGTTGTCCTGACAAATTGAATCCAGGAGGACAG
       rabbit3_COWP  (241) AATTCAGCTGATTCTGGTGCCATACATTGTTGTCCTGACAAATTGAATCCAGGAGGACAG
       rabbit4_COWP  (241) AATTCAGCTGATTCTGGTGCCATACATTGTTGTCCTGACAAATTGAATCCAGGAGGACAG
           cp2_COWP  (241) AATTCAGCTGATTCTGGTGCCATACATTGTTGTCCTGACAAATTGAATCCAGGAGGACAG
           cp3_COWP  (241) AATTCAGCTGATTCTGGTGCCATACATTGTTGTCCTGACAAATTGAATCCAGGAGGACAG
           cp4_COWP  (241) AATTCAGCTGATTCTGGTGCCATACATTGTTGTCCTGACAAATTGAATCCAGGAGGACAG
          Iowa_COWP  (241) AATTCAGCTGATTCTGGTGCCATACATTGTTGTCCTGACAAATTGAATCCAGGAGGACAG
       Moredun_COWP  (241) AATTCAGCTGATTCTGGTGCCATACATTGTTGTCCTGACAAATTGAATCCAGGAGGACAG
           w65_COWP  (241) AATTCAGCTGATTCTGGTGCCATACATTGTTGTCCTGACAAATTGAATCCAGGAGGACAG
           w66_COWP  (241) AATTCAGCTGATTCTGGTGCCATACATTGTTGTCCTGACAAATTGAATCCAGGAGGACAG
           w67_COWP  (241) AATTCAGCTGATTCTGGTGCCATACATTGTTGTCCTGACAAATTGAATCCAGGAGGACAG
           w70_COWP  (241) AATTCAGCTGATTCTGGTGCCATACATTGTTGTCCTGACAAATTGAATCCAGGAGGACAG
 C.meleagridis_COWP  (241) AATTCAGCAGATTCTGGTGCCATACATTGTTGTCCTGACAAATTAAATCCAGGAGGACAG
                           301                                                       360
           ch2_COWP  (301) ATTTTGTTTGCAGGGAGATATTGAACACATCTATTTCCTTCTTCAACAAAACCAGAAGGA
           ch3_COWP  (301) ATTTTGTTTGCAGGGAGATATTGAACACATCTATTTCCTTCTTCAACAAAACCAGAAGGA
           ch4_COWP  (301) ATTTTGTTTGCAGGGAGATATTGAACACATCTATTTCCTTCTTCAACAAAACCAGAAGGA
         TU502_COWP  (301) ATTTTGTTTGCAGGGAGATATTGAACACATCTATTTCCTTCTTCAACAAAACCAGAAGGA
       rabbit1_COWP  (301) ATTTTGTTTGCAGGGAGATATTGAACACATCTATTTCCTTCTTCAACAAAACCAGAAGGA
       rabbit2_COWP  (301) ATTTTGTTTGCAGGGAGATATTGAACACATCTATTTCCTTCTTCAACAAAACCAGAAGGA
       rabbit3_COWP  (301) ATTTTGTTTGCAGGGAGATATTGAACACATCTATTTCCTTCTTCAACAAAACCAGAAGGA
       rabbit4_COWP  (301) ATTTTGTTTGCAGGGAGATATTGAACACATCTATTTCCTTCTTCAACAAAACCAGAAGGA
           cp2_COWP  (301) ATTTTATTTGCAGGGAGATATTGAACACATCTATTTCCTTCTTCAACAAAACCAGAAGGA
           cp3_COWP  (301) ATTTTATTTGCAGGGAGATATTGAACACATCTATTTCCTTCTTCAACAAAACCAGAAGGA
           cp4_COWP  (301) ATTTTATTTGCAGGGAGATATTGAACACATCTATTTCCTTCTTCAACAAAACCAGAAGGA
          Iowa_COWP  (301) ATTTTATTTGCAGGGAGATATTGAACACATCTATTTCCTTCTTCAACAAAACCAGAAGGA
       Moredun_COWP  (301) ATTTTATTTGCAGGGAGATATTGAACACATCTATTTCCTTCTTCAACAAAACCAGAAGGA
           w65_COWP  (301) ATTTTATTTGCAGGGAGATATTGAACACATCTATTTCCTTCTTCAACAAAACCAGAAGGA
           w66_COWP  (301) ATTTTATTTGCAGGGAGATATTGAACACATCTATTTCCTTCTTCAACAAAACCAGAAGGA
           w67_COWP  (301) ATTTTATTTGCAGGGAGATATTGAACACATCTATTTCCTTCTTCAACAAAACCAGAAGGA
           w70_COWP  (301) ATTTTATTTGCAGGGAGATATTGAACACATCTATTTCCTTCTTCAACAAAACCAGAAGGA
 C.meleagridis_COWP  (301) ATTTTGTTTGCAGGGAGATATTGAACACATCTATTTCCTTCTTCAACAAAACCAGAAGGA
                           361                                                       420
           ch2_COWP  (361) CAAACGGTATCAATTTGTTGAATTAATTTACATGTGCCATTTTCCAGTATAGTACCTGGA
           ch3_COWP  (361) CAAACGGTATCAATTTGTTGAATTAATTTACATGTGCCATTTTCCAGTATAGTACCTGGA
           ch4_COWP  (361) CAAACGGTATCAATTTGTTGAATTAATTTACATGTGCCATTTTCCAGTATAGTACCTGGA
         TU502_COWP  (361) CAAACGGTATCAATTTGTTGAATTAATTTACATGTGCCATTTTCCAGTATAGTACCTGGA
       rabbit1_COWP  (361) CAAACGGTATCAATTTGTTGAATTAATTTACATGTGCCATTTTCCAGTATAGTACCTGGA
       rabbit2_COWP  (361) CAAACGGTATCAATTTGTTGAATTAATTTACATGTGCCATTTTCCAGTATAGTACCTGGA
       rabbit3_COWP  (361) CAAACGGTATCAATTTGTTGAATTAATTTACATGTGCCATTTTCCAGTATAGTACCTGGA
       rabbit4_COWP  (361) CAAACGGTATCAATTTGTTGAATTAATTTACATGTGCCATTTTCCAGTATAGTACCTGGA
           cp2_COWP  (361) CAAACGGTATCAATTTGTTGAATTAATTTACATGTGCCATTCTCCAGTATAGTGCCTGGA
           cp3_COWP  (361) CAAACGGTATCAATTTGTTGAATTAATTTACATGTGCCATTCTCCAGTATAGTGCCTGGA
```

```
          cp4_COWP  (361) CAAACGGTATCAATTTGTTGAATTAATTTACATGTGCCATTCTCCAGTATAGTGCCTGGA
          Iowa_COWP  (361) CAAACGGTATCAATTTGTTGAATTAATTTACATGTGCCATTCTCCAGTATAGTGCCTGGA
        Moredun_COWP  (361) CAAACGGTATCAATTTGTTGAATTAATTTACATGTGCCATTCTCCAGTATAGTGCCTGGA
          w65_COWP  (361) CAAACGGTATCAATTTGTTGAATTAATTTACATGTGCCATTCTCCAGTATAGTGCCTGGA
          w66_COWP  (361) CAAACGGTATCAATTTGTTGAATTAATTTACATGTGCCATTCTCCAGTATAGTGCCTGGA
          w67_COWP  (361) CAAACGGTATCAATTTGTTGAATTAATTTACATGTGCCATTCTCCAGTATAGTGCCTGGA
          w70_COWP  (361) CAAACGGTATCAATTTGTTGAATTAATTTACATGTGCCATTCTCCAGTATAGTGCCTGGA
C.meleagridis_COWP  (361) CAAACGGTATCAATTTGTTGAATTAATTTACATGTACCATTCTCTAGTATAGTGCCTGGA
                          421                                                 480
          ch2_COWP  (421) GGACATTCTGGATTAGGAGGAGCTGTGTCTGATTGAACACATTGTTTTCCAGAAAAGACG
          ch3_COWP  (421) GGACATTCTGGATTAGGAGGAGCTGTGTCTGATTGAACACATTGTTTTCCAGAAAAGACG
          ch4_COWP  (421) GGACATTCTGGATTAGGAGGAGCTGTGTCTGATTGAACACATTGTTTTCCAGAAAAGACG
        TU502_COWP  (421) GGACATTCTGGATTAGGAGGAGCTGTGTCTGATTGAACACATTGTTTTCCAGAAAAGACG
        rabbit1_COWP  (421) GGACATTCTGGATTAGGAGGAGCTGTGTCTGATTGAACACATTGTTTTCCAGAAAAGACG
        rabbit2_COWP  (421) GGACATTCTGGATTAGGAGGAGCTGTGTCTGATTGAACACATTGTTTTCCAGAAAAGACG
        rabbit3_COWP  (421) GGACATTCTGGATTAGGAGGAGCTGTGTCTGATTGAACACATTGTTTTCCAGAAAAGACG
        rabbit4_COWP  (421) GGACATTCTGGATTAGGAGGAGCTGTGTCTGATTGAACACATTGTTTTCCAGAAAAGACG
          cp2_COWP  (421) GGACATTCTGGATTAGGAGGAGCTGTGTCTGATTGAACACATTGTTTTCCAGAAAAAACG
          cp3_COWP  (421) GGACATTCTGGATTAGGAGGAGCTGTGTCTGATTGAACACATTGTTTTCCAGAAAAAACG
          cp4_COWP  (421) GGACATTCTGGATTAGGAGGAGCTGTGTCTGATTGAACACATTGTTTTCCAGAAAAAACG
          Iowa_COWP  (421) GGACATTCTGGATTAGGAGGAGCTGTGTCTGATTGAACACATTGTTTTCCAGAAAAAACG
        Moredun_COWP  (421) GGACATTCTGGATTAGGAGGAGCTGTGTCTGATTGAACACATTGTTTTCCAGAAAAAACG
          w65_COWP  (421) GGACATTCTGGATTAGGAGGAGCTGTGTCTGATTGAACACATTGTTTTCCAGAAAAAACG
          w66_COWP  (421) GGACATTCTGGATTAGGAGGAGCTGTGTCTGATTGAACACATTGTTTTCCAGAAAAAACG
          w67_COWP  (421) GGACATTCTGGATTAGGAGGAGCTGTGTCTGATTGAACACATTGTTTTCCAGAAAAAACG
          w70_COWP  (421) GGACATTCTGGATTAGGAGGAGCTGTGTCTGATTGAACACATTGTTTTCCAGAAAAAACG
C.meleagridis_COWP  (421) GGACATTCTGGATTAGGAGGAGCTGTGTCTGATTGAACACATTGCTTTCCAGAAAAAACG
                          481                                                 529
          ch2_COWP  (481) AATCCTGGGGGACATGATTTTTCAGGCATAGTGAATGCAACACAATCTC
          ch3_COWP  (481) AATCCTGGGGGACATGATTTTTCAGGCATAGTGAATGCAACACAATCTC
          ch4_COWP  (481) AATCCTGGGGGACATGATTTTTCAGGCATAGTGAATGCAACACAATCTC
        TU502_COWP  (481) AATCCTGGGGGACATGATTTTTCAGGCATAGTGAATGCAACACAATCTC
        rabbit1_COWP  (481) AATCCTGGGGGACATGATTTTTCAGGCATAGTGAATGCAACACAATCTC
        rabbit2_COWP  (481) AATCCTGGGGGACATGATTTTTCAGGCATAGTGAATGCAACACAATCTC
        rabbit3_COWP  (481) AATCCTGGGGGACATGATTTTTCAGGCATAGTGAATGCAACACAATCTC
        rabbit4_COWP  (481) AATCCTGGGGGACATGATTTTTCAGGCATAGTGAATGCAACACAATCTC
          cp2_COWP  (481) AATCCTGGGGGACATGATTTTTCAGGCATAGTGAATGCAACACAATCTC
          cp3_COWP  (481) AATCCTGGGGGACATGATTTTTCAGGCATAGTGAATGCAACACAATCTC
          cp4_COWP  (481) AATCCTGGGGGACATGATTTTTCAGGCATAGTGAATGCAACACAATCTC
          Iowa_COWP  (481) AATCCTGGGGGACATGATTTTTCAGGCATAGTGAATGCAACACAATCTC
        Moredun_COWP  (481) AATCCTGGGGGACATGATTTTTCAGGCATAGTGAATGCAACACAATCTC
          w65_COWP  (481) AATCCTGGGGGACATGATTTTTCAGGCATAGTGAATGCAACACAATCTC
          w66_COWP  (481) AATCCTGGGGGACATGATTTTTCAGGCATAGTGAATGCAACACAATCTC
          w67_COWP  (481) AATCCTGGGGGACATGATTTTTCAGGCATAGTGAATGCAACACAATCTC
          w70_COWP  (481) AATCCTGGGGGACATGATTTTTCAGGCATAGTGAATGCAACACAATCTC
C.meleagridis_COWP  (481) AATCCTGGAGGACATGATTTTTCAGGCATAGTAAATGCAACACAATCTC
```

Appendix V

Appendix V: Summary of SNP type and position for each gene tested. Information for each gene was presented on each table, including gene size, chromosome number and location, expected PCR product size and number of high quality base pair sequence retrieved.
The SNPS were then used to identify species and subtypes pattern using colour legend. Each SNP was labelled based on its gene location and a colour was used to attribute the SNP to a species/subtype, the same colour legend was used to determine the species/subtype profile:



Cgd2_80 (4431 bp)
Chr2 (23.97k- 28.85k)
PCR product 266 bp
total number of bp 248

Isolates	34	41	115	151	154	160	163	200	210
Ch2	G	G	G	A	A	T	C	T	T
Ch3	G	G	G	A	A	T	C	T	T
Ch4	G	G	G	A	A	T	C	T	T
TU502	G	G	G	A	A	T	C	T	
Cp2	G	A	A	G	G	C	T	T	T
Cp3	G	A	A	G	G	C	T	T	T
Cp4	G	A	A	G	G	C	T	T	T
Iowa	G	A	A	G	G	C	T	T	T
Moredun	G	A	A	G	G	C	T	T	T
w65	G	A	A	G	G	C	T	T	T
w66	G	A	A	G	G	C	T	T	T
w67	G	A	A	G	G	C	T	T	T
w70	G	A	A	G	G	C	T	T	T
Rabbit1	A	G	G	A	A	T	C	T	T
Rabbit2	A	G	G	A	A	T	C	T	T
Rabbit3	A	G	G	A	A	T	C	T	T
Rabbit4	A	G	G	A	A	T	C	T	T
C.meleagridis	G	A	A	G	G	C	T	G	C

Cgd2_2430 (873 bp)
Chr2 (492.39k- 493.35k)
PCR product 389 bp
total number of bp 361

Isolates	40	73	74	79	94	148	170	190	217	247	295	320	346
Ch2	A	G	A	T	A	A	T	T	T	C	C	G	G
Ch3	A	G	A	T	A	A	T	T	T	C	C	G	G
Ch4	A	G	A	T	A	A	T	T	T	C	C	G	G
TU502	A	G	A	T	A	A	T	T	T	C	C	G	G
Cp2	G	A	G	C	G	G	A	T	C	C	T	A	A
Cp3	G	A	G	C	G	G	A	T	C	C	T	A	A
Cp4	G	A	G	C	G	G	A	T	C	C	T	A	A
Iowa	G	A	G	C	G	G	A	T	C	C	T	A	A
Moredun	G	A	G	C	G	G	A	T	C	C	T	A	A
w65	G	A	G	C	G	G	A	T	C	C	T	A	A
w66	G	A	G	C	G	G	A	T	C	C	T	A	A
w67	G	A	G	C	G	G	A	T	C	C	T	A	A
w70	G	A	G	C	G	G	A	T	C	C	T	A	A
Rabbit1	G	G	A	T	A	G	T	C	T	T	C	G	G
Rabbit2	G	G	A	T	A	G	T	C	T	T	C	G	G
Rabbit3	G	G	A	T	A	G	T	C	T	T	C	G	G
Rabbit4	G	G	A	T	A	G	T	C	T	T	C	G	G

cgd6_200 (1374 pb)
chr.6 (50.41k- 51.92k)
PCR product 447 bp
total number of bp 420

Isolates	37	55	102	103	109	112	116	136	145	151	202	256	259	265	277	301	318	319	325	331
Ch2	A	G	G	T	G	C	A	C	T	G	T	C	C	A	C	T	C	T	T	A
Ch3	A	G	G	T	G	C	A	C	T	G	T	C	C	A	C	T	C	T	T	A
Ch4	A	G	G	T	G	C	A	C	T	G	T	C	C	A	C	T	C	T	T	A
TU502	A	G	G	T	G	C	A	C	T	G	T	C	C	A	C	T	C	T	T	A
Cp2	A	G	G	C	T	T	A	C	T	G	C	T	C	G	A	T	A	T	T	A
Cp3	A	G	G	C	T	T	A	C	T	G	C	T	C	G	A	T	A	T	T	A
Cp4	A	G	G	C	T	T	A	C	T	G	C	T	C	G	A	T	A	T	T	A
Iowa	A	G	G	C	T	T	A	C	T	G	C	T	C	G	A	T	A	T	T	A
Moredun	A	G	G	C	T	T	A	C	T	G	C	T	C	G	A	T	A	T	T	A
w65	A	G	G	T	T	T	A	C	T	G	C	T	C	G	A	T	A	T	T	A
w66	A	G	G	T	T	T	A	C	T	G	C	T	C	G	A	T	A	T	T	A
w67	A	G	G	T	T	T	A	C	T	G	C	T	C	G	A	T	A	T	T	A
w70	A	G	G	T	T	T	A	C	T	G	C	T	C	G	A	T	A	T	T	A
Rabbit1	A	G	G	T	G	C	A	C	T	G	T	T	C	A	C	T	C	T	T	A
Rabbit2	A	G	G	T	G	C	A	C	T	G	T	T	C	A	C	T	C	T	T	A
Rabbit3	A	G	G	T	G	C	A	C	T	G	T	T	C	A	C	T	C	T	T	A
Rabbit4	A	G	G	T	G	C	A	C	T	G	T	T	C	A	C	T	C	T	T	A
C.meleagridis	G	A	A	T	T	T	T	A	C	A	T	T	T	G	A	C	C	A	C	G

cgd6_5020 (651 bp)
Chr.6 (1189.88k- 1190.60k)
PCR product 271 bp
total number of bp 248

Isolates	26	40	43	53	134	166	181
Ch2	T	G	C	A	T	T	C
Ch3	T	G	C	A	T	T	C
Ch4	T	G	C	A	T	T	C
TU502	T	G	C	A	T	T	C
Cp2	T	G	T	A	A	T	C
Cp3	T	G	T	A	A	T	C
Cp4	T	G	T	A	A	T	C
Iowa	T	G	T	A	A	T	C
Moredun	T	G	T	A	A	T	C
w65	T	G	T	A	A	T	C
w66	T	G	T	A	A	T	C
w67	T	G	T	A	A	T	C
w70	T	G	T	A	A	T	C
Rabbit1	T	G	C	A	T	T	C
Rabbit2	T	G	C	A	T	T	C
Rabbit3	T	G	C	A	T	T	C
Rabbit4	T	G	C	A	T	T	C
C.meleagridis	C	T	C	G	A	C	T

cgd8_2370 (1188 bp)
chr.8 (631.04K- 632.35K)
PCR product 685 bp
total number of bp 661

Isolates	72	102	132	279	291	312	414	450	513	543	579	596
Ch2	T	A	A	T	A	G	T	A	A	T	C	A
Ch3	T	A	A	T	A	G	T	A	A	T	C	A
Ch4	T	A	A	T	A	G	T	A	A	T	C	A
TU502	T	A	G	T	A	G	T	A	A	T	C	A
Cp2	T	C	G	A	A	A	C	G	G	C	T	G
Cp3	T	C	G	A	A	A	C	G	G	C	T	G
Cp4	T	C	G	A	A	A	C	G	G	C	T	G
Iowa	T	C	G	A	A	A	C	G	G	C	T	G
Moredun	T	C	G	A	A	A	C	G	G	C	T	G
w65	T	C	G	A	G	A	C	G	G	C	T	G
w66	T	C	G	A	G	A	C	G	G	C	T	G
w67	T	C	G	A	G	A	C	G	G	C	T	G
w70	T	C	G	A	G	A	C	G	G	C	T	G
Rabbit1	A	C	G	T	A	G	T	A	A	T	C	A
Rabbit2	A	C	G	T	A	G	T	A	A	T	C	A
Rabbit3	A	C	G	T	A	G	T	A	A	T	C	A
Rabbit4	A	C	G	T	A	G	T	A	A	T	C	A

Chro.20156 (795 bp)
Chr.2 (310.76- 311.551K)
PCR product 247 bp
total number of bp 248

Isolates	107	186	196	208	227	228
Ch2	G	A	A	T	A	A
Ch3	G	A	A	T	A	A
Ch4	G	A	A	T	A	A
TU502	G	A	A	T	A	A
Cp2	A	G	A	C	C	C
Cp3	A	G	A	C	C	C
Cp4	A	G	A	C	C	C
Iowa	A	G	A	C	C	C
Moredun	A	G	A	C	C	C
w65	A	G	A	C	C	C
w66	A	G	A	C	C	C
w67	A	G	A	C	C	C
w70	A	G	A	C	C	C
Rabbit1	G	A	T	T	A	A
Rabbit2	G	A	T	T	A	A
Rabbit3	G	A	T	T	A	A
Rabbit4	G	A	T	T	A	A

chro.50330 (681 bp)
chr.5 (148.97K- 151.05K)
PCR product 368 bp
total number of bp 346

Isolates	9	69	210
Ch2	G	T	A
Ch3	G	T	A
Ch4	G	T	A
TU502	G	T	A
Cp2	A	C	G
Cp3	A	C	G
Cp4	A	C	G
Iowa	A	C	G
Moredun	A	C	G
w65	A	C	G
w66	A	C	G
w67	A	C	G
w70	A	C	G
Rabbit1	G	T	A
Rabbit2	G	T	A
Rabbit3	G	T	A
Rabbit4	G	T	A

chro.50317 (1419 bp)
chr.5 (173.60K- 179.56K)
PCR product 752 bp
total number of bp 727

Isolates	9	35	46	65	71	74	82	110	126	134	152	153	203	214	219	221	228	230	239	245	263	293	298
Ch2	T	T	G	G	A	T	C	T	A	T	G	G	T	C	A	G	G	A	T	G	G	T	G
Ch3	T	T	G	G	A	T	C	T	A	T	G	G	T	C	A	G	G	A	T	G	G	T	G
Ch4	T	T	G	G	A	T	C	T	A	T	G	G	T	C	A	G	G	A	T	G	G	T	G
TU502	T	T	G	G	A	T	C	T	A	T	G	G	T	C	A	G	G	A	T	G	G	T	G
Cp2	C	T	G	G	A	T	T	T	A	T	G	G	T	C	A	G	G	A	T	A	A	T	G
Cp3	C	T	G	G	A	T	T	T	A	T	G	G	T	C	A	G	G	A	T	A	A	T	G
Cp4	C	T	G	G	A	T	T	T	A	T	G	G	T	C	A	G	G	A	T	A	A	T	G
Iowa	C	T	G	G	A	T	T	T	A	T	G	G	T	C	A	G	G	A	T	A	A	T	G
Moredun	C	T	G	G	A	T	T	T	A	T	G	G	T	C	A	G	G	A	T	A	A	T	G
w65	C	T	G	G	A	T	T	T	A	T	G	G	T	C	A	G	A	A	T	A	A	T	G
w66	C	T	G	G	A	T	T	T	A	T	G	G	T	C	A	G	A	A	T	A	A	T	G
w67	C	T	G	G	A	T	T	T	A	T	G	G	T	C	A	G	A	A	T	A	A	T	G
w70	C	T	G	G	A	T	T	T	A	T	G	G	T	C	A	G	A	A	T	A	A	T	G
Rabbit1	T	T	G	G	A	T	C	T	A	T	G	G	T	C	A	G	G	A	T	G	G	T	G
Rabbit2	T	T	G	G	A	T	C	T	A	T	G	G	T	C	A	G	G	A	T	G	G	T	G
Rabbit3	T	T	G	G	A	T	C	T	A	T	G	G	T	C	A	G	G	A	T	G	G	T	G
Rabbit4	T	T	G	G	A	T	C	T	A	T	G	G	T	C	A	G	G	A	T	G	G	T	G
C.meleagridis		C	A	A	G	G	G	C	G	A	T	T	C	G	G	T	G	G	A	G	G	C	A

chro.50317 (1419 bp)
chr.5 (173.60K-179.56K)
PCR product 752 bp
total number of bp 727

Isolate	299	307	310	319	324	331	340	364	396	413	424	431	458	464	476	494	523	524	533	536	569	581	584	596	620	632	638	645
Ch2	G	A	G	A	A	A	C	G	T	T	G	A	G	A	G	G	G	G	C	G	T	A	A	A	T	C	A	G
Ch3	G	A	G	A	A	A	C	G	T	T	G	A	G	A	G	G	G	G	C	G	T	A	A	A	T	C	A	G
Ch4	G	A	G	A	A	A	C	G	T	T	G	A	G	A	G	G	G	G	C	G	T	A	A	A	T	C	A	G
TU502	G	A	G	A	A	A	C	G	T	T	G	A	G	A	G	G	G	G	C	G	T	A	A	A	T	C	A	G
Cp2	G	G	C	G	A	A	T	G	T	T	A	A	G	G	G	G	A	G	C	G	T	A	A	A	C	C	A	A
Cp3	G	G	C	G	A	A	T	G	T	T	A	A	G	G	G	G	A	G	C	G	T	A	A	A	C	C	A	A
Cp4	G	G	C	G	A	A	T	G	T	T	A	A	G	G	G	G	A	G	C	G	T	A	A	A	C	C	A	A
Iowa	G	G	C	G	A	A	T	G	T	T	A	A	G	G	G	G	A	G	C	G	T	A	A	A	C	C	A	A
Moredun	G	G	C	G	A	A	T	G	T	T	A	A	G	G	G	G	A	G	C	G	T	A	A	A	C	C	A	A
w65	G	A	C	G	A	A	T	G	T	T	A	A	G	G	G	G	A	G	T	G	T	A	A	A	C	C	A	A
w66	G	A	C	G	A	A	T	G	T	T	A	A	G	G	G	G	A	G	T	G	T	A	A	A	C	C	A	A
w67	G	A	C	G	A	A	T	G	T	T	A	A	G	G	G	G	A	G	T	G	T	A	A	A	C	C	A	A
w70	G	A	C	G	A	A	T	G	T	T	A	A	G	G	G	G	A	G	T	G	T	A	A	A	C	C	A	A
Rabbit1	G	A	C	A	A	A	C	G	T	T	A	A	G	A	G	G	G	G	C	G	T	A	A	A	T	C	A	G
Rabbit2	G	A	C	A	A	A	C	G	T	T	A	A	G	A	G	G	G	G	C	G	T	A	A	A	T	C	A	G
Rabbit3	G	A	C	A	A	A	C	G	T	T	A	A	G	A	G	G	G	G	C	G	T	A	A	A	T	C	A	G
Rabbit4	G	A	C	A	A	A	C	G	T	T	A	A	G	A	G	G	G	G	C	G	T	A	A	A	T	C	A	G
C.melea-gridis	A	A	C	G	G	G	T	C	C	C	A	G	A	G	T	T	A	A	C	T	C	G	G	G	T	T	G	A

chro.50457 (1080 bp)
Chr.5 (903.26K- 905.12K)
PCR product 394 bp
total number of bp 373

Isolates	12	13	28	31	38	43	44	45	49	50	51	73	85	91	94	97	98	108	142	146	154	158	161	162	169	178
Ch2	G	T	A	C	G	A	T	T	A	G	A	A	T	T	A	C	G	A	T	A	G	G	A	A	C	A
Ch3	G	T	A	C	G	A	T	T	A	G	A	A	T	T	A	C	G	A	T	A	G	G	A	A	C	A
Ch4	G	T	A	C	G	A	T	T	A	G	A	A	T	T	A	C	G	A	T	A	G	G	A	A	C	A
TU502	G	T	A	C	G	A	T	T	A	G	A	A	T	T	A	C	G	A	T	A	G	G	A	A	C	A
Cp2	G	T	A	C	G	A	T	C	G	G	A	A	T	T	A	C	A	A	T	A	A	G	G	G	T	A
Cp3	G	T	A	C	G	A	T	C	G	G	A	A	T	T	A	C	A	A	T	A	A	G	G	G	T	A
Cp4	G	T	A	C	G	A	T	C	G	G	A	A	T	T	A	C	A	A	T	A	A	G	G	G	T	A
Iowa	G	T	A	C	G	A	T	C	G	G	A	A	T	T	A	C	A	A	T	A	A	G	G	G	T	A
Moredun	G	T	A	C	G	A	T	C	G	G	A	A	T	T	A	C	A	A	T	A	A	G	G	G	T	A
w65	G	T	A	C	G	A	T	C	G	G	A	A	T	T	A	C	A	A	T	A	A	G	G	G	T	A
w66	G	T	A	C	G	A	T	C	G	G	A	A	T	T	A	C	A	A	T	A	A	G	G	G	T	A
w67	G	T	A	C	G	A	T	C	G	G	A	A	T	T	A	C	A	A	T	A	A	G	G	G	T	A
w70	G	T	A	C	G	A	T	C	G	G	A	A	T	T	A	C	A	A	T	A	A	G	G	G	T	A
Rabbit1	G	T	A	C	G	A	T	T	A	G	A	A	T	T	A	C	G	A	T	A	G	G	A	A	C	A
Rabbit2	G	T	A	C	G	A	T	T	A	G	A	A	T	T	A	C	G	A	T	A	G	G	A	A	C	A
Rabbit3	G	T	A	C	G	A	T	T	A	G	A	A	T	T	A	C	G	A	T	A	G	G	A	A	C	A
Rabbit4	G	T	A	C	G	A	T	T	A	G	A	A	T	T	A	C	G	A	T	A	G	G	A	A	C	A
C.meleagridis	A	C	G	T	A	G	A	C	A	A	T	G	C	C	G	T	G	G	C	G	A	A	G	A	G	G

chro.50457 (1080 bp)

Chr.5 (903.26K- 905.12K)
PCR product 394 bp
total number of bp 373

Isolates	196	227	229	238	241	247	265	266	268	271	283	307	310	313	316	331	335	340	350
Ch2	T	C	A	C	T	A	A	G	T	T	C	C	C	T	T	T	A	C	T
Ch3	T	C	A	C	T	A	A	G	T	T	C	C	C	T	T	T	A	C	T
Ch4	T	C	A	C	T	A	A	G	T	T	C	C	C	T	T	T	A	C	T
TU502	T	C	A	C	T	A	A	G	T	T	C	C	C	T	T	T	A	C	T
Cp2	T	C	A	T	T	A	A	A	T	T	C	C	T	T	T	T	A	T	C
Cp3	T	C	A	T	T	A	A	A	T	T	C	C	T	T	T	T	A	T	C
Cp4	T	C	A	T	T	A	A	A	T	T	C	C	T	T	T	T	A	T	C
Iowa	T	C	A	T	T	A	A	A	T	T	C	C	T	T	T	T	A	T	C
Moredun	T	C	A	T	T	A	A	A	T	T	C	C	T	T	T	T	A	T	C
w65	T	C	A	T	T	A	A	A	T	T	C	C	T	T	T	T	A	T	C
w66	T	C	A	T	T	A	A	A	T	T	C	C	T	T	T	T	A	T	C
w67	T	C	A	T	T	A	A	A	T	T	C	C	T	T	T	T	A	T	C
w70	T	C	A	T	T	A	A	A	T	T	C	C	T	T	T	T	A	T	C
Rabbit1	T	C	A	C	T	A	A	G	T	T	C	C	C	T	T	T	A	C	T
Rabbit2	T	C	A	C	T	A	A	G	T	T	C	C	C	T	T	T	A	C	T
Rabbit3	T	C	A	C	T	A	A	G	T	T	C	C	C	T	T	T	A	C	T
Rabbit4	T	C	A	C	T	A	A	G	T	T	C	C	C	T	T	T	A	C	T
C.meleagridis	C	T	G	T	C	T	G	G	C	C	T	T	T	C	G	C	G	T	T

Chro.30149 (1500 bp)
chr.3 (9.7K- 11.5 K)
PCR product 331 bp

	total number of bp 308																			
Isolates	14	17	41	44	50	56	59	62	68	71	110	152	158	167	173	179	233	245	257	263
Ch2	T	A	A	A	C	T	T	T	A	T	T	A	T	A	G	T	T	A	A	T
Ch3	T	A	A	A	C	T	T	T	A	T	T	A	T	A	G	T	T	A	A	T
Ch4	T	A	A	A	C	T	T	T	A	T	T	A	T	A	G	T	T	A	A	T
TU502	T	A	A	A	C	T	T	T	A	T	T	A	T	A	G	T	T	A	A	T
Cp2	T	A	A	A	C	T	T	T	A	T	T	A	T	A	G	T	T	A	A	T
Cp3	T	A	A	A	C	T	T	T	A	T	T	A	T	A	G	T	T	A	A	T
Cp4																				
Iowa	T	A	A	A	C	T	T	T	A	T	T	A	T	A	G	T	T	A	A	T
Moredun	T	A	A	A	C	T	T	T	A	T	T	A	T	A	G	T	T	A	A	T
w65	T	A	A	A	C	T	T	T	A	T	T	A	T	A	G	T	T	A	A	T
w66	T	A	A	A	C	T	T	T	A	T	T	A	T	A	G	T	T	A	A	T
w67	T	A	A	A	C	T	T	T	A	T	T	A	T	A	G	T	T	A	A	T
w70	T	A	A	A	C	T	T	T	A	T	T	A	T	A	G	T	T	A	A	T
Rabbit1	T	A	A	A	C	T	T	T	A	T	T	A	T	A	G	T	T	A	A	T
Rabbit2	T	A	A	A	C	T	T	T	A	T	T	A	T	A	G	T	T	A	A	T
Rabbit3	T	A	A	A	C	T	T	T	A	T	T	A	T	A	G	T	T	A	A	T
Rabbit4	T	A	A	A	C	T	T	T	A	T	T	A	T	A	G	T	T	A	A	T
C.meleagridis	A	G	G	G	T	C	C	C	G	C	C	G	C	C	T	C	C	G	G	C

COWP (*Cryptosporidium* oocyst wall protein) 4890 bp
Chr.6 (493.95K- 499.32K)
PCR product 555 bp

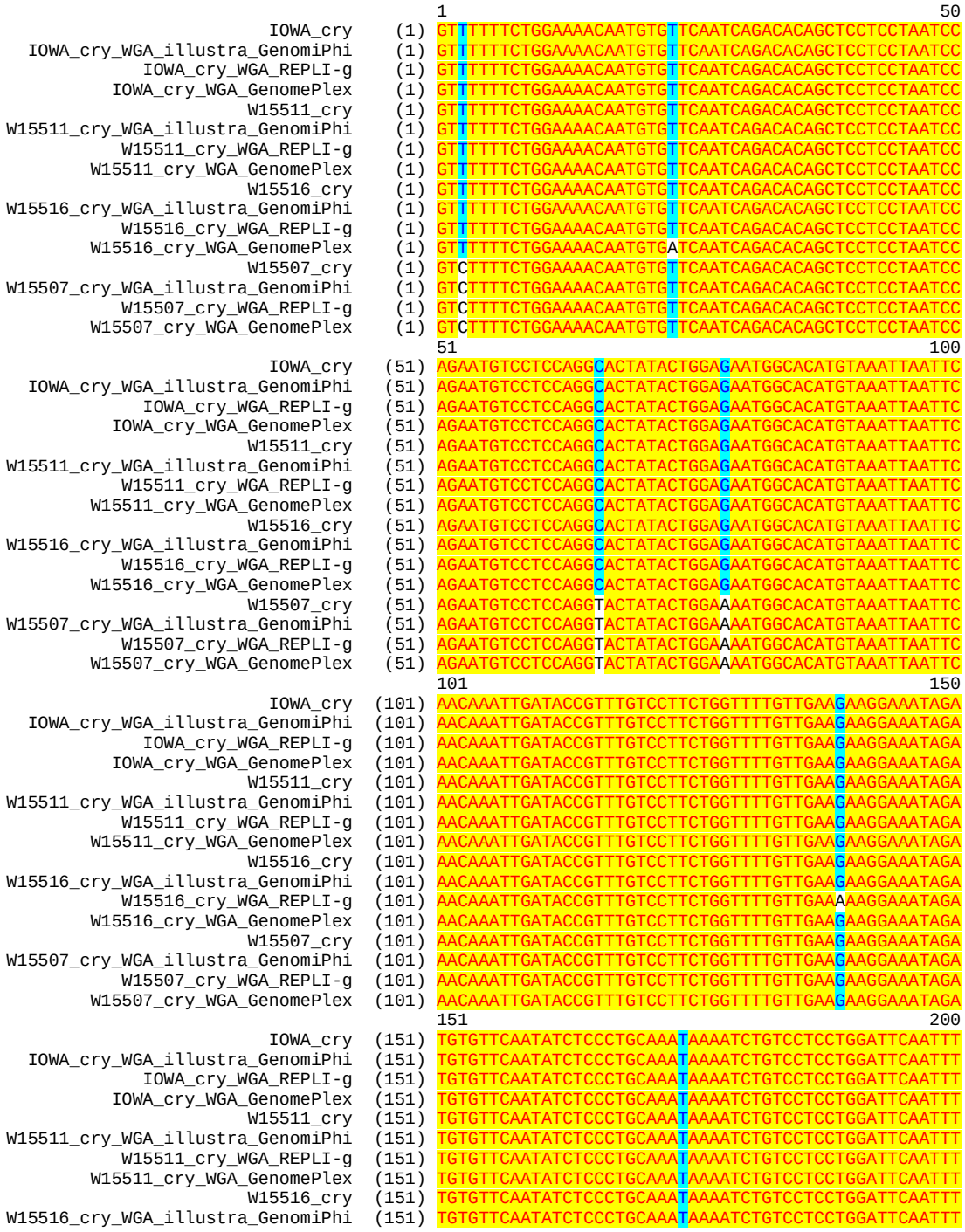
total number of bp 529

Isolates	17	18	29	36	45	54	66	72	87	129	147	195	234	249	285	306	396	402	405	414	465	477	489	513
Ch2	A	T	G	G	A	A	A	C	A	T	A	A	C	T	G	G	G	T	C	A	T	G	G	G
Ch3	A	T	G	G	A	A	A	C	A	T	A	A	C	T	G	G	G	T	C	A	T	G	G	G
Ch4	A	T	G	G	A	A	A	C	A	T	A	A	C	T	G	G	G	T	C	A	T	G	G	G
TU502	A	T	G	G	A	A	A	C	A	T	A	A	C	T	G	G	G	T	C	A	T	G	G	G
Cp2	A	T	G	G	A	A	T	T	A	T	G	A	C	T	G	A	G	C	C	G	T	A	G	G
Cp3	A	T	G	G	A	A	T	T	A	T	G	A	C	T	G	A	G	C	C	G	T	A	G	G
Cp4	A	T	G	G	A	A	T	T	A	T	G	A	C	T	G	A	G	C	C	G	T	A	G	G
Iowa	A	T	G	G	A	A	T	T	A	T	G	A	C	T	G	A	G	C	C	G	T	A	G	G
Moredun	A	T	G	G	A	A	T	T	A	T	G	A	C	T	G	A	G	C	C	G	T	A	G	G
w65	A	T	G	G	A	A	T	T	A	T	G	A	C	T	G	A	G	C	C	G	T	A	G	G
w66	A	T	G	G	A	A	T	T	A	T	G	A	C	T	G	A	G	C	C	G	T	A	G	G
w67	A	T	G	G	A	A	T	T	A	T	G	A	C	T	G	A	G	C	C	G	T	A	G	G
w70	A	T	G	G	A	A	T	T	A	T	G	A	C	T	G	A	G	C	C	G	T	A	G	G
Rabbit1	A	T	G	G	A	A	A	C	A	T	A	A	C	T	G	G	G	T	C	A	T	G	G	G
Rabbit2	A	T	G	G	A	A	A	C	A	T	A	A	C	T	G	G	G	T	C	A	T	G	G	G
Rabbit3	A	T	G	G	A	A	A	C	A	T	A	A	C	T	G	G	G	T	C	A	T	G	G	G
Rabbit4	A	T	G	G	A	A	A	C	A	T	A	A	C	T	G	G	G	T	C	A	T	G	G	G
C.meleagridis	G	C	T	T	T	G	A	C	T	A	G	G	A	A	A	G	A	C	T	G	C	A	A	A

Appendix VI

Appendix VI: Alignment of PCR products sequences for W15507, W15519, W15511, W15516 and *C. parvum* IOWA isolates before and after WGA using illustra GenomiPhi, REPLI-g and GenomePlex kits.

COWP gene PCR product



```
          W15516_cry_WGA_REPLI-g  (151) TGTGTTCAATATCTCCCTGCAAATAAAATCTGTCCTCCTGGATTCAATTT
        W15516_cry_WGA_GenomePlex  (151) TGTGTTCAATATCTCCCTGCAAATAAAATCTGTCCTCCTGGATTCAATTT
                      W15507_cry  (151) TGTGTTCAATATCTCCCTGCAAACAAAATCTGTCCTCCTGGATTCAATTT
W15507_cry_WGA_illustra_GenomiPhi  (151) TGTGTTCAATATCTCCCTGCAAACAAAATCTGTCCTCCTGGATTCAATTT
          W15507_cry_WGA_REPLI-g  (151) TGTGTTCAATATCTCCCTGCAAACAAAATCTGTCCTCCTGGATTCAATTT
        W15507_cry_WGA_GenomePlex  (151) TGTGTTCAATATCTCCCTGCAAACAAAATCTGTCCTCCTGGATTCAATTT
                                         201                                       250
                      IOWA_cry  (201) GTCAGGACAACAATGTATGGCACCAGAATCAGCTGAATTAGAATCGACAT
  IOWA_cry_WGA_illustra_GenomiPhi  (201) GTCAGGACAACAATGTATGGCACCAGAATCAGCTGAATTAGAATCGACAT
            IOWA_cry_WGA_REPLI-g  (201) GTCAGGACAACAATGTATGGCACCAGAATCAGCTGAATTAGAATCGACAT
          IOWA_cry_WGA_GenomePlex  (201) GTCAGGACAACAATGTATGGCACCAGAATCAGCTGAATTAGAATCGACAT
                      W15511_cry  (201) GTCAGGACAACAATGTATGGCACCAGAATCAGCTGAATTAGAATCGACAT
W15511_cry_WGA_illustra_GenomiPhi  (201) GTCAGGACAACAATGTATGGCACCAGAATCAGCTGAATTAGAATCGACAT
          W15511_cry_WGA_REPLI-g  (201) GTCAGGACAACAATGTATGGCACCAGAATCAGCTGAATTAGAATCGACAT
        W15511_cry_WGA_GenomePlex  (201) GTCAGGACAACAATGTATGGCACCAGAATCAGCTGAATTAGAATCGACAT
                      W15516_cry  (201) GTCAGGACAACAATGTATGGCACCAGAATCAGCTGAATTAGAATCGACAT
W15516_cry_WGA_illustra_GenomiPhi  (201) GTCAGGACAACAATGTATGGCACCAGAATCAGCTGAATTAGAATCGACAT
          W15516_cry_WGA_REPLI-g  (201) GTCAGGACAACAATGTATGGCACCAGAATCAGCTGAATTAAAATCGACAT
        W15516_cry_WGA_GenomePlex  (201) GTCAGGACAACAATGTATGGCACCAGAATCAGCTGAATTAGAATCGACAT
                      W15507_cry  (201) GTCAGGACAACAATGTATGGCACCAGAATCAGCTGAATTAGAATCGACAT
W15507_cry_WGA_illustra_GenomiPhi  (201) GTCAGGACAACAATGTATGGCACCAGAATCAGCTGAATTAGAATCGACAT
          W15507_cry_WGA_REPLI-g  (201) GTCAGGACAACAATGTATGGCACCAGAATCAGCTGAATTAGAATCGACAT
        W15507_cry_WGA_GenomePlex  (201) GTCAGGACAACAATGTATGGCACCAGAATCAGCTGAATTAGAATCGACAT
                                         251                                       300
                      IOWA_cry  (251) GCCCACCTAATTCAATATTTGAAAATGGAAAATGTAAAGTGATTAAAAAT
  IOWA_cry_WGA_illustra_GenomiPhi  (251) GCCCACCTAATTCAATATTTGAAAATGGAAAATGTAAAGTGATTAAAAAT
            IOWA_cry_WGA_REPLI-g  (251) GCCCACCTAATTCAATATTTGAAAATGGAAAATGTAAAGTGATTAAAAAT
          IOWA_cry_WGA_GenomePlex  (251) GCCCACCTAATTCAATATTTGAAAATGGAAAATGTAAAGTGATTAAAAAT
                      W15511_cry  (251) GCCCACCTAATTCAATATTTGAAAATGGAAAATGTAAAGTGATTAAAAAT
W15511_cry_WGA_illustra_GenomiPhi  (251) GCCCACCTAATTCAATATTTGAAAATGGAAAATGTAAAGTGATTAAAAAT
          W15511_cry_WGA_REPLI-g  (251) GCCCACCTAATTCAATATTTGAAAATGGAAAATGTAAAGTGATTAAAAAT
        W15511_cry_WGA_GenomePlex  (251) GCCCACCTAATTCAATATTTGAAAATGGAAAATGTAAAGTGATTAAAAAT
                      W15516_cry  (251) GCCCACCTAATTCAATATTTGAAAATGGAAAATGTAAAGTGATTAAAAAT
W15516_cry_WGA_illustra_GenomiPhi  (251) GCCCACCTAATTCAATATTTGAAAATGGAAAATGTAAAGTGATTAAAAAT
          W15516_cry_WGA_REPLI-g  (251) GCCCACCTAATTCAATATTTGAAAATGGAAAATGTAAAGTGATTAAAAAT
        W15516_cry_WGA_GenomePlex  (251) GCCCACCTAATTCAATATTTGAAAATGGAAAATGTAAAGTGATTAAAAAT
                      W15507_cry  (251) GCCCACCTAATTCAATATTTGAAAATGGAAAATGTAAAGTGATTAAAAAT
W15507_cry_WGA_illustra_GenomiPhi  (251) GCCCACCTAATTCAATATTTGAAAATGGAAAATGTAAAGTGATTAAAAAT
          W15507_cry_WGA_REPLI-g  (251) GCCCACCTAATTCAATATTTGAAAATGGAAAATGTAAAGTGATTAAAAAT
        W15507_cry_WGA_GenomePlex  (251) GCCCACCTAATTCAATATTTGAAAATGGAAAATGTAAAGTGATTAAAAAT
                                         301                                       350
                      IOWA_cry  (301) ATTGATATGGTCTGCCCACCAGGATATACAGACTCAGGAGATGATTGTGT
  IOWA_cry_WGA_illustra_GenomiPhi  (301) ATTGATATGGTCTGCCCACCAGGATATACAGACTCAGGAGATGATTGTGT
            IOWA_cry_WGA_REPLI-g  (301) ATTGATATGGTCTGCCCACCAGGATATACAGACTCAGGAGATGATTGTGT
          IOWA_cry_WGA_GenomePlex  (301) ATTGATATGGTCTGCCCACCAGGATATACAGACTCAGGAGATGATTGTGT
                      W15511_cry  (301) ATTGATATGGTCTGCCCACCAGGATATACAGACTCAGGAGATGATTGTGT
W15511_cry_WGA_illustra_GenomiPhi  (301) ATTGATATGGTCTGCCCACCAGGATATACAGACTCAGGAGATGATTGTGT
          W15511_cry_WGA_REPLI-g  (301) ATTGATATGGTCTGCCCACCAGGATATACAGACTCAGGAGATGATTGTGT
        W15511_cry_WGA_GenomePlex  (301) ATTGATATGGTCTGCCCACCAGGATATACAGACTCAGGAGATGATTGTGT
                      W15516_cry  (301) ATTGATATGGTCTGCCCACCAGGATATACAGACTCAGGAGATGATTGTGT
W15516_cry_WGA_illustra_GenomiPhi  (301) ATTGATATGGTCTGCCCACCAGGATATACAGACTCAGGAGATGATTGTGT
          W15516_cry_WGA_REPLI-g  (301) ATTGATATGGTCTGCCCACCAGGATATACAGACTCAGGAAATGATTGTGT
        W15516_cry_WGA_GenomePlex  (301) ATTGATATGGTCTGCCCACCAGGATATACAGACTCAGGAGATGATTGTGT
                      W15507_cry  (301) ATTGATATGGTCTGCCCACCAGGATATACAGATTCAGGAGATGATTGTGT
W15507_cry_WGA_illustra_GenomiPhi  (301) ATTGATATGGTCTGCCCACCAGGATATACAGATTCAGGAGATGATTGTGT
          W15507_cry_WGA_REPLI-g  (301) ATTGATATGGTCTGCCCACCAGGATATACAGATTCAGGAGATGATTGTGT
        W15507_cry_WGA_GenomePlex  (301) ATTGATATGGTCTGCCCACCAGGATATACAGATTCAGGAGATGATTGTGT
                                         351                                       400
                      IOWA_cry  (351) ACTATATGTAGCTCCTGCAAAGGAATGCCCACCAAATTTCATTTTACAAG
  IOWA_cry_WGA_illustra_GenomiPhi  (351) ACTATATGTAGCTCCTGCAAAGGAATGCCCACCAAATTTCATTTTACAAG
            IOWA_cry_WGA_REPLI-g  (351) ACTATATGTAGCTCCTGCAAAGGAATGCCCACCAAATTTCATTTTACAAG
          IOWA_cry_WGA_GenomePlex  (351) ACTATATGTAGCTCCTGCAAAGGAATGCCCACCAAATTTCATTTTACAAG
                      W15511_cry  (351) ACTATATGTAGCTCCTGCAAAGGAATGCCCACCAAATTTCATTTTACAAG
W15511_cry_WGA_illustra_GenomiPhi  (351) ACTATATGTAGCTCCTGCAAAGGAATGCCCACCAAATTTCATTTTACAAG
          W15511_cry_WGA_REPLI-g  (351) ACTATATGTAGCTCCTGCAAAGGAATGCCCACCAAATTTCATTTTACAAG
        W15511_cry_WGA_GenomePlex  (351) ACTATATGTAGCTCCTGCAAAGGAATGCCCACCAAATTTCATTTTACAAG
                      W15516_cry  (351) ACTATATGTAGCTCCTGCAAAGGAATGCCCACCAAATTTCATTTTACAAG
W15516_cry_WGA_illustra_GenomiPhi  (351) ACTATATGTAGCTCCTGCAAAGGAATGCCCACCAAATTTCATTTTACAAG
          W15516_cry_WGA_REPLI-g  (351) ACTATATGTAGCTCCTGCAAAGGAATGCCCACCAAATTTCATTTTACAAG
        W15516_cry_WGA_GenomePlex  (351) ACTATATGTAGCTCCTGCAAAGGAATGCCCACCAAATTTCATTTTACAAG
                      W15507_cry  (351) ACTATATGTAGCTCCTGCAAAGGAATGCCCACCAAATTTCATTTTACAAG
W15507_cry_WGA_illustra_GenomiPhi  (351) ACTATATGTAGCTCCTGCAAAGGAATGCCCACCAAATTTCATTTTACAAG
          W15507_cry_WGA_REPLI-g  (351) ACTATATGTAGCTCCTGCAAAGGAATGCCCACCAAATTTCATTTTACAAG
        W15507_cry_WGA_GenomePlex  (351) ACTATATGTAGCTCCTGCAAAGGAATGCCCACCAAATTTCATTTTACAAG
                                         401                                       450
                      IOWA_cry  (401) GCCTCCAATGTATACAAACTAGTTCTGCTCCAACTCAACCTGTCTGCCCT
  IOWA_cry_WGA_illustra_GenomiPhi  (401) GCCTCCAATGTATACAAACTAGTTCTGCTCCAACTCAACCTGTCTGCCCT
            IOWA_cry_WGA_REPLI-g  (401) GCCTCCAATGTATACAAACTAGTTCTGCTCCAACTCAACCTGTCTGCCCT
          IOWA_cry_WGA_GenomePlex  (401) GCCTCCAATGTATACAAACTAGTTCTGCTCCAACTCAACCTGTCTGCCCT
                      W15511_cry  (401) GCCTCCAATGTATACAAACTAGTTCTGCTCCAACTCAACCTGTCTGCCCT
W15511_cry_WGA_illustra_GenomiPhi  (401) GCCTCCAATGTATACAAACTAGTTCTGCTCCAACTCAACCTGTCTGCCCT
```

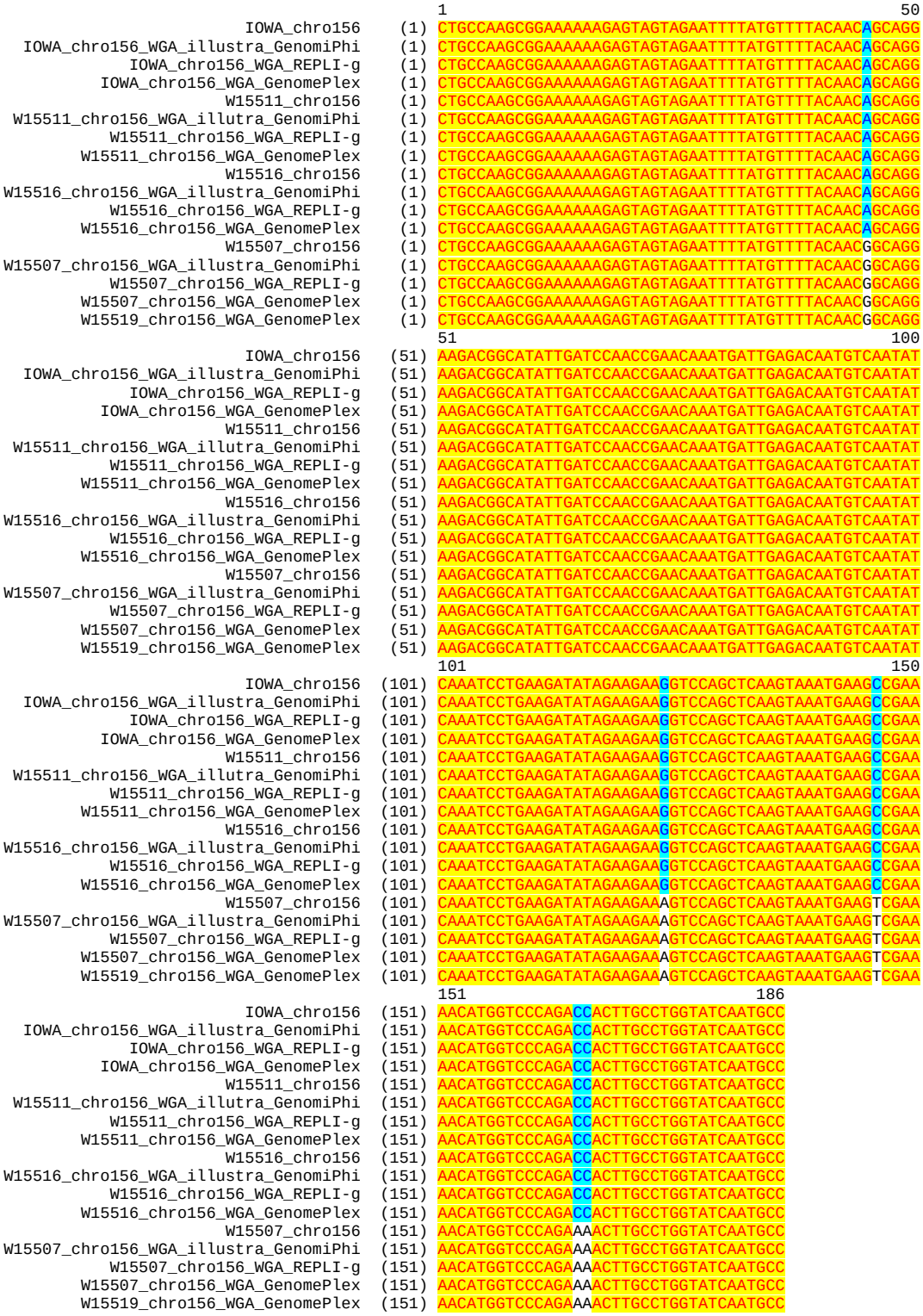
```
                W15511_cry_WGA_REPLI-g  (401) GCCTCCAATGTATACAAACTAGTTCTGCTCCAACTCAACCTGTCTGCCCT
              W15511_cry_WGA_GenomePlex  (401) GCCTCCAATGTATACAAACTAGTTCTGCTCCAACTCAACCTGTCTGCCCT
                         W15516_cry  (401) GCCTCCAATGTATACAAACTAGTTCTGCTCCAACTCAACCTGTCTGCCCT
  W15516_cry_WGA_illustra_GenomiPhi  (401) GCCTCCAATGTATACAAACTAGTTCTGCTCCAACTCAACCTGTCTGCCCT
                W15516_cry_WGA_REPLI-g  (401) GCCTCCAATGTATACAAACTAGTTCTGCTCCAACTCAACCTGTCTGCCCT
              W15516_cry_WGA_GenomePlex  (401) GCCTCCAATGTATACAAACTAGTTCTGCTCCAACTCAACCTGTCTGCCCT
                         W15507_cry  (401) GCCTCCAGTGTATTCAAACTAGTTCTGCTCCAACTCAACCTGTCTGCCCT
  W15507_cry_WGA_illustra_GenomiPhi  (401) GCCTCCAGTGTATTCAAACTAGTTCTGCTCCAACTCAACCTGTCTGCCCT
                W15507_cry_WGA_REPLI-g  (401) GCCTCCAGTGTATTCAAACTAGTTCTGCTCCAACTCAACCTGTCTGCCCT
              W15507_cry_WGA_GenomePlex  (401) GCCTCCAGTGTATTCAAACTAGTTCTGCTCCAACTCAACCTGTCTGCCCT
                                         451               470
                          IOWA_cry  (451) CCAGGTACAGTATTACAAGA
    IOWA_cry_WGA_illustra_GenomiPhi  (451) CCAGGTACAGTATTACAAGA
                IOWA_cry_WGA_REPLI-g  (451) CCAGGTACAGTATTACAAGA
              IOWA_cry_WGA_GenomePlex  (451) CCAGGTACAGTATTACAAGA
                         W15511_cry  (451) CCAGGTACAGTATTACAAGA
  W15511_cry_WGA_illustra_GenomiPhi  (451) CCAGGTACAGTATTACAAGA
                W15511_cry_WGA_REPLI-g  (451) CCAGGTACAGTATTACAAGA
              W15511_cry_WGA_GenomePlex  (451) CCAGGTACAGTATTACAAGA
                         W15516_cry  (451) CCAGGTACAGTATTACAAGA
  W15516_cry_WGA_illustra_GenomiPhi  (451) CCAGGTACAGTATTACAAGA
                W15516_cry_WGA_REPLI-g  (451) CCAGGTACAGTATTACAAGA
              W15516_cry_WGA_GenomePlex  (451) CCAGGTACAGTATTACAAGA
                         W15507_cry  (451) CCAGGTACAGTATTACAAGA
  W15507_cry_WGA_illustra_GenomiPhi  (451) CCAGGTACAGTATTACAAGA
                W15507_cry_WGA_REPLI-g  (451) CCAGGTACAGTATTACAAGA
              W15507_cry_WGA_GenomePlex  (451) CCAGGTACAGTATTACAAGA
```

Cgd6_5020 gene PCR product

```
                                         1                                                 50
                          IOWA_cgd5020    (1) ATGAGCCAATAATGAAGAACAAGAAATCCCATTCTATGGGTGTAACTTGT
    IOWA_cgd5020_WGA_illustra_GenomiPhi    (1) ATGAGCCAATAATGAAGAACAAGAAATCCCATTCTATGGGTGTAACTTGT
                IOWA_cgd5020_WGA_REPLI-g    (1) ATGAGCCAATAATGAAGAACAAGAAATCCCATTCTATGGGTGTAACTTGT
              IOWA_cgd5020_WGA_GenomePlex    (1) ATGAGCCAATAATGAAGAACAAGAAATCCCATTCTATGGGTGTAACTTGT
                         W15511_cgd5020    (1) ATGAGCCAATAATGAAGAACAAGAAATCCCATTCTATGGGTGTAACTTGT
  W15511_cgd5020_WGA_illustra_GenomiPhi    (1) ATGAGCCAATAATGAAGAACAAGAAATCCCATTCTATGGGTGTAACTTGT
                W15511_cgd5020_WGA_REPLI-g    (1) ATGAGCCAATAATGAAGAACAAGAAATCCCATTCTATGGGTGTAACTTGT
              W15511_cgd5020_WGA_GenomePlex    (1) ATGAGCCAATAATGAAGAACAAGAAATCCCATTCTATGGGTGTAACTTGT
                         W15516_cgd5020    (1) ATGAGCCAATAATGAAGAACAAGAAATCCCATTCTATGGGTGTAACTTGT
  W15516_cgd5020_WGA_illustra_GenomiPhi    (1) ATGAGCCAATAATGAAGAACAAGAAATCCCATTCTATGGGTGTAACTTGT
                W15516_cgd5020_WGA_REPLI-g    (1) ATGAGCCAATAATGAAGAACAAGAAATCCCATTCTATGGGTGTAACTTGT
              W15516_cgd5020_WGA_GenomePlex    (1) ATGAGCCAATAATGAAGAACAAGAAATCCCATTCTATGGGTGTAACTTGT
                         W15507_cgd5020    (1) ATGAGCCAATAATGAAGAACAAGAAATCCCATTCTATGGGTGTAACTTGT
  W15507_cgd5020_WGA_illustra_GenomiPhi    (1) ATGAGCCAATAATGAAGAACAAGAAATCCCATTCTATGGGTGTAACTTGT
                W15507_cgd5020_WGA_REPLI-g    (1) ATGAGCCAATAATGAAGAACAAGAAATCCCATTCTATGGGTGTAACTTGT
              W15507_cgd5020_WGA_GenomePlex    (1) ATGAGCCAATAATGAAGAACAAGAAATCCCATTCTATGGGTGTAACTTGT
                         W15519_cgd5020    (1) ATGAGCCAATAATGAAGAACAAGAAATCCCATTCTATGGGTGTAACTTGT
  WW15519_cgd5020_WGA_illustraGenomiPhi    (1) ATGAGCCAATAATGAAGAACAAGAAATCCCATTCTATGGGTGTAACTTGT
        WW15519_cgd5020_WGA_GenomePlex    (1) ATGAGCCAATAATGAAGAACAAGAAATCCCATTCTATGGGTGTAACTTGT
                                        51                                                100
                          IOWA_cgd5020   (51) ATCCAAAAGTCGAATAGAAATCATCAATTTTGGTCAGGAAGTTACGACGA
    IOWA_cgd5020_WGA_illustra_GenomiPhi   (51) ATCCAAAAGTCGAATAGAAATCATCAATTTTGGTCAGGAAGTTACGACGA
                IOWA_cgd5020_WGA_REPLI-g   (51) ATCCAAAAGTCGAATAGAAATCATCAATTTTGGTCAGGAAGTTACGACGA
              IOWA_cgd5020_WGA_GenomePlex   (51) ATCCAAAAGTCGAATAGAAATCATCAATTTTGGTCAGGAAGTTACGACGA
                         W15511_cgd5020   (51) ATCCAAAAGTCGAATAGAAATCATCAATTTTGGTCAGGAAGTTACGACGA
  W15511_cgd5020_WGA_illustra_GenomiPhi   (51) ATCCAAAAGTCGAATAGAAATCATCAATTTTGGTCAGGAAGTTACGACGA
                W15511_cgd5020_WGA_REPLI-g   (51) ATCCAAAAGTCGAATAGAAATCATCAATTTTGGTCAGGAAGTTACGACGA
              W15511_cgd5020_WGA_GenomePlex   (51) ATCCAAAAGTCGAATAGAAATCATCAATTTTGGTCAGGAAGTTACGACGA
                         W15516_cgd5020   (51) ATCCAAAAGTCGAATAGAAATCATCAATTTTGGTCAGGAAGTTACGACGA
  W15516_cgd5020_WGA_illustra_GenomiPhi   (51) ATCCAAAAGTCGAATAGAAATCATCAATTTTGGTCAGGAAGTTACGACGA
                W15516_cgd5020_WGA_REPLI-g   (51) ATCCAAAAGTCGAATAGAAATCATCAATTTTGGTCAGGAAGTTACGACGA
              W15516_cgd5020_WGA_GenomePlex   (51) ATCCAAAAGTCGAATAGAAATCATCAATTTTGGTCAGGAAGTTACGACGA
                         W15507_cgd5020   (51) ATCCAAAAGTCGAATAGAAATCATCAATTTTGGTCAGGTAGTTACGACGA
  W15507_cgd5020_WGA_illustra_GenomiPhi   (51) ATCCAAAAGTCGAATAGAAATCATCAATTTTGGTCAGGTAGTTACGACGA
                W15507_cgd5020_WGA_REPLI-g   (51) ATCCAAAAGTCGAATAGAAATCATCAATTTTGGTCAGGTAGTTACGACGA
              W15507_cgd5020_WGA_GenomePlex   (51) ATCCAAAAGTCGAATAGAAATCATCAATTTTGGTCAGGTAGTTACGACGA
                         W15519_cgd5020   (51) ATCCAAAAGTCGAATAGAAATCATCAATTTTGGTCAGGTAGTTACGACGA
```

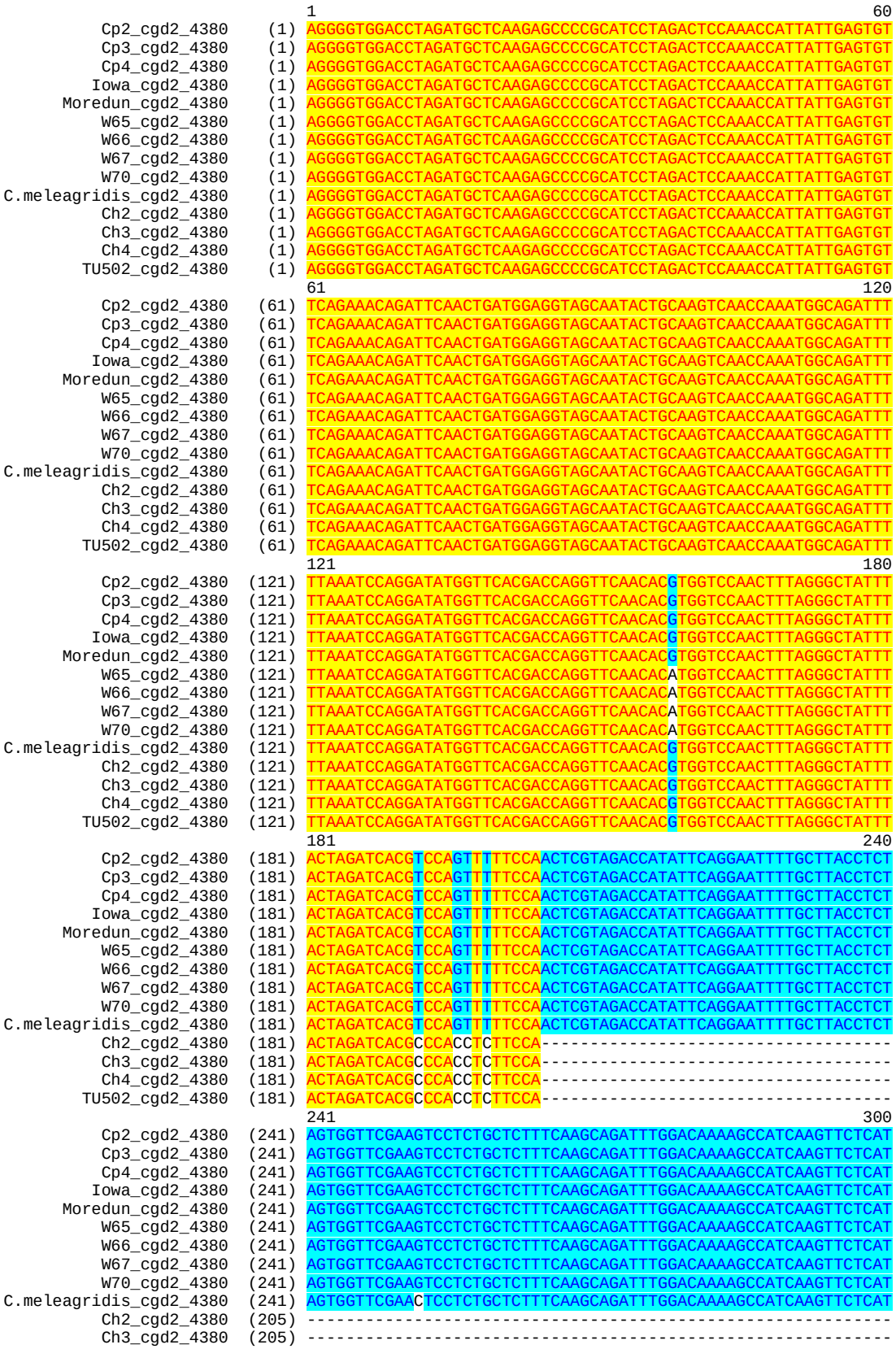
```
WW15519_cgd5020_WGA_illustraGenomiPhi   (51) ATCCAAAAGTCGAATAGAAATCATCAATTTTGGTCAGGTAGTTACGACGA
       WW15519_cgd5020_WGA_GenomePlex   (51) ATCCAAAAGTCGAATAGAAATCATCAATTTTGGTCAGGTAGTTACGACGA
                                             101                                             150
                         IOWA_cgd5020  (101) AACCCTTAGATTCTGGGATTTTAGGATGATAAACTCTCCTATTTATGAGC
  IOWA_cgd5020_WGA_illustra_GenomiPhi  (101) AACCCTTAGATTCTGGGATTTTAGGATGATAAACTCTCCTATTTATGAGC
             IOWA_cgd5020_WGA_REPLI-g  (101) AACCCTTAGATTCTGGGATTTTAGGATGATAAACTCTCCTATTTATGAGC
          IOWA_cgd5020_WGA_GenomePlex  (101) AACCCTTAGATTCTGGGATTTTAGGATGATAAACTCTCCTATTTATGAGC
                         W15511_cgd5020  (101) AACCCTTAGATTCTGGGATTTTAGGATGATAAACTCTCCTATTTATGAGC
W15511_cgd5020_WGA_illustra_GenomiPhi  (101) AACCCTTAGATTCTGGGATTTTAGGATGATAAACTCTCCTATTTATGAGC
           W15511_cgd5020_WGA_REPLI-g  (101) AACCCTTAGATTCTGGGATTTTAGGATGATAAACTCTCCTATTTATGAGC
        W15511_cgd5020_WGA_GenomePlex  (101) AACCCTTAGATTCTGGGATTTTAGGATGATAAACTCTCCTATTTATGAGC
                         W15516_cgd5020  (101) AACCCTTAGATTCTGGGATTTTAGGATGATAAACTCTCCTATTTATGAGC
W15516_cgd5020_WGA_illustra_GenomiPhi  (101) AACCCTTAGATTCTGGGATTTTAGGATGATAAACTCTCCTATTTATGAGC
           W15516_cgd5020_WGA_REPLI-g  (101) AACCCTTAGATTCTGGGATTTTAGGATGATAAACTCTCCTATTTATGAGC
        W15516_cgd5020_WGA_GenomePlex  (101) AACCCTTAGATTCTGGGATTTTAGGATGATAAACTCTCCTATTTATGAGC
                         W15507_cgd5020  (101) AACCCTTAGATTCTGGGATTTTAGGATGATAAACTCTCCTATTTATGAGC
W15507_cgd5020_WGA_illustra_GenomiPhi  (101) AACCCTTAGATTCTGGGATTTTAGGATGATAAACTCTCCTATTTATGAGC
           W15507_cgd5020_WGA_REPLI-g  (101) AACCCTTAGATTCTGGGATTTTAGGATGATAAACTCTCCTATTTATGAGC
        W15507_cgd5020_WGA_GenomePlex  (101) AACCCTTAGATTCTGGGATTTTAGGATGATAAACTCTCCTATTTATGAGC
                         W15519_cgd5020  (101) AACCCTTAGATTCTGGGATTTTAGGATGATAAACTCTCCTATTTATGAGC
WW15519_cgd5020_WGA_illustraGenomiPhi  (101) AACCCTTAGATTCTGGGATTTTAGGATGATAAACTCTCCTATTTATGAGC
       WW15519_cgd5020_WGA_GenomePlex  (101) AACCCTTAGATTCTGGGATTTTAGGATGATAAACTCTCCTATTTATGAGC
                                             151                                             200
                         IOWA_cgd5020  (151) ACAAAACTAATGGAGGAATTTGGAGAATTAACCAATTTGAAGATTACCTT
  IOWA_cgd5020_WGA_illustra_GenomiPhi  (151) ACAAAACTAATGGAGGAATTTGGAGAATTAACCAATTTGAAGATTACCTT
             IOWA_cgd5020_WGA_REPLI-g  (151) ACAAAACTAATGGAGGAATTTGGAGAATTAACCAATTTGAAGATTACCTT
          IOWA_cgd5020_WGA_GenomePlex  (151) ACAAAACTAATGGAGGAATTTGGAGAATTAACCAATTTGAAGATTACCTT
                         W15511_cgd5020  (151) ACAAAACTAATGGAGGAATTTGGAGAATTAACCAATTTGAAGATTACCTT
W15511_cgd5020_WGA_illustra_GenomiPhi  (151) ACAAAACTAATGGAGGAATTTGGAGAATTAACCAATTTGAAGATTACCTT
           W15511_cgd5020_WGA_REPLI-g  (151) ACAAAACTAATGGAGGAATTTGGAGAATTAACCAATTTGAAGATTACCTT
        W15511_cgd5020_WGA_GenomePlex  (151) ACAAAACTAATGGAGGAATTTGGAGAATTAACCAATTTGAAGATTACCTT
                         W15516_cgd5020  (151) ACAAAACTAATGGAGGAATTTGGAGAATTAACCAATTTGAAGATTACCTT
W15516_cgd5020_WGA_illustra_GenomiPhi  (151) ACAAAACTAATGGAGGAATTTGGAGAATTAACCAATTTGAAGATTACCTT
           W15516_cgd5020_WGA_REPLI-g  (151) ACAAAACTAATGGAGGAATTTGGAGAATTAACCAATTTGAAGATTACCTT
        W15516_cgd5020_WGA_GenomePlex  (151) ACAAAACTAATGGAGGAATTTGGAGAATTAACCAATTTGAAGATTACCTT
                         W15507_cgd5020  (151) ACAAAACTAATGGAGGAATTTGGAGAATTAACCAATTTGAAGATTACCTT
W15507_cgd5020_WGA_illustra_GenomiPhi  (151) ACAAAACTAATGGAGGAATTTGGAGAATTAACCAATTTGAAGATTACCTT
           W15507_cgd5020_WGA_REPLI-g  (151) ACAAAACTAATGGAGGAATTTGGAGAATTAACCAATTTGAAGATTACCTT
        W15507_cgd5020_WGA_GenomePlex  (151) ACAAAACTAATGGAGGAATTTGGAGAATTAACCAATTTGAAGATTACCTT
                         W15519_cgd5020  (151) ACAAAACTAATGGAGGAATTTGGAGAATTAACCAATTTGAAGATTACCTT
WW15519_cgd5020_WGA_illustraGenomiPhi  (151) ACAAAACTAATGGAGGAATTTGGAGAATTAACCAATTTGAAGATTACCTT
       WW15519_cgd5020_WGA_GenomePlex  (151) ACAAAACTAATGGAGGAATTTGGAGAATTAACCAATTTGAAGATTACCTT
                                             201
                         IOWA_cgd5020  (201) GGA
  IOWA_cgd5020_WGA_illustra_GenomiPhi  (201) GGA
             IOWA_cgd5020_WGA_REPLI-g  (201) GGA
          IOWA_cgd5020_WGA_GenomePlex  (201) GGA
                         W15511_cgd5020  (201) GGA
W15511_cgd5020_WGA_illustra_GenomiPhi  (201) GGA
           W15511_cgd5020_WGA_REPLI-g  (201) GGA
        W15511_cgd5020_WGA_GenomePlex  (201) GGA
                         W15516_cgd5020  (201) GGA
W15516_cgd5020_WGA_illustra_GenomiPhi  (201) GGA
           W15516_cgd5020_WGA_REPLI-g  (201) GGA
        W15516_cgd5020_WGA_GenomePlex  (201) GGA
                         W15507_cgd5020  (201) GGA
W15507_cgd5020_WGA_illustra_GenomiPhi  (201) GGA
           W15507_cgd5020_WGA_REPLI-g  (201) GGA
        W15507_cgd5020_WGA_GenomePlex  (201) GGA
                         W15519_cgd5020  (201) GGA
WW15519_cgd5020_WGA_illustraGenomiPhi  (201) GGA
       WW15519_cgd5020_WGA_GenomePlex  (201) GGA
```

Chro.20156 gene PCR products



Appendix VII

Appendix VII: Alignment of Cops-1 PCR product sequences from clinical isolates and reference strains amplified using Cgd2_4380 F and R primers.



```
          Ch4_cgd2_4380  (205) ------------------------------------------------------------
        TU502_cgd2_4380  (205) ------------------------------------------------------------
                                301                                                       360
          Cp2_cgd2_4380  (301) TCTACAAGTACAGGAACTCGTGCTCTACAAAGCGGTGTAGGAAGCAGATTTTTGAGTCCA
          Cp3_cgd2_4380  (301) TCTACAAGTACAGGAACTCGTGCTCTACAAAGCGGTGTAGGAAGCAGATTTTTGAGTCCA
          Cp4_cgd2_4380  (301) TCTACAAGTACAGGAACTCGTGCTCTACAAAGCGGTGTAGGAAGCAGATTTTTGAGTCCA
         Iowa_cgd2_4380  (301) TCTACAAGTACAGGAACTCGTGCTCTACAAAGCGGTGTAGGAAGCAGATTTTTGAGTCCA
      Moredun_cgd2_4380  (301) TCTACAAGTACAGGAACTCGTGCTCTACAAAGCGGTGTAGGAAGCAGATTTTTGAGTCCA
          W65_cgd2_4380  (301) TCTACAAGTACAGGAACTCGTGCTCCACAAAGCGGTGTAGGAAGCAGATTTTTGAGTCCA
          W66_cgd2_4380  (301) TCTACAAGTACAGGAACTCGTGCTCCACAAAGCGGTGTAGGAAGCAGATTTTTGAGTCCA
          W67_cgd2_4380  (301) TCTACAAGTACAGGAACTCGTGCTCCACAAAGCGGTGTAGGAAGCAGATTTTTGAGTCCA
          W70_cgd2_4380  (301) TCTACAAGTACAGGAACTCGTGCTCCACAAAGCGGTGTAGGAAGCAGATTTTTGAGTCCA
C.meleagridis_cgd2_4380  (301) TCTACAAGTACAGGAACTCGTGCTCTACAAAGCGGTGTAGGAAGCAGATTTTTGAGTCCA
          Ch2_cgd2_4380  (205) ------------------------------------------------------------
          Ch3_cgd2_4380  (205) ------------------------------------------------------------
          Ch4_cgd2_4380  (205) ------------------------------------------------------------
        TU502_cgd2_4380  (205) ------------------------------------------------------------
                                361                                                       420
          Cp2_cgd2_4380  (361) GGATATGGTTCAAGACCAGGTTCAACACGTGGTCCAACTTTAGGGCTATTTACTAGATCA
          Cp3_cgd2_4380  (361) GGATATGGTTCAAGACCAGGTTCAACACGTGGTCCAACTTTAGGGCTATTTACTAGATCA
          Cp4_cgd2_4380  (361) GGATATGGTTCAAGACCAGGTTCAACACGTGGTCCAACTTTAGGGCTATTTACTAGATCA
         Iowa_cgd2_4380  (361) GGATATGGTTCAAGACCAGGTTCAACACGTGGTCCAACTTTAGGGCTATTTACTAGATCA
      Moredun_cgd2_4380  (361) GGATATGGTTCAAGACCAGGTTCAACACGTGGTCCAACTTTAGGGCTATTTACTAGATCA
          W65_cgd2_4380  (361) GGATATGGTTCAAGACCAGGTTCAACACGTGGTCCAACTTTAGGGCTATTTACTAGATCA
          W66_cgd2_4380  (361) GGATATGGTTCAAGACCAGGTTCAACACGTGGTCCAACTTTAGGGCTATTTACTAGATCA
          W67_cgd2_4380  (361) GGATATGGTTCAAGACCAGGTTCAACACGTGGTCCAACTTTAGGGCTATTTACTAGATCA
          W70_cgd2_4380  (361) GGATATGGTTCAAGACCAGGTTCAACACGTGGTCCAACTTTAGGGCTATTTACTAGATCA
C.meleagridis_cgd2_4380  (361) GGATATGGTTCAAGACCAGGTTCAACACGTGGTCCAACTTTAGGGCTATTTACTAGATCA
          Ch2_cgd2_4380  (205) ------------------------------------------------------------
          Ch3_cgd2_4380  (205) ------------------------------------------------------------
          Ch4_cgd2_4380  (205) ------------------------------------------------------------
        TU502_cgd2_4380  (205) ------------------------------------------------------------
                                421                                                       480
          Cp2_cgd2_4380  (421) CGTCCAGTTTTTCCAACTCGTAGACCATATTCAGGAAGTTTGTTTACTTCTAGCAGTTTT
          Cp3_cgd2_4380  (421) CGTCCAGTTTTTCCAACTCGTAGACCATATTCAGGAAGTTTGTTTACTTCTAGCAGTTTT
          Cp4_cgd2_4380  (421) CGTCCAGTTTTTCCAACTCGTAGACCATATTCAGGAAGTTTGTTTACTTCTAGCAGTTTT
         Iowa_cgd2_4380  (421) CGTCCAGTTTTTCCAACTCGTAGACCATATTCAGGAAGTTTGTTTACTTCTAGCAGTTTT
      Moredun_cgd2_4380  (421) CGTCCAGTTTTTCCAACTCGTAGACCATATTCAGGAAGTTTGTTTACTTCTAGCAGTTTT
          W65_cgd2_4380  (421) CGTCCAGTTTTTCCAACTCGTAGACCATATTCAGGAAGTTTGCTTACTTCTAGCAGTTTT
          W66_cgd2_4380  (421) CGTCCAGTTTTTCCAACTCGTAGACCATATTCAGGAAGTTTGCTTACTTCTAGCAGTTTT
          W67_cgd2_4380  (421) CGTCCAGTTTTTCCAACTCGTAGACCATATTCAGGAAGTTTGCTTACTTCTAGCAGTTTT
          W70_cgd2_4380  (421) CGTCCAGTTTTTCCAACTCGTAGACCATATTCAGGAAGTTTGCTTACTTCTAGCAGTTTT
C.meleagridis_cgd2_4380  (421) CGTCCAGTTTTTCCAACTCGTAGACCATATTCAGGAAGTTTGTTTACTTCTAGCAGTTTT
          Ch2_cgd2_4380  (205) ------------------------------------------------------------
          Ch3_cgd2_4380  (205) ------------------------------------------------------------
          Ch4_cgd2_4380  (205) ------------------------------------------------------------
        TU502_cgd2_4380  (205) ------------------------------------------------------------
                                481                                                       540
          Cp2_cgd2_4380  (481) AGATCTTCTAATGCTTCAGACGGATCAGGAGATTCGTCATATAGTTCTCGTTTTACAGGT
          Cp3_cgd2_4380  (481) AGATCTTCTAATGCTTCAGACGGATCAGGAGATTCGTCATATAGTTCTCGTTTTACAGGT
          Cp4_cgd2_4380  (481) AGATCTTCTAATGCTTCAGACGGATCAGGAGATTCGTCATATAGTTCTCGTTTTACAGGT
         Iowa_cgd2_4380  (481) AGATCTTCTAATGCTTCAGACGGATCAGGAGATTCGTCATATAGTTCTCGTTTTACAGGT
      Moredun_cgd2_4380  (481) AGATCTTCTAATGCTTCAGACGGATCAGGAGATTCGTCATATAGTTCTCGTTTTACAGGT
          W65_cgd2_4380  (481) AGATCTTCTAATGCTTCAGGCGGATCAGGAGATTCGTCATATAGTTCTCGTTTTACAGGT
          W66_cgd2_4380  (481) AGATCTTCTAATGCTTCAGGCGGATCAGGAGATTCGTCATATAGTTCTCGTTTTACAGGT
          W67_cgd2_4380  (481) AGATCTTCTAATGCTTCAGGCGGATCAGGAGATTCGTCATATAGTTCTCGTTTTACAGGT
          W70_cgd2_4380  (481) AGATCTTCTAATGCTTCAGGCGGATCAGGAGATTCGTCATATAGTTCTCGTTTTACAGGT
C.meleagridis_cgd2_4380  (481) AGATCTTCTAATGCTTCAGACGGATCAGGAGATTCGTCATATAGTTCTCGTTTTACAGGT
          Ch2_cgd2_4380  (205) ------------------------------------------------------------
          Ch3_cgd2_4380  (205) ------------------------------------------------------------
          Ch4_cgd2_4380  (205) ------------------------------------------------------------
        TU502_cgd2_4380  (205) ------------------------------------------------------------
                                541                                                       600
          Cp2_cgd2_4380  (541) ACAGGAACTCGTGGTTCACAAGGCGGTGTAGGAAGCAGATTTTTGAGTCCAGGATACGGT
          Cp3_cgd2_4380  (541) ACAGGAACTCGTGGTTCACAAGGCGGTGTAGGAAGCAGATTTTTGAGTCCAGGATACGGT
          Cp4_cgd2_4380  (541) ACAGGAACTCGTGGTTCACAAGGCGGTGTAGGAAGCAGATTTTTGAGTCCAGGATACGGT
         Iowa_cgd2_4380  (541) ACAGGAACTCGTGGTTCACAAGGCGGTGTAGGAAGCAGATTTTTGAGTCCAGGATACGGT
      Moredun_cgd2_4380  (541) ACAGGAACTCGTGGTTCACAAGGCGGTGTAGGAAGCAGATTTTTGAGTCCAGGATACGGT
          W65_cgd2_4380  (541) ACAGGAACTCGTGGTTCACAAGGCGGTGTAGGAAGCAGATTTTTGAGTCCAGGATACGGT
          W66_cgd2_4380  (541) ACAGGAACTCGTGGTTCACAAGGCGGTGTAGGAAGCAGATTTTTGAGTCCAGGATACGGT
          W67_cgd2_4380  (541) ACAGGAACTCGTGGTTCACAAGGCGGTGTAGGAAGCAGATTTTTGAGTCCAGGATACGGT
          W70_cgd2_4380  (541) ACAGGAACTCGTGGTTCACAAGGCGGTGTAGGAAGCAGATTTTTGAGTCCAGGATACGGT
C.meleagridis_cgd2_4380  (541) ACAGGAACTCGTGGTTCACAAGGCGGTGTAGGAAGCAGATTTTTGAGTCCAGGATACGGT
          Ch2_cgd2_4380  (205) ------------------------------------------------------------
          Ch3_cgd2_4380  (205) ------------------------------------------------------------
          Ch4_cgd2_4380  (205) ------------------------------------------------------------
        TU502_cgd2_4380  (205) ------------------------------------------------------------
                                601                                                   652
          Cp2_cgd2_4380  (601) TTACAACCAGGTTCAGCACGCGGTCCAACTTTGGGGCTATTTACTAGATCAC
          Cp3_cgd2_4380  (601) TTACAACCAGGTTCAGCACGCGGTCCAACTTTGGGGCTATTTACTAGATCAC
          Cp4_cgd2_4380  (601) TTACAACCAGGTTCAGCACGCGGTCCAACTTTGGGGCTATTTACTAGATCAC
```

```
          Iowa_cgd2_4380  (601) TTACAACCAGGTTCAGCACGCGGTCCAACTTTGGGGCTATTTACTAGATCAC
       Moredun_cgd2_4380  (601) TTACAACCAGGTTCAGCACGCGGTCCAACTTTGGGGCTATTTACTAGATCAC
           W65_cgd2_4380  (601) TTACAACCAGGTTCAGCACGCGGTCCAACTTTGGGGCTATTTACTAGATCAC
           W66_cgd2_4380  (601) TTACAACCAGGTTCAGCACGCGGTCCAACTTTGGGGCTATTTACTAGATCAC
           W67_cgd2_4380  (601) TTACAACCAGGTTCAGCACGCGGTCCAACTTTGGGGCTATTTACTAGATCAC
           W70_cgd2_4380  (601) TTACAACCAGGTTCAGCACGCGGTCCAACTTTGGGGCTATTTACTAGATCAC
C.meleagridis_cgd2_4380  (601) TTACAACCAGGTTCAGCACGCGGTCCA------------------------
          Ch2_cgd2_4380  (205) --------------------------------------------------
          Ch3_cgd2_4380  (205) --------------------------------------------------
          Ch4_cgd2_4380  (205) --------------------------------------------------
        TU502_cgd2_4380  (205) --------------------------------------------------
```

Appendix VIII

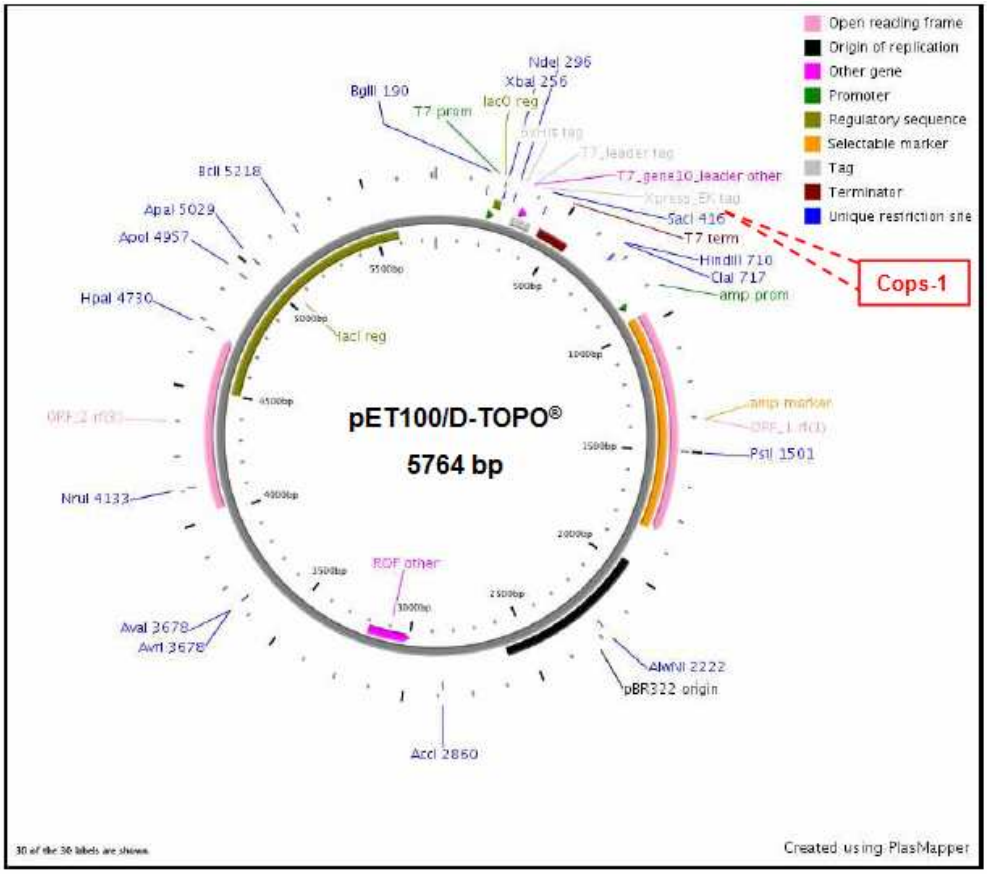
Appendix VIII: Alignment of Cops-1 gene and downstream sequences from *C. parvum* and *C. hominis*. The gene sequence was retrieved using a primer walking approach. The 3' end of the gene in *C. hominis* was determined by the identification of a stop codon in the right open reading frame (ORF). The stop codon is highlighted in blue.

```
                    1                                                    60
 C. parvum      (1) ATGGGTAATAGTTTAAATGTTTTTTTTGCTATTTTTCTAGTTGTTTTTCTTAATTTTTTG
C. hominis      (1) ATGGGTAATAGTTTAAATGTTTTTTTTGTTATTTTTCTAGTTGTTTTTCTTAATTTTTTG
                   61                                                   120
 C. parvum     (61) GGGTTACAAACTTTTTTTAATACCAGAAATGTTGAAAGCAATTTATTTTTAATTTTAAAT
C. hominis     (61) GGGTTCCAAACTTTTTTTAATACCGGAAATGTTGAAAGCAATTTATTTTTTATTTTAAAT
                  121                                                   180
 C. parvum    (121) TCTTATCCATCATTCATTAAGCTTGGAGGAATAAATGGTAGAGAAGGGAGTTCTAGTGGA
C. hominis    (121) TCTTATCCATCATTCATTAAGCTTGGAGGAATAAAAGGTAGAGAAGGGAGTTCTAACGGA
                  181                                                   240
 C. parvum    (181) TTTTCGTCTGGAGGTAGACATGGTTCATTGCAAGGAGGATCATTAAGAGATAGTGCGCGT
C. hominis    (181) TTTTCGTCTGGAGGTGGACATAGTTCATTTCAAGGAAGATCATTAAGAGATAGTGGGCGT
                  241                                                   300
 C. parvum    (241) TCAAGGGGTGGACCTAGATGCTCAAGAGCCCCGCATCCTAGACTCCAAACCATTATTGAG
C. hominis    (241) TCAAGGGGTAGATCTAGATGCTCAAGAGCTCCGGATCCTAAACTCCAAACCATTATTGAG
                  301                                                   360
 C. parvum    (301) TGTTCAGAAACAGATTCAACTGATGGAGGTAGCAATACTGCAAGTCAACCAAATGGCAGA
C. hominis    (301) TGTTCAGAAACAGATTCACCTGATGGAGTTAGCAATACTGGAAATCAACCAACTGGCAGA
                  361                                                   420
 C. parvum    (361) TTTTTAAATCCAGGATATGGTTCACGACCAGG---------TTCAACACGTGGTCCAACT
C. hominis    (361) TTTTCAAATCCAGGAGGTGGCTTGCAACCAGGCCCACGTGTTGGATCACGTGGGCCAACT
                  421                                                   480
 C. parvum    (412) TTAGGGCTATTTACTAGATCACGTCCAGTTTTTCCAACTCGTAGACCATATTCAGGAATT
C. hominis    (421) TTAGGGCTATTTACTAGATCACGTCCAACTATTCCAACTCGTAGACCATATTCAGGAATT
                  481                                                   540
 C. parvum    (472) TTGCTTACCTCTAGTGGTTCGAAGTCCTCTGCTCTTTCAAGCAGATTTGGACAAAAGCCA
C. hominis    (481) TTGCTTACCTCTGGTGGGTTGAACTCTTCTGGTCTTTCAAGCAGATTAGGAAAGCATCA
                  541                                                   600
 C. parvum    (532) TCAAGTTCTCATTCTACAAGTACAGGAACTCGTGCTCTACAAAGCGGTGTAGGAAGCAGA
C. hominis    (541) TCGAGTTCTCATTCTACAAGTGTAGGAACTCGTGGTCCACAAAGCGGTGTAGGAAGCAAA
                  601                                                   660
 C. parvum    (592) TTTTTGAGTCCAGGATATGGTTCAAGACCAGGTTCAACACGTGGTCCAACTTTAGGGCTA
C. hominis    (601) TTTTTGAGTCCAGGATATGGTTCACGACCAGGTTCAGCACGCGGTCCAACTTTAGGGCTA
                  661                                                   720
 C. parvum    (652) TTTACTAGATCACGTCCAGTTTTTCCAACTCGTAGACCATATTCAGGAAGTTTGTTTACT
C. hominis    (661) TTTACTAGATCAAGTCCAACTCTTCCAACTCGTAGACCATATTCAGGAAGTTTGCTTACT
                  721                                                   780
 C. parvum    (712) TCTAGCAGTTTTAGATCTTCTAATGCTTCAGACGGATCAGGAGATTCGTCATATAGTTCT
C. hominis    (721) TCTAGCAGTTTAAGTTCTTCTAATGCTTCAGGTGGATTAGGACAATCGTCATATAGTTCT
                  781                                                   840
 C. parvum    (772) CGTTTTACAGGTACAGGAACTCGTGGTTCACAAGGCGGTGTAGGAAGCAGATTTTTGAGT
C. hominis    (781) CGTTTTACAAGTACAGGAACTC---------AAGG------------------------
                  841                                                   900
 C. parvum    (832) CCAGGATACGGTTTACAACCAGGTTCAGCACGCGGTCCAACTTTGGGGCTATTTACTAGA
C. hominis    (807) ------------------------------------------------------------
                  901                                                   960
 C. parvum    (892) TCACGCCCACCTCTTCCAACTCGTAAACCATATTCAGGAAGTTTGCTTACTTCTAGCAGA
C. hominis    (807) ---------------------------ACCATAT---GGAA--------------------
                  961                                                  1020
 C. parvum    (952) TTGAGTTCTTCTAATGCTTCAGGTGGATTAGGACAATCGTCATCTAGTTCTCGTTTACA
C. hominis    (818) ----------------CTTCAGGT--AATAGA----------------------TTTTCT
                 1021                                                  1080
 C. parvum   (1012) AGTACAGGACCTCAAGGACCA--TATGGAACTTCAGGAGTTGGTACACCGTTGGGCCACT
C. hominis    (838) AGTATTGGA-----AGGACCAGTTCTGTAACTTCAGGAGCTGGTACAGCATCGGGCCAGT
                 1081                                                  1140
 C. parvum   (1070) CTGTTTCACCTGAAGGAAAACCACAAGGTTTACTAGCTAGAGGATATATAACATCAAATT
C. hominis    (893) CTGTTTCATCTGAAAGAATACCACAAGGTTTACAAGCTAGAGGATATATGACATCAAATT
                 1141                                                  1200
 C. parvum   (1130) GTCCACGTGGTATTCCAGGTGAGCATCGAGTAGATGTTACTAGTAACGGTTCTTTGATAT
C. hominis    (953) GTCCACGTGGCATTCCAGGTGAGCATCGAGTAGATATTACTAGTAACGGTTCTTTGATGT
                 1201                                                  1260
 C. parvum   (1190) GCTGTTATTGTTATAATAGATGTGATCATGAAGGTTTTAAGCCACCAAGACGAACAACAA
C. hominis   (1013) GGTGTTATTGTTATAATAGATGTGATCATGAAGGTTTTAAGCCACCAAGACGAACAACGA
                 1261                                                  1320
 C. parvum   (1250) CAACAACAACACAATCACCACCATATTCTTCTAGGGGGTACCTGACATTAGATTGTCCAC
C. hominis   (1073) CAACAACTACACAATCACCACCATATTCTTCTAGGGGGTACTTGACATTAGATTGTCCAC
                 1321                                                  1380
```

```
 C. parvum  (1310) TTGGTACCCCAGGTGAACATCGCTTAGATGTTGATAATTCTGGCGTCTTATTTTGCGCTA
C. hominis  (1133) TTGGTACCCCAGGCGAACATCGCTTAGATGTTACTAGTGATGGCGTGTTATTTTGTGCTA
                   1381                                                 1440
 C. parvum  (1370) CTTGTGGTAACAGGTTTAGTCATCAAGGGTGCCCACCACCGAAAATACCGTTATGCCGAA
C. hominis  (1193) CTTGTGGTAACAAGTTTAGTCATCAAGGGTGCCCACCACCGAAAATACCGAAATGCCGAA
                   1441                                                 1500
 C. parvum  (1430) AATAAAACATGAAACAGCACCCATTTCATTATTATTTTCAAACATTTTACACCAAAAACA
C. hominis  (1253) AA-AAAACATGAAACAGCACCTATTTCATTATTATTTTCGATCATCTTACACTATAAACA
                   1501                                                 1560
 C. parvum  (1490) GTTTGCGGTCTAAATACGGCTGCTATATTTTATTTTATATAGATTCCGTTTTTACTTTTG
C. hominis  (1312) GTTTGCGGTCTAAATACGGCTGCTATATTTTATTTTATATAGATTCCGTTTTTACTTTTG
                   1561                                                 1620
 C. parvum  (1550) TAGCACAAAAAATGATTGGATTGAAGTCACCGAGTTTTGTAATAGTTTATTTCTAAAAAGG
C. hominis  (1372) TAGCGCAAAAAATAATTGGATTGAAGTTACCGAGTTTTTTAATGATTTGTTTCTAAAAATG
                   1621                                                 1680
 C. parvum  (1610) CTAGTAAAGTTGCTTATTTTAATATCTATCCTTTACAATGTTATGTTATAAGTTTGTGCT
C. hominis  (1432) CTAGTAAAGTTGCTTATTTTAATATCTTTCCTTTACAATGTTAT-----AAGTTTATATT
                   1681                                                 1740
 C. parvum  (1670) ATAATTTCGAG---ATCTTAATGCAAAGATTCTTGGATTATAACTTTCGAAACTTTAGGG
C. hominis  (1487) ATCATTTCAAGAAGATTTTAATGCAAAGATTCTTGGATTATAGCTTTCAGCCAATTAGTG
                   1741     1753
 C. parvum  (1727) TT--TTAGGGTTT
C. hominis  (1547) ATCGTAATGGTTT
```

Appendix IX

Appendix IX: Diagram of the recombinant pET100/D-TOPO® plasmid (Invitrogen) containing Cops-1 gene. The cloning site corresponds to the Topoisomerase recognition site (position: 396-400) and the overhang sequence (401-404). The position of Cops-1 is shown in red. The plasmid map was drawn using PlasMapper software (http://wishart.biology.ualberta.ca/PlasMapper/).



References

Abrahamsen, M., Templeton, TJ., Enomoto, S., Abrahante, JE., Zhu, G., Lancto, CA., Deng, M, Liu, C, Widmer, G, Tzipori, S, Buck, GA, Xu, P, Bankier, AT, Dear, PH, Konfortov, BA, Spriggs, HF, Iyer, L, Anantharaman, V, Aravind, L, Kapur, V. (2004). "Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*." Science. **304**(5669): 441-5.

Abubakar, I., Aliyu, SH, Arumugam, C, Usman, NK, Hunter, PR. (2007). "Treatment of cryptosporidiosis in immunocompromised individuals: systematic review and meta-analysis." Br J Clin Pharmacol. **63**(4): 387-93.

Adams, R. B., Guerrant, R. L., Zu, S., Fang, G., Roche, J. K. (1994). "*Cryptosporidium parvum* infection of intestinal epithelium: morphologic and functional studies in an *in vitro* model." J Infect Dis. **169**(1): 170-7.

Ahn, S., Costa, J, Emanuel, JR. (1996). "PicoGreen quantitation of DNA: effective evaluation of samples pre- or post-PCR." Nucleic Acids Res. **24**(13): 2623-5.

Alles, A., Waldron, MA, Sierra, LS, Mattia, AR. (1995). "Prospective comparison of direct immunofluorescence and conventional staining methods for detection of *Giardia* and *Cryptosporidium* spp. in human fecal specimens." J Clin Microbiol. **33**(6): 1632-4.

Alonso-Monge, R., Navarro-García, F, Román, E, Eisman, B, Nombela, C, Pla, J. (2003). "Strategies for the identification of virulence determinants in human pathogenic fungi." Curr Genet. **42**(6): 301-12.

Altschul, S., Gish, W, Miller, W, Myers, EW, Lipman, DJ. (1990). "Basic local alignment search tool." J Mol Biol. **215**(3): 403-10.

Alvarez-Pellitero, P., Quiroga, M. I., Sitja-Bobadilla, A., Redondo, M. J., Palenzuela, O., Padros, F., Vazquez, S., Nieto, J. M. (2004). "*Cryptosporidium scophthalm*i n. sp. (Apicomplexa: Cryptosporidiidae) from cultured turbot *Scophthalmus maximus*. Light and electron microscope description and histopathological study." Dis Aquat Organ. **62**(1-2): 133-45.

Alvarez-Pellitero, P., Sitjà-Bobadilla, A. (2002). "*Cryptosporidium molnari* n. sp. (Apicomplexa: Cryptosporidiidae) infecting two marine fish species, *Sparus aurata* L. and *Dicentrarchus labrax* L." Int J Parasitol. **32**(8): 1007-21.

Alves, M., Xiao, L, Antunes, F, Matos, O. (2006). "Distribution of *Cryptosporidium* subtypes in humans and domestic and wild ruminants in Portugal." Parasitol Res. **99**(3): 287-92.

Alves, M., Xiao, L, Sulaiman, I, Lal, AA, Matos, O, Antunes, F. (2003 a). "Subgenotype analysis of *Cryptosporidium* isolates from humans, cattle, and zoo ruminants in Portugal." J Clin Microbiol. **41**(6): 2744-7.

Alves, M., Matos, O., Antunes, F. (2003 b). "Microsatellite analysis of *Cryptosporidium hominis* and *C. parvum* in Portugal: a preliminary study." J Eukaryot Microbiol. **50**: 529-30.

Amadi, B., Mwiya, M, Sianongo, S, Payne, L, Watuka, A, Katubulushi, M, Kelly, P. (2009). "High dose prolonged treatment with nitazoxanide is not effective for cryptosporidiosis in HIV positive Zambian children: a randomised controlled trial." BMC Infect Dis. **9**(1): 195.

Amadi, B., Mwiya, M, Musuku, J, Watuka, A, Sianongo, S, Ayoub, A, Kelly, P. (2002). "Effect of nitazoxanide on morbidity and mortality in Zambian children with cryptosporidiosis: a randomised controlled trial." Lancet. **360**(9343): 1375-80.

Anderson, T., Haubold, B, Williams, JT, Estrada-Franco, JG, Richardson, L, Mollinedo, R, Bockarie, M, Mokili, J, Mharakurwa, S, French, N, Whitworth, J, Velez, ID, Brockman, AH, Nosten, F, Ferreira, MU, Day, KP. (2000).

"Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*." Mol Biol Evol. **17**(10): 1467-82.

Anderson, R. M., May, R. M. (1982). "Coevolution of hosts and parasites." Parasitology **85**: 411-26.

Anusz, K., Mason, PH, Riggs, MW, Perryman, LE. (1990). "Detection of *Cryptosporidium parvum* oocysts in bovine feces by monoclonal antibody capture enzyme-linked immunosorbent assay." J Clin Microbiol. **28**(12): 2770-4.

Arrowood, M., Xie, LT, Hurd, MR. (1994). "*In vitro* assays of maduramicin activity against *Cryptosporidium parvum*." J Eukaryot Microbiol. **41**(5): 23S.

Arrowood, M. J., Sterling, C. R., Healey, M. C. (1991). "Immunofluorescent microscopical visualization of trails left by gliding *Cryptosporidium parvum* sporozoites." J Parasitol. **77**(2): 315-7.

Arrowood, M., Sterling, CR. (1989). "Comparison of conventional staining methods and monoclonal antibody-based methods for *Cryptosporidium* oocyst detection." J Clin Microbiol. **27**(7): 1490-5.

Awad-El-Kariem, F. (1999). "Does *Cryptosporidium parvum* have a clonal population structure?" Parasitol Today. **15**(12): 502-4.

Awad-el-Kariem, F., Warhurst, DC, McDonald, V. (1994). "Detection and species identification of *Cryptosporidium* oocysts using a system based on PCR and endonuclease restriction." Parasitology. **109**: 19-22.

Balatbat, A., Jordan, GW, Tang, YJ, Silva, J Jr. (1996). "Detection of *Cryptosporidium parvum* DNA in human feces by nested PCR." J Clin Microbiol. **34**(7): 1769-72.

Barker, D., Hansen, MS, Faruqi, AF, Giannola, D, Irsula, OR, Lasken, RS, Latterich, M, Makarov, V, Oliphant, A, Pinter, JH, Shen, R, Sleptsova, I, Ziehler, W, Lai, E. (2004). "Two methods of whole-genome amplification enable accurate genotyping across a 2320-SNP linkage panel." Genome Res. **14**(5): 901-7.

Barnes, D. A., Bonnin, A., Huang, J. X., Gousset, L., Wu, J., Gut, J., Doyle, P., Dubremetz, J. F., Ward, H., Petersen, C. (1998). "A novel multi-domain mucin-like glycoprotein of *Cryptosporidium parvum* mediates invasion." Mol Biochem Parasitol. **96**(1-2): 93-110.

Barry, J. D., Ginger, M. L., Burton, P., McCulloch, R. (2003). "Why are parasite contingency genes often associated with telomeres?" Int J Parasitol. **33**(1): 29-45.

Barta, J. R., Thompson, R.C. (2006). "What is *Cryptosporidium*? Reappraising its biology and phylogenetic affinities." Trends Parasitol. **22**(10): 463-8.

Barta, J., Jenkins, MC, Danforth, HD. (1991). "Evolutionary relationships of avian *Eimeria* species among other Apicomplexan protozoa: monophyly of the apicomplexa is supported." Mol Biol Evol. **8**(3): 345-55.

Baxby, D., Blundell, N, Hart, CA. (1984). "The development and performance of a simple, sensitive method for the detection of *Cryptosporidium* oocysts in faeces." J Hyg (Lond). **93**(2): 317-23.

Becher, K. A., Robertson, I. D., Fraser, D. M., Palmer, D. G., Thompson, R. C. (2004). "Molecular epidemiology of *Giardia* and *Cryptosporidium* infections in dairy calves originating from three sources in Western Australia." Vet Parasitol. **123**(1-2): 1-9.

Beck, H. P., Blake, D., Darde, M. L., Felger, I., Pedraza-Diaz, S., Regidor-Cerrillo, J., Gomez-Bautista, M., Ortega-Mora, L. M., Putignani, L., Shiels, B., Tait,

A., Weir, W. (2009). "Molecular approaches to diversity of populations of apicomplexan parasites." Int J Parasitol. **39**(2): 175-89.

Belosevic, M., Guy, RA, Taghi-Kilani, R, Neumann, NF, Gyürék, LL, Liyanage, LR, Millard, PJ, Finch, GR. (1997). "Nucleic acid stains as indicators of *Cryptosporidium parvum* oocyst viability." Int J Parasitol. **27**(7): 787-98.

Bergen, A., Haque, KA, Qi, Y, Beerman, MB, Garcia-Closas, M, Rothman, N, Chanock, SJ. (2005). "Comparison of yield and genotyping performance of multiple displacement amplification and OmniPlex whole genome amplified DNA generated from multiple DNA sources." Hum Mutat. **26**(3): 262-70.

Bjorneby, J. M., Riggs, M. W., Perryman, L. E. (1990). "*Cryptosporidium parvum* merozoites share neutralization-sensitive epitopes with sporozoites." J Immunol. **145**(1): 298-304.

Black, E., Finch, GR, Taghi-Kilani, R, Belosevic, M. (1996). "Comparison of assays for *Cryptosporidium parvum* oocysts viability after chemical disinfection." FEMS Microbiol Lett. **135**(2-3): 187-9.

Blagburn, B., Lindsay, DS, Giambrone, JJ, Sundermann, CA, Hoerr, FJ. (1987). "Experimental cryptosporidiosis in broiler chickens." Poult Sci. **66**(3): 442-9.

Blanco, M. A., Iborra, A., Vargas, A., Nsie, E., Mba, L., Fuentes, I. (2009). "Molecular characterization of *Cryptosporidium* isolates from humans in Equatorial Guinea." Trans R Soc Trop Med Hyg. **103**(12): 1282-4.

Blanco, L., Bernad, A, Lázaro, JM, Martín, G, Garmendia, C, Salas, M. (1989). "Highly efficient DNA synthesis by the phage phi 29 DNA polymerase. Symmetrical mode of DNA replication." J Biol Chem. **264**(15): 8935-40.

Bloch, H. (1950). "A component of tubercle bacilli concerned with their virulence." Bull N Y Acad Med. **26**(7): 506-7.

Blunt, D., Khramtsov, NV, Upton, SJ, Montelone, BA. (1997). "Molecular karyotype analysis of *Cryptosporidium parvum*: evidence for eight chromosomes and a low-molecular-size molecule." Clin Diagn Lab Immunol. **4**(1): 11-3.

Bonnin, A., Ojcius, D. M., Souque, P., Barnes, D. A., Doyle, P. S., Gut, J., Nelson, R. G., Petersen, C., Dubremetz, J. F. (2001). "Characterization of a monoclonal antibody reacting with antigen-4 domain of gp900 in *Cryptosporidium parvum* invasive stages." Parasitol Res. **87**(8): 589-92.

Bonnin, A., Fourmaux, M. N., Dubremetz, J. F., Nelson, R. G., Gobet, P., Harly, G., Buisson, M., Puygauthier-Toubas, D., Gabriel-Pospisil, G., Naciri, M., Camerlynck, P. (1996). "Genotyping human and bovine isolates of *Cryptosporidium parvum* by polymerase chain reaction-restriction fragment length polymorphism analysis of a repetitive DNA sequence." FEMS Microbiol Lett. **137**(2-3): 207-11.

Bonnin, A., Dubremetz, J. F., Camerlynck, P. (1993). "A new antigen of *Cryptosporidium parvum* micronemes possessing epitopes cross-reactive with macrogamete granules." Parasitol Res. **79**(1): 8-14.

Boulter-Bitzer, J., Lee, H, Trevors, JT. (2007). "Molecular targets for detection and immunotherapy in *Cryptosporidium parvum*." Biotechnol Adv. **25**(1): 13-44.

Bouzid, M., Heavens, D., Elwin, K., Chalmers, R. M., Hadfield, S. J., Hunter, P. R., Tyler, K. M. (2010). "Whole genome amplification (WGA) for archiving and genotyping of clinical isolates of *Cryptosporidium* species." Parasitology. **137**(1): 27-36.

Bouzid, M., Steverding, D., Tyler, K. M. (2008). "Detection and surveillance of waterborne protozoan parasites." Curr Opin Biotechnol. **19**(3): 302-6.

Boxell, A., Hijjawi, N, Monis, P, Ryan, U. (2008). "Comparison of various staining

methods for the detection of *Cryptosporidium* in cell-free culture." Exp Parasitol. **120**(1): 67-72.

Brook, E. J., Anthony Hart, C., French, N. P., Christley, R. M. (2009). "Molecular epidemiology of *Cryptosporidium* subtypes in cattle in England." Vet J. **179**(3): 378-82.

Brownstein, D., Strandberg, JD, Montali, RJ, Bush, M, Fortner, J. (1977). "*Cryptosporidium* in snakes with hypertrophic gastritis." Vet Pathol. **14**(6): 606-17.

Brubaker, R. R. (1985). "Mechanisms of bacterial virulence." Annu Rev Microbiol. **39**: 21-50.

Bukhari, Z., McCuin, RM, Fricker, CR, Clancy, JL. (1998). "Immunomagnetic separation of *Cryptosporidium parvum* from source water samples of various turbidities." Appl Environ Microbiol. **64**(11): 4495-9.

Bull, S., Chalmers, R, Sturdee, AP, Curry, A, Kennaugh, J. (1998). "Cross-reaction of an anti-*Cryptosporidium* monoclonal antibody with sporocysts of Monocystis species." Vet Parasitol. **77**(2-3): 195-7.

Burgos, M., Méndez, JC, Ribon, W. (2004). "Molecular epidemiology of tuberculosis: methodology and applications." Biomedica. **24**(Supp 1): 188-201.

Cacciò, S., Pozio, E. (2006). "Advances in the epidemiology, diagnosis and treatment of cryptosporidiosis." Expert Rev Anti Infect Ther. **4**(3): 429-43.

Cacciò, S. (2005). "Molecular epidemiology of human cryptosporidiosis." Parassitologia. **47**(2): 185-92.

Caccio, S., Spano, F., Pozio, E. (2001). "Large sequence variation at two microsatellite loci among zoonotic (genotype C) isolates of *Cryptosporidium parvum*." Int J Parasitol. **31**(10): 1082-6.

Cacciò, S., Homan, W, Camilli, R, Traldi, G, Kortbeek, T, Pozio, E. (2000). "A microsatellite marker reveals population heterogeneity within human and animal genotypes of *Cryptosporidium parvum*." Parasitology. **120**: 237-44.

Cama, V. A., Ross, J. M., Crawford, S., Kawai, V., Chavez-Valdez, R., Vargas, D., Vivar, A., Ticona, E., Navincopa, M., Williamson, J., Ortega, Y., Gilman, R. H., Bern, C., Xiao, L. (2007). "Differences in clinical manifestations among *Cryptosporidium* species and subtypes in HIV-infected persons." J Infect Dis. **196**(5): 684-91.

Cama, V., Arrowood, MJ, Ortega, YR, Xiao, L. (2006). "Molecular characterization of the *Cryptosporidium parvum* IOWA isolate kept in different laboratories." J Eukaryot Microbiol. **53**(Suppl 1): S40-2.

Camero, L., Shulaw, W. P., Xiao, L. (2003). "Characterization of a *Cryptosporidium parvum* gene encoding a protein with homology to long chain fatty acid synthetase." J Eukaryot Microbiol. **50 Suppl**: 534-8.

Campbell, A., Robertson, LJ, Smith, HV. (1992). "Viability of *Cryptosporidium parvum* oocysts: correlation of *in vitro* excystation with inclusion or exclusion of fluorogenic vital dyes." Appl Environ Microbiol. **58**(11): 3488-93.

Carey, C., Lee, H, Trevors, JT. (2004). "Biology, persistence and detection of *Cryptosporidium parvum* and *Cryptosporidium hominis* oocyst." Water Res. **38**(4): 818-62.

Carraway, M., Tzipori, S, Widmer, G. (1997). "A new restriction fragment length polymorphism from *Cryptosporidium parvum* identifies genetically heterogeneous parasite populations and genotypic changes following

transmission from bovine to human hosts." Infect Immun. **65**(9): 3958-60.

Carraway, M., Tzipori, S, Widmer, G. (1996). "Identification of genetic heterogeneity in the *Cryptosporidium parvum* ribosomal repeat." Appl Environ Microbiol. **62**(2): 712-6.

Carraway, M., Widmer, G., Tzipori, S. (1994). "Genetic markers differentiate *C. parvum* isolates." J Eukaryot Microbiol. **41**(5): 26S.

Carreno, R., Martin, DS, Barta, JR. (1999). "*Cryptosporidium* is more closely related to the gregarines than to coccidia as shown by phylogenetic analysis of apicomplexan parasites inferred using small-subunit ribosomal RNA gene sequences." Parasitol Res. **85**(11): 899-904.

Casadevall, A., Pirofski, L. (2001). "Host-pathogen interactions: the attributes of virulence." J Infect Dis. **184**(3): 337-44.

Casadevall, A., Pirofski, L. A. (1999). "Host-pathogen interactions: redefining the basic concepts of virulence and pathogenicity." Infect Immun. **67**(8): 3703-13.

Casemore, D., Sands, RL, Curry, A. (1985). "*Cryptosporidium* species a "new" human pathogen." J Clin Pathol. **38**(12): 1321-36.

Cevallos, A., Zhang, X, Waldor, MK, Jaison, S, Zhou, X, Tzipori, S, Neutra, MR, Ward, HD. (2000). "Molecular cloning and expression of a gene encoding *Cryptosporidium parvum* glycoproteins gp40 and gp15." Infect Immun. **68**(7): 4108-16. .

Cevallos, A. M., Bhat, N., Verdon, R., Hamer, D. H., Stein, B., Tzipori, S., Pereira, M. E., Keusch, G. T., Ward, H. D. (2000 b). "Mediation of *Cryptosporidium parvum* infection *in vitro* by mucin-like glycoproteins defined by a neutralizing monoclonal antibody." Infect Immun. **68**(9): 5167-75.

Chalmers, R. M., Davies, A. P. (2010). "Minireview: clinical cryptosporidiosis." Exp Parasitol. **124**(1): 138-46.

Chalmers, R. M., Elwin, K., Thomas, A. L., Guy, E. C., Mason, B. (2009 c). "Long-term *Cryptosporidium* typing reveals the aetiology and species-specific epidemiology of human cryptosporidiosis in England and Wales, 2000 to 2003." Euro Surveill. **14**(2).

Chalmers, R. M., Robinson, G., Wright, S., Hunter, P.R., Elwin, E., Bouzid, M., Innes, E.A., Hadfield, S., Tyler, M., Katzer, F. (2009 b). Comparative biology of *Cryptosporidium hominis* and the *Cryptosporidium* rabbit genotype. III International *Giardia* and *Cryptosporidium* Conference. 11-15 October 2009, Orvieto, Italy.

Chalmers, R., Robinson, G, Elwin, K, Hadfield, SJ, Xiao, L, Ryan, U, Modha, D, Mallaghan, C. (2009). "*Cryptosporidium* sp. rabbit genotype, a newly identified human pathogen." Emerg Infect Dis. **15**(5): 829-30.

Chalmers, R. M. (2008 b). "*Cryptosporidium*: from laboratory diagnosis to surveillance and outbreaks." Parasite. **15**(3): 372-8.

Chalmers, R., Hadfield, SJ, Jackson, CJ, Elwin, K, Xiao, L, Hunter P. (2008). "Geographic linkage and variation in *Cryptosporidium hominis*." Emerg Infect Dis. **14**(3): 496-8.

Chalmers, R., Elwin, K, Thomas, AL, Joynson, DH. (2002). "Infection with unusual types of *Cryptosporidium* is not restricted to immunocompromised patients." J Infect Dis. **185**(2): 270-1.

Chappell, C., Okhuysen, PC. (2002). "Cryptosporidiosis." Curr Opin Infect Dis. **15**(5): 523-7.

Char, S., Kelly, P., Naeem, A., Farthing, M. J. (1996). "Codon usage in

Cryptosporidium parvum differs from that in other Eimeriorina." Parasitology. **112 (Pt 4)**: 357-62.

Chen, X. M., O'Hara, S. P., Huang, B. Q., Nelson, J. B., Lin, J. J., Zhu, G., Ward, H. D., LaRusso, N. F. (2004). "Apical organelle discharge by *Cryptosporidium parvum* is temperature, cytoskeleton, and intracellular calcium dependent and required for host cell invasion." Infect Immun. **72**(12): 6806-16.

Chen, X., Keithly, JS, Paya, CV, LaRusso, NF. (2002). "Cryptosporidiosis." N Engl J Med. **346**(22): 1723-31.

Cheung, V., Nelson, SF. (1996). "Whole genome amplification using a degenerate oligonucleotide primer allows hundreds of genotypes to be performed on less than one nanogram of genomic DNA." Proc Natl Acad Sci U S A. **93**(25): 14676-9.

Choudhry, N., Bajaj-Elliott, M, McDonald, V. (2008). "The terminal sialic acid of glycoconjugates on the surface of intestinal epithelial cells activates excystation of *Cryptosporidium parvum*." Infect Immun. **76**(8): 3735-41.

Clark, D. (1999). "New insights into human cryptosporidiosis." Clin Microbiol Rev. **12**(4): 554-63.

Clarke, J. J. (1895). "A study of coccidia met with in mice." J. Microsc. Soc. **37**: 277-302.

Cohen, S. (2005). "Use of nitazoxanide as a new therapeutic option for persistent diarrhea: a pediatric perspective." Curr Med Res Opin. **21**(7): 999-1004.

Colford, J. M., Jr., Saha, S. R., Wade, T. J., Wright, C. C., Vu, M., Charles, S., Jensen, P., Hubbard, A., Levy, D. A., Eisenberg, J. N. (2005). "A pilot randomized, controlled trial of an in-home drinking water intervention among HIV + persons." J Water Health. **3**(2): 173-84.

Collins, F. S., Green, E. D., Guttmacher, A. E., Guyer, M. S. (2003). "A vision for the future of genomics research." Nature. **422**(6934): 835-47.

Coombs, G. (1999). "Biochemical peculiarities and drug targets in *Cryptosporidium parvum*: lessons from other coccidian parasites." Parasitol Today. **15**(8): 333-8. .

Cordell, R. L., Addiss, D. G. (1994). "Cryptosporidiosis in child care settings: a review of the literature and recommendations for prevention and control." Pediatr Infect Dis J. **13**(4): 310-7.

Corso, P., Kramer, MH, Blair, KA, Addiss, DG, Davis, JP, Haddix, AC. (2003). "Cost of illness in the 1993 waterborne *Cryptosporidium* outbreak, Milwaukee, Wisconsin." Emerg Infect Dis. **9**(4): 426-31.

Coutinho, B. P., R. B. Oria, *et al.* (2008). "*Cryptosporidium* infection causes undernutrition and, conversely, weanling undernutrition intensifies infection." J Parasitol. **94**(6): 1225-32.

Crawford, F., Vermund, SH. (1988). "Human cryptosporidiosis." Crit Rev Microbiol. **16**(2): 113-59.

Current, W., Garcia, LS. (1991). "Cryptosporidiosis." Clin Microbiol Rev. **4**(3): 325-58.

Current, W., Snyder, DB. (1988). "Development of and serologic evaluation of acquired immunity to *Cryptosporidium baileyi* by broiler chickens." Poult Sci. **67**(5): 720-9.

Current, W., Reese, NC. (1986). "A comparison of endogenous development of three isolates of *Cryptosporidium* in suckling mice." J Protozool. **33**(1): 98-108.

Current, W., Upton, SJ, Haynes, TB. (1986). "The life cycle of *Cryptosporidium baileyi* n. sp. (Apicomplexa, Cryptosporidiidae) infecting chickens." J

Protozool. **33**(2): 289-96.

Current, W., Reese, NC, Ernst, JV, Bailey, WS, Heyman, MB, Weinstein, WM. (1983). "Human cryptosporidiosis in immunocompetent and immunodeficient persons. Studies of an outbreak and experimental transmission." N Engl J Med. **308**(21): 1252-7.

D'Antonio, R., Winn, RE, Taylor, JP, Gustafson, TL, Current, WL, Rhodes, MM, Gary, GW Jr, Zajac, RA. (1985). "A waterborne outbreak of cryptosporidiosis in normal hosts." Ann Intern Med. **103**(6 (Pt 1)): 886-8.

Davies, A. P., Campbell, B., Evans, M. R., Bone, A., Roche, A., Chalmers, R. M. (2009). "Asymptomatic carriage of protozoan parasites in children in day care centers in the United kingdom." Pediatr Infect Dis J. **28**(9): 838-40.

De Graaf, D., Vanopdenbosch, E, Ortega-Mora, LM, Abbassi, H, Peeters, JE. (1999). "A review of the importance of cryptosporidiosis in farm animals." Int J Parasitol. **29**(8): 1269-87.

Dean, F., Hosono, S, Fang, L, Wu, X, Faruqi, AF, Bray-Ward, P, Sun, Z, Zong, Q, Du, Y, Du, J, Driscoll, M, Song, W, Kingsmore, SF, Egholm, M, Lasken, RS. (2002). "Comprehensive human genome amplification using multiple displacement amplification." Proc Natl Acad Sci U S A. **99**(8): 5261-6.

Deitsch, K. W., Moxon, E. R., Wellems, T. E. (1997). "Shared themes of antigenic variation and virulence in bacterial, protozoal, and fungal infections." Microbiol Mol Biol Rev. **61**(3): 281-93.

Deng, M., Cliver, DO. (1998). "*Cryptosporidium parvum* development in the BS-C-1 cell line." J Parasitol. **84**(1): 8-15.

Deng, M., Cliver, DO, Mariam, TW. (1997). "Immunomagnetic capture PCR to detect viable *Cryptosporidium parvum* oocysts from environmental samples." Appl Environ Microbiol. **63**(8): 3134-8.

Detter, J., Jett, JM, Lucas, SM, Dalin, E, Arellano, AR, Wang, M, Nelson, JR, Chapman, J, Lou, Y, Rokhsar, D, Hawkins, TL, Richardson, PM. (2002). "Isothermal strand-displacement amplification applications for high-throughput genomics." Genomics. **80**(6): 691-8.

Di Giovanni, G., Hashemi, FH, Shaw, NJ, Abrams, FA, LeChevallier, MW, Abbaszadegan, M. (1999). "Detection of infectious *Cryptosporidium parvum* oocysts in surface and filter backwash water samples by immunomagnetic separation and integrated cell culture-PCR." Appl Environ Microbiol. **65**(8): 3427-32.

DuPont, H., Chappell, CL, Sterling, CR, Okhuysen, PC, Rose, JB, Jakubowski, W. (1995). "The infectivity of *Cryptosporidium parvum* in healthy volunteers." N Engl J Med. **332**(13): 855-9.

Dwivedi, K. K., Prasad, G., Saini, S., Mahajan, S., Lal, S., Baveja, U. K. (2007). "Enteric opportunistic parasites among HIV infected individuals: associated risk factors and immune status." Jpn J Infect Dis. **60**(2-3): 76-81.

Ebert, D., Herre, EA. (1996). "The evolution of parasitic diseases." Parasitol Today. **12**(3): 96-101.

Ebert, D. (1994). "Virulence and Local Adaptation of a Horizontally Transmitted Parasite." Science. **265**(5175): 1084-1086.

Eckert, K., Kunkel, TA. (1991). "DNA polymerase fidelity and the polymerase chain reaction." PCR Methods Appl. **1**(1): 17-24.

Edberg, S. C. (2009). "Does the possession of virulence factor genes mean that those genes will be active?" J Water Health. **7 Suppl 1**: S19-28.

Egger, M., Mäusezahl, D, Odermatt, P, Marti, HP, Tanner, M. (1990). "Symptoms

and transmission of intestinal cryptosporidiosis." Arch Dis Child. **65**(4): 445-7.

Egyed, Z., Sréter, T, Széll, Z, Varga, I. (2003). "Characterization of *Cryptosporidium* spp.--recent developments and future needs." Vet Parasitol. **111**(2-3): 103-14.

Elliot, B., Wisnewski, AV, Johnson, J, Fenwick-Smith, D, Wiest, P, Hamer, D, Kresina, T, Flanigan, TP. (1997). "*In vitro* inhibition of *Cryptosporidium parvum* infection by human monoclonal antibodies." Infect Immun. **65**(9): 3933-5.

Elwin, K., Chalmers, RM., Roberts, R., Guy, EC., Casemore, DP. (2001). "Modification of a rapid method for the identification of gene-specific polymorphisms in *Cryptosporidium parvum* and its application to clinical and epidemiological investigations." Appl Environ Microbiol. **67**(12): 5581-4.

Enemark, H. L., Ahrens, P., Juel, C. D., Petersen, E., Petersen, R. F., Andersen, J. S., Lind, P., Thamsborg, S. M. (2002). "Molecular characterization of Danish *Cryptosporidium parvum* isolates." Parasitology. **125**(Pt 4): 331-41.

Engelstadter, J., Bonhoeffer, S. (2009). "Red Queen dynamics with non-standard fitness interactions." PLoS Comput Biol. **5**(8): e1000469.

Esteban, J., Salas, M, Blanco, L. (1993). "Fidelity of phi 29 DNA polymerase. Comparison between protein-primed initiation and DNA polymerization." J Biol Chem. **268**(4): 2719-26.

Farazi, T. A., Waksman, G., Gordon, J. I. (2001). "The biology and enzymology of protein N-myristoylation." J Biol Chem. **276**(43): 39501-4.

Fayer, R., Orlandi, P., Perdue, M. L. (2009). "Virulence factor activity relationships for hepatitis E and *Cryptosporidium*." J Water Health. **7 Suppl 1**: S55-63.

Fayer, R., Santín, M. (2009). "*Cryptosporidium xiaoi* n. sp. (Apicomplexa: Cryptosporidiidae) in sheep (Ovis aries)." Vet Parasitol. **164**(2-4): 192-200.

Fayer, R., Santín, M, Trout, JM. (2008). "*Cryptosporidium ryanae* n. sp. (Apicomplexa: Cryptosporidiidae) in cattle (*Bos taurus*)." Vet Parasitol. **156**(3-4): 191-8.

Fayer, R., Santín, M, Xiao, L. (2005). "*Cryptosporidium bovis* n. sp. (Apicomplexa: Cryptosporidiidae) in cattle (*Bos taurus*)." J Parasitol. **91**(3): 624-9.

Fayer, R. (2004). "*Cryptosporidium*: a water-borne zoonotic parasite." Vet Parasitol. **126**(1-2): 37-56.

Fayer, R., Trout, JM, Xiao, L, Morgan, UM, Lai, AA, Dubey, JP. (2001). "*Cryptosporidium canis* n. sp. from domestic dogs." J Parasitol. **87**(6): 1415-22.

Fayer, R., Morgan, U, Upton, SJ. (2000). "Epidemiology of *Cryptosporidium*: transmission, detection and identification." Int J Parasitol. **30**(12-13): 1305-22.

Fayer, R., Trout, JM, Jenkins, MC. (1998). "Infectivity of *Cryptosporidium parvum* oocysts stored in water at environmental temperatures." J Parasitol. **84**(6): 1165-9.

Fayer, R. (1997). *Cryptosporidium* And Cryptosporidiosis, CRC Press, Boca Raton.

Fayer, R., Ungar, BL. (1986). "*Cryptosporidium* spp. and cryptosporidiosis." Microbiol Rev. **50**(4): 458-83.

Fayer, R., Leek, RG. (1984). "The effects of reducing conditions, medium, pH, temperature, and time on *in vitro* excystation of *Cryptosporidium*." J Protozool. **31**(4): 567-9.

Feltus, D., Giddings, CW, Schneck, BL, Monson, T, Warshauer, D, McEvoy, JM. (2006). "Evidence supporting zoonotic transmission of *Cryptosporidium* spp.

in Wisconsin." J Clin Microbiol. **44**(12): 4303-8.

Feng, H., Nie, W, Sheoran, A, Zhang, Q, Tzipori, S. (2006). "Bile acids enhance invasiveness of *Cryptosporidium* spp. into cultured cells." Infect Immun. **74**(6): 3342-6.

Feng, X., Rich, S. M., Akiyoshi, D., Tumwine, J. K., Kekitiinwa, A., Nabukeera, N., Tzipori, S., Widmer, G. (2000). "Extensive polymorphism in *Cryptosporidium parvum* identified by multilocus microsatellite analysis." Appl Environ Microbiol. **66**(8): 3344-9.

Feng, X., Rich, SM, Tzipori, S, Widmer, G. (2002). "Experimental evidence for genetic recombination in the opportunistic pathogen *Cryptosporidium parvum*." Mol Biochem Parasitol. **119**(1): 55-62.

Fiegler, H., Geigl, JB, Langer, S, Rigler, D, Porter, K, Unger, K, Carter, NP, Speicher, MR. (2007). "High resolution array-CGH analysis of single cells." Nucleic Acids Research. **35**(3): e15.

Fischer, P., Taraschewski, H, Ringelmann, R, Eing, B. (1998). "Detection of *Cryptosporidium parvum* in human feces by PCR." Tokai J Exp Clin Med. **23**(6): 309-11.

Fischer, O. (1982). "Cryptosporidiosis in calves during the period of milk nutrition." Vet Med (Praha). **27**(8): 465-71.

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A., Merrick, J. M., *et al.* (1995). "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd." Science. **269**(5223): 496-512.

Fontaine, M., Guillot, E. (2002). "Development of a TaqMan quantitative PCR assay specific for *Cryptosporidium parvum*." FEMS Microbiol Lett. **214**(1): 13-7.

Forgacs, P., Tarshis, A, Ma, P, Federman, M, Mele, L, Silverman, ML, Shea, JA. (1983). "Intestinal and bronchial cryptosporidiosis in an immunodeficient homosexual man." Ann Intern Med. **99**(6): 793-4.

Forney, J. R., Yang, S., Healey, M. C. (1996 a). "Protease activity associated with excystation of *Cryptosporidium parvum* oocysts." J Parasitol. **82**(6): 889-92.

Forney, J. R., Yang, S., Du, C., Healey, M. C. (1996 b). "Efficacy of serine protease inhibitors against *Cryptosporidium parvum* infection in a bovine fallopian tube epithelial cell culture system." J Parasitol. **82**(4): 638-40.

Fraser, C. M., Eisen, J. A., Salzberg, S. L. (2000). "Microbial genome sequencing." Nature. **406**(6797): 799-803.

Frost, F. J., Roberts, M., Kunde, T. R., Craun, G., Tollestrup, K., Harter, L., Muller, T. (2005). "How clean must our drinking water be: the importance of protective immunity." J Infect Dis. **191**(5): 809-14.

Garcia, L., Shimizu, RY. (1997). "Evaluation of nine immunoassay kits (enzyme immunoassay and direct fluorescence) for detection of *Giardia lamblia* and *Cryptosporidium parvum* in human fecal specimens." J Clin Microbiol. **35**(6): 1526-9.

Garcia, L., Shum, AC, Bruckner, DA. (1992). "Evaluation of a new monoclonal antibody combination reagent for direct fluorescence detection of *Giardia* cysts and *Cryptosporidium* oocysts in human fecal specimens." J Clin Microbiol. **30**(12): 3255-7.

Garcia, L., Brewer, TC, Bruckner, DA. (1987). "Fluorescence detection of *Cryptosporidium* oocysts in human fecal specimens by using monoclonal antibodies." J Clin Microbiol. **25**(1): 119-21.

Garcia, L., Bruckner, DA, Brewer, TC, Shimizu, RY. (1983). "Techniques for the

recovery and identification of *Cryptosporidium* oocysts from stool specimens." J Clin Microbiol. **18**(1): 185-90.

Gargala, G. (2008). "Drug treatment and novel drug target against *Cryptosporidium*." Parasite. **15**(3): 275-81.

Garnick, E. (1992). "Parasite virulence and parasite-host coevolution: a reappraisal " J. Parasitol. **78**(2): 381-386.

Geurden, T., Levecke, B., Caccio, S. M., Visser, A., De Groote, G., Casaert, S., Vercruysse, J., Claerebout, E. (2009). "Multilocus genotyping of *Cryptosporidium* and *Giardia* in non-outbreak related cases of diarrhoea in human patients in Belgium." Parasitology. **136**(10): 1161-8.

Gibbons-Matthews, C., Prescott, A. M. (2003). "Intra-isolate variation of *Cryptosporidium parvum* small subunit ribosomal RNA genes from human hosts in England." Parasitol Res. **90**(6): 439-44.

Gibbons, C. L., Gazzard, B G, Ibrahim, M, Morris-Jones, S, Ong, C S L, Awad-El-Kariem, F M. (1998). "Correlation between markers of strain variation in *Cryptosporidium parvum*: evidence of clonality." Parasitol Int. **47**: 139-147.

Girouard, D., Gallant, J., Akiyoshi, D. E., Nunnari, J., Tzipori, S. (2006). "Failure to propagate *Cryptosporidium* spp. in cell-free culture." J Parasitol. **92**(2): 399-400.

Glaberman, S., Moore, JE, Lowery, CJ, Chalmers, RM, Sulaiman, I, Elwin, K, Rooney, PJ, Millar, BC, Dooley, JS, Lal, AA, Xiao, L. (2002). "Three drinking-water-associated cryptosporidiosis outbreaks, Northern Ireland." Emerg Infect Dis. **8**(6): 631-3.

Glaberman, S., Sulaiman, I. M., Bern, C., Limor, J., Peng, M. M., Morgan, U., Gilman, R., Lal, A. A., Xiao, L. (2001). "A multilocus genotypic analysis of *Cryptosporidium meleagridis*." J Eukaryot Microbiol. **Suppl**: 19S-22S.

Göbel, E., Brändler, U. (1982). "Ultrastructure of microgametogenesis, microgametes and gametogony of *Cryptosporidium* sp. in the small intestine of mice." Protistologica. **18**: 331-344.

Gobet, P., Toze, S. (2001). "Sensitive genotyping of *Cryptosporidium parvum* by PCR-RFLP analysis of the 70-kilodalton heat shock protein (HSP70) gene." FEMS Microbiol Lett. **200**(1): 37-41.

Goh, S., Reacher, M, Casemore, DP, Verlander, NQ, Chalmers, R, Knowles, M, Williams, J, Osborn, K, Richards, S. (2004). "Sporadic cryptosporidiosis, North Cumbria, England, 1996-2000." Emerg Infect Dis. **10**(6): 1007-15.

Goodgame, R., Genta, RM, White, AC, Chappell, CL. (1993). "Intensity of infection in AIDS-associated cryptosporidiosis." J Infect Dis. **167**(3): 704-9.

Goodwin, M. (1989). "Cryptosporidiosis in birds--a review." Avian Pathol. **18**(3): 365-84.

Gordon, J., Sibley, LD. (2005). " Comparative genome analysis reveals a conserved family of actin-like proteins in apicomplexan parasites.." BMC Genomics. **6**: 179.

Goumenou, M., Machera, K. (2004). "Measurement of DNA single-strand breaks by alkaline elution and fluorometric DNA quantification." Anal Biochem. **326**(2): 146-52.

Grocock, R. J., Sharp, P. M. (2001). "Synonymous codon usage in *Cryptosporidium parvum*: identification of two distinct trends among genes." Int J Parasitol. **31**(4): 402-12.

Guerrant, R. (1997). "Cryptosporidiosis: an emerging, highly infectious threat." Emerg Infect Dis. **3**(1): 51-7.

Gut, J., Nelson, RG. (1994). "*Cryptosporidium parvum* sporozoites deposit trails of 11A5 antigen during gliding locomotion and shed 11A5 antigen during invasion of MDCK cells *in vitro*." J Eukaryot Microbiol. **41**(5): 42S-43S.

Gut, J., Petersen, C, Nelson R, Leech, J. (1991). "*Cryptosporidium parvum*: *in vitro* cultivation in Madin-Darby canine kidney cells." J Protozool. **38**(6): 72S-73S.

Guttmacher, A. E., Collins, F. S. (2003). "Welcome to the genomic era." N Engl J Med. **349**(10): 996-8.

Guy, R., Payment, P, Krull, UJ, Horgen, PA. (2003). "Real-time PCR for quantification of *Giardia* and *Cryptosporidium* in environmental water samples and sewage." Appl Environ Microbiol. **69**(9): 5178-85. .

Guyot, K., Follet-Dumoulin, A, Recourt, C, Lelièvre, E, Cailliez, JC, Dei-Cas, E. (2002). "PCR-restriction fragment length polymorphism analysis of a diagnostic 452-base-pair DNA fragment discriminates between *Cryptosporidium parvum* and *C. meleagridis* and between *C. parvum* isolates of human and animal origin." Appl Environ Microbiol. **68**(4): 2071-6.

Han, S., Zschausch, HC, Meyer, HG, Schneider, T, Loos, M, Bhakdi, S, Maeurer, MJ. (2000). "*Helicobacter pylori*: clonal population structure and restricted transmission within families revealed by molecular typing." J Clin Microbiol. **38**(10): 3646-51.

Harari, M., West, B, Dwyer, B. (1986). "*Cryptosporidium* as cause of laryngotracheitis in an infant." Lancet. **1**(8491): 1207.

Hawkins, T., Detter, JC, Richardson, PM. (2002). "Whole genome amplification--applications and advances." Curr Opin Biotechnol. **13**(1): 65-7.

Hayes, E., Matte, TD, O'Brien, TR, McKinley, TW, Logsdon, GS, Rose, JB, Ungar, BL, Word, DM, Pinsky, PF, Cummings, ML, *et al.* (1989). "Large community outbreak of cryptosporidiosis due to contamination of a filtered public water supply." N Engl J Med. **320**(21): 1372-6.

Hays, M., Mosier, DA, Oberst, RD. (1995). "Enhanced karyotype resolution of *Cryptosporidium parvum* by contour-clamped homogeneous electric fields." Vet Parasitol. **58**(4): 273-80.

Hellard, M., Hocking, J., Willis, J., Dore, G., Fairley, C. (2003). "Risk factors leading to *Cryptosporidium* infection in men who have sex with men." Sex Transm Infect. **79**(5): 412-4.

Henderson, I. R., Owen, P.,Nataro, J. P. (1999). "Molecular switches--the ON and OFF of bacterial phase variation." Mol Microbiol. **33**(5): 919-32.

Henriksen, S. A., Pohlenz, J. F. (1981). "Staining of cryptosporidia by a modified Ziehl-Neelsen technique." Acta Vet Scand. **22**(3-4): 594-6.

Hewitt, R., Yiannoutsos, CT, Higgs, ES, Carey, JT, Geiseler, PJ, Soave, R, Rosenberg, R, Vazquez, GJ, Wheat, LJ, Fass, RJ, Antoninievic, Z, Walawander, AL, Flanigan, TP, Bender, JF. (2000). "Paromomycin: no more effective than placebo for treatment of cryptosporidiosis in patients with advanced human immunodeficiency virus infection. AIDS Clinical Trial Group." Clin Infect Dis. **31**(4): 1084-92.

Higgins, J., Fayer, R, Trout, JM, Xiao, L, Lal, AA, Kerby, S, Jenkins, MC. (2001). "Real-time PCR for the detection of *Cryptosporidium parvum*." J Microbiol Methods. **47**(3): 323-37.

Hijjawi, N. (2010). "*Cryptosporidium*: new developments in cell culture." Exp Parasitol. **124**(1): 54-60.

Hijjawi, N., Meloni, BP, Ng'anzo, M, Ryan, UM, Olson, ME, Cox, PT, Monis, PT, Thompson, RC. (2004). "Complete development of *Cryptosporidium parvum*

in host cell-free culture." Int J Parasitol. **34**(7): 769-77.

Hijjawi, N., Meloni, BP, Ryan, UM, Olson, ME, Thompson, RC. (2002). "Successful *in vitro* cultivation of *Cryptosporidium andersoni*: evidence for the existence of novel extracellular stages in the life cycle and implications for the classification of *Cryptosporidium*." Int J Parasitol. **32**(14): 1719-26.

Hijjawi, N., Meloni, BP, Morgan, UM, Thompson, RC. (2001). "Complete development and long-term maintenance of *Cryptosporidium parvum* human and cattle genotypes in cell culture." Int J Parasitol. **31**(10): 1048-55.

Holland, R. (1990). "Some infectious causes of diarrhea in young farm animals." Clin Microbiol Rev. **3**(4): 345-75.

Homan, W., van Gorkom, T, Kan, YY, Hepener, J. (1999). "Characterization of *Cryptosporidium parvum* in human and animal feces by single-tube nested polymerase chain reaction and restriction analysis." Parasitol Res. **85**(8-9): 707-12.

Horman, A., Korpela, H., Sutinen, J., Wedel, H., Hanninen, M. L. (2004). "Meta-analysis in assessment of the prevalence and annual incidence of *Giardia* spp. and *Cryptosporidium* spp. infections in humans in the Nordic countries." Int J Parasitol. **34**(12): 1337-46.

Huang, X., Madan, A. (1999). "CAP3: A DNA sequence assembly program." Genome Res. **9**(9): 868-77.

Huetink, R. E., van der Giessen, J. W., Noordhuizen, J. P., Ploeger, H. W. (2001). "Epidemiology of *Cryptosporidium* spp. and *Giardia duodenalis* on a dairy farm." Vet Parasitol. **102**(1-2): 53-67.

Hunt, R., Sauna, Z. E., Ambudkar, S. V., Gottesman, M. M., Kimchi-Sarfaty, C. (2009). "Silent (synonymous) SNPs: should we care about them?" Methods Mol Biol. **578**: 23-39.

Hunter, P. R., Wilkinson, D. C., Lake, I. R., Harrison, F. C., Syed, Q., Hadfield, S. J., Chalmers, R. M. (2008). "Microsatellite typing of *Cryptosporidium parvum* in isolates from a waterborne outbreak." J Clin Microbiol. **46**(11): 3866-7.

Hunter, P., Hadfield, SJ, Wilkinson, D, Lake, IR, Harrison, FC, Chalmers, RM. (2007). "Subtypes of *Cryptosporidium parvum* in humans and disease risk." Emerg Infect Dis. **13**(1): 82-8.

Hunter, P., Thompson, RC. (2005). "The zoonotic transmission of *Giardia* and *Cryptosporidium*." Int J Parasitol. **35**(11-12): 1181-90.

Hunter, P., Hughes, S, Woodhouse, S, Syed, Q, Verlander, NQ, Chalmers, RM, Morgan, K, Nichols, G, Beeching, N, Osborn, K. (2004 b). "Sporadic cryptosporidiosis case-control study with genotyping." Emerg Infect Dis. **10**(7): 1241-9.

Hunter, P., Hughes, S, Woodhouse, S, Raj, N, Syed, Q, Chalmers, RM, Verlander, NQ, Goodacre, J. (2004). "Health sequelae of human cryptosporidiosis in immunocompetent patients." Clin Infect Dis. **39**(4): 504-10.

Hunter, P., Chalmers, RM, Syed, Q, Hughes, LS, Woodhouse, S, Swift, L. (2003). "Foot and mouth disease and cryptosporidiosis: possible interaction between two emerging infectious diseases." Emerg Infect Dis. **9**(1): 109-12.

Hunter, P., Nichols, G. (2002). "Epidemiology and clinical features of *Cryptosporidium* infection in immunocompromised patients." Clin Microbiol Rev. **15**(1): 145-54.

Hunter, P. R., Syed, Q. (2001). "Community surveys of self-reported diarrhoea can dramatically overestimate the size of outbreaks of waterborne cryptosporidiosis." Water Sci Technol. **43**(12): 27-30.

Hunter, P. R., Fraser, C. A. (1990). "Application of the theory of adaptive polymorphism to the ecology and epidemiology of pathogenic yeasts." Appl Environ Microbiol. **56**(7): 2219-22.

Ignatius, R., Eisenblatter, M., Regnath, T., Mansmann, U., Futh, U., Hahn, H., Wagner, J. (1997). "Efficacy of different methods for detection of low *Cryptosporidium parvum* oocyst numbers or antigen concentrations in stool specimens." Eur J Clin Microbiol Infect Dis. **16**(10): 732-6.

Iseki, M. (1979). "*Cryptosporidium felis* sp.n. (Protozoa: Eimeriorina) from the domestic cat." Jpn. J. Parasitol. **28**: 285-307.

Jenkins, M., Trout, J, Abrahamsen, MS, Lancto, CA, Higgins, J, Fayer, R. (2000). "Estimating viability of *Cryptosporidium parvum* oocysts using reverse transcriptase-polymerase chain reaction (RT-PCR) directed at mRNA encoding amyloglucosidase." J Microbiol Methods. **43**(2): 97-106.

Jex, A., Gasser, RB. (2010). "Genetic richness and diversity in *Cryptosporidium hominis* and *C. parvum* reveals major knowledge gaps and a need for the application of "next generation" technologies - Research review." Biotechnol Adv. **28**: 17-26.

Jex, A. R., Smith, H. V., Monis, P. T., Campbell, B. E., Gasser, R. B. (2008). "*Cryptosporidium*--biotechnological advances in the detection, diagnosis and analysis of genetic variation." Biotechnol Adv. **26**(4): 304-17.

Jiang, J., Alderisio, KA, Xiao, L. (2005). "Distribution of *Cryptosporidium* genotypes in storm event water samples from three watersheds in New York." Appl Environ Microbiol. **71**(8): 4446-54.

Jiang, J., Xiao, L. (2003). "An evaluation of molecular diagnostic tools for the detection and differentiation of human-pathogenic *Cryptosporidium* spp." J Eukaryot Microbiol. **50**(Suppl): 542-7.

Jokipii, L., Jokipii, AM. (1986). "Timing of symptoms and oocyst excretion in human cryptosporidiosis." N Engl J Med. **315**(26): 1643-7.

Jokipii, L., Pohjola, S, Jokipii, AM. (1983). "*Cryptosporidium*: a frequent finding in patients with gastrointestinal symptoms." Lancet. **2**(8346): 358-61.

Juranek, D. (1995). "Cryptosporidiosis: sources of infection and guidelines for prevention." Clin Infect Dis. **22**(1): S57-61.

Karanis, P., Kimura, A, Nagasawa, H, Igarashi, I, Suzuki, N. (2008). "Observations on *Cryptosporidium* life cycle stages during excystation." J Parasitol. **94**(1): 298-300.

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): 1-38.

Keithly, J., Langreth, SG, Buttle, KF, Mannella, CA. (2005). "Electron tomographic and ultrastructural analysis of the *Cryptosporidium parvum* relict mitochondrion, its associated membranes, and organelles." J Eukaryot Microbiol. **52**(2): 132-40.

Keithly, J., Zhu, G, Upton, SJ, Woods, KM, Martinez, MP, Yarlett, N. (1997). "Polyamine biosynthesis in *Cryptosporidium parvum* and its implications for chemotherapy." Mol Biochem Parasitol. **88**(1-2): 35-42.

Khramtsov, N., Woods, KM, Nesterenko, MV, Dykstra, CC, Upton, SJ. (1997). "Virus-like, double-stranded RNAs in the parasitic protozoan *Cryptosporidium parvum*." Mol Microbiol. **26**(2): 289-300.

Khramtsov, N. V., Tilley, M., Blunt, D. S., Montelone, B. A., Upton, S. J. (1995). "Cloning and analysis of a *Cryptosporidium parvum* gene encoding a protein

with homology to cytoplasmic form Hsp70." J Eukaryot Microbiol. **42**(4): 416-22.

Kiang, K., Scheftel, JM, Leano, FT, Taylor, CM, Belle-Isle, PA, Cebelinski, EA, Danila, R, Smith, KE. (2006). "Recurrent outbreaks of cryptosporidiosis associated with calves among students at an educational farm programme, Minnesota, 2003." Epidemiol Infect. **134**(4): 878-86.

Korich, D., Mead, JR, Madore, MS, Sinclair, NA, Sterling, CR. (1990). "Effects of ozone, chlorine dioxide, chlorine, and monochloramine on *Cryptosporidium parvum* oocyst viability." Appl Environ Microbiol. **56**(5): 1423-8.

Kosek, M., Alcantara, C, Lima, AA, Guerrant, RL. (2001). "Cryptosporidiosis: an update." Lancet Infect Dis. **1**(4): 262-9.

Kostrzynska, M., Sankey, M, Haack, E, Power, C, Aldom, JE, Chagla, AH, Unger, S, Palmateer, G, Lee, H, Trevors, JT, De Grandis, SA. (1999). "Three sample preparation protocols for polymerase chain reaction based detection of *Cryptosporidium parvum* in environmental samples." J Microbiol Methods. **35**(1): 65-71.

Koudela, B., Modry, D. (1998). "New species of *Cryptosporidium* (Apicomplexa : Cryptosporidiidae) from lizards." Folia Parasitologica. **45**: 93-100.

Kuo, C., Wares, JP, Kissinger, JC. (2008). "The Apicomplexan whole-genome phylogeny: an analysis of incongruence among gene trees." Mol Biol Evol. **25**(12): 2689-98.

Kyes, S., Horrocks, P., Newbold, C. (2001). "Antigenic variation at the infected red cell surface in malaria." Annu Rev Microbiol. **55**: 673-707.

Laemmli, U., Beguin, F, Gujer-Kellenberger, G. (1970). "A factor preventing the major head protein of bacteriophage T4 from random aggregation." J Mol Biol. **47**(1): 69-85.

LaGier, M., Tachezy, J, Stejskal, F, Kutisova, K, Keithly, JS. (2003). "Mitochondrial-type iron-sulfur cluster biosynthesis genes (IscS and IscU) in the apicomplexan *Cryptosporidium parvum*." Microbiology. **149**(12): 3519-30.

LaGier, M. J., Keithly, J. S., Zhu, G. (2002). "Characterisation of a novel transporter from *Cryptosporidium parvum*." Int J Parasitol. **32**(7): 877-87.

LaGier, M. J., Zhu, G., Keithly, J. S. (2001). "Characterization of a heavy metal ATPase from the apicomplexan *Cryptosporidium parvum*." Gene. **266**(1-2): 25-34.

Langer, R. C., Schaefer, D. A., Riggs, M. W. (2001). "Characterization of an intestinal epithelial cell receptor recognized by the *Cryptosporidium parvum* sporozoite ligand CSL." Infect Immun. **69**(3): 1661-70.

Langer, R. C., Riggs, M. W. (1999). "*Cryptosporidium parvum* apical complex glycoprotein CSL contains a sporozoite ligand for intestinal epithelial cells." Infect Immun. **67**(10): 5282-91.

Langmore, J. (2002). "Rubicon Genomics, Inc." Pharmacogenomics. **3**(4): 557-60.

Le Blancq, S., Khramtsov, NV, Zamani, F, Upton, SJ, Wu, TW. (1997). "Ribosomal RNA gene organization in *Cryptosporidium parvum*." Mol Biochem Parasitol. **90**(2): 463-78.

Leander, B., Harper, JT, Keeling, PJ. (2003). "Molecular phylogeny and surface morphology of marine aseptate gregarines (Apicomplexa): *Selenidium* spp. and *Lecudina* spp." J Parasitol. **89**(6): 1191-205.

Learmonth, J., Ionas, G, Ebbett, KA, Kwan, ES. (2004). "Genetic characterization and transmission cycles of *Cryptosporidium* species isolated from humans in

New Zealand." Appl Environ Microbiol. **70**(7): 3973-8.

Leav, B., Mackay, MR, Anyanwu, A, O' Connor, RM, Cevallos, AM, Kindra, G, Rollins, NC, Bennish, ML, Nelson, RG, Ward, HD. (2002). "Analysis of sequence diversity at the highly polymorphic Cpgp40/15 locus among *Cryptosporidium* isolates from human immunodeficiency virus-infected children in South Africa." Infect Immun. **70**(7): 3881-90.

LeChevallier, M. W., Di Giovanni, G. D., Clancy, J. L., Bukhari, Z., Bukhari, S., Rosen, J. S., Sobrinho, J., Frey, M. M. (2003). "Comparison of method 1623 and cell culture-PCR for detection of *Cryptosporidium* spp. in source waters." Appl Environ Microbiol. **69**(2): 971-9.

Lee, S., Kolodziej, P. A. (2002). "Short Stop provides an essential link between F-actin and microtubules during axon extension." Development. **129**(5): 1195-204.

Lee, J. J., Leedale, G.F. and Bradbury, P. (2000 a). The Illustrated Guide to the Protozoa. Lawrence, Kansas, Wiley-Blackwell.

Lee, S., Harris, K. L., Whitington, P. M., Kolodziej, P. A. (2000 b). "short stop is allelic to kakapo, and encodes rod-like cytoskeletal-associated proteins required for axon extension." J Neurosci. **20**(3): 1096-108.

Leng, X., Mosier, DA, Oberst, RD. (1996). "Differentiation of *Cryptosporidium parvum*, *C. muris*, and *C. baileyi* by PCR-RFLP analysis of the 18S rRNA gene." Vet Parasitol. **62**(1-2): 1-7.

Leoni, F., Amar, C, Nichols, G, Pedraza-Díaz, S, McLauchlin, J. (2006). "Genetic analysis of *Cryptosporidium* from 2414 humans with diarrhoea in England between 1985 and 2000." J Med Microbiol. **55**(6): 703-7.

Leoni, F., Gallimore, CI, Green, J, McLauchlin, J. (2003). "Molecular epidemiological analysis of *Cryptosporidium* isolates from humans and animals by using a heteroduplex mobility assay and nucleic acid sequencing based on a small double-stranded RNA element." J Clin Microbiol. **41**(3): 981-92.

Levine, N. (1980). "Some corrections of coccidian (Apicomplexa: Protozoa) nomenclature." J Parasitol. **66**(5): 830-4.

Limor, J., Lal, AA, Xiao, L. (2002). "Detection and differentiation of *Cryptosporidium* parasites that are pathogenic for humans by real-time PCR." J Clin Microbiol. **40**(7): 2335-8. .

Lindsay, D., Upton, SJ, Owens, DS, Morgan, UM, Mead, JR, Blagburn, BL. (2000). "*Cryptosporidium andersoni* n. sp. (Apicomplexa: Cryptosporiidae) from cattle, *Bos taurus*." J Eukaryot Microbiol. **47**(1): 91-5.

Lively, C. M., Dybdahl, M. F. (2000). "Parasite adaptation to locally common host genotypes." Nature. **405**(6787): 679-81.

Luthra, R., Medeiros, LJ. (2004). "Isothermal multiple displacement amplification: a highly reliable approach for generating unlimited high molecular weight genomic DNA from clinical specimens." J Mol Diagn. **6**(3): 236-42.

Lyons, R. E., Johnson, A. M. (1998). "Gene sequence and transcription differences in 70 kDa heat shock protein correlate with murine virulence of *Toxoplasma gondii*." Int J Parasitol. **28**(7): 1041-51.

Lyons, R. E., Johnson, A. M. (1995). "Heat shock proteins of *Toxoplasma gondii*." Parasite Immunol. **17**(7): 353-9.

Ma, P., Villanueva, T.G., Kaufman, D., Gillooley, J.F. (1984). "Respiratory cryptosporidiosis in the acquired immune deficiency syndrome. Use of modified cold kinyoun and hemacolor stains for rapid diagnoses." J. Am.

Med. Assoc. **252**(10): 1298-1301.

Ma, P., Soave, R. (1983). "Three-step stool examination for cryptosporidiosis in 10 homosexual men with protracted watery diarrhea." J Infect Dis. **147**(5): 824-8.

Mac Kenzie, W., Hoxie, NJ, Proctor, ME, Gradus, MS, Blair, KA, Peterson, DE, Kazmierczak, JJ, Addiss, DG, Fox, KR, Rose, JB, *et al.* (1994). "A massive outbreak in Milwaukee of *Cryptosporidium* infection transmitted through the public water supply." N Engl J Med. **331**(3): 161-7.

MacDonald, L., Sargent, K, Armson, A, Thompson, RC, Reynoldson, JA. (2002). "The development of a real-time quantitative-PCR method for characterisation of a *Cryptosporidium parvum in vitro* culturing system and assessment of drug efficacy." Mol Biochem Parasitol. **121**(2): 279-82.

MacPherson, D., McQueen, R. (1993). "Cryptosporidiosis: multiattribute evaluation of six diagnostic methods." J Clin Microbiol. **31**(2): 198-202.

Mallon, M., MacLeod, A, Wastling, J, Smith, H, Reilly, B, Tait, A. (2003 b). "Population structures and the role of genetic exchange in the zoonotic pathogen *Cryptosporidium parvum*." J Mol Evol. **56**(4): 407-17.

Mallon, M., MacLeod, A, Wastling, JM, Smith, H, Tait, A. (2003). "Multilocus genotyping of *Cryptosporidium parvum* Type 2: population genetics and sub-structuring." Infect Genet Evol. **3**(3): 207-18.

Marks, S., Hanson, TE, Melli, AC. (2004). "Comparison of direct immunofluorescence, modified acid-fast staining, and enzyme immunoassay techniques for detection of *Cryptosporidium* spp in naturally exposed kittens." J Am Vet Med Assoc. **225**(10): 1549-53.

Mayr, E. (1942). Systematics and the origin of species. Columbia University Press. New York, N.Y.

McClelland, E., Bernhardt, P, Casadevall, A. (2006). "Estimating the relative contributions of virulence factors for pathogenic microbes." Infect Immun. **74**(3): 1500-4.

McLauchlin, J., Amar, C, Pedraza-Díaz, S, Nichols, GL. (2000). "Molecular epidemiological analysis of *Cryptosporidium* spp. in the United Kingdom: results of genotyping *Cryptosporidium* spp. in 1,705 fecal samples from humans and 105 fecal samples from livestock animals." J Clin Microbiol. **38**(11): 3984-90.

Mead, J. R., Arrowood, M. J., Sterling, C. R. (1988). "Antigens of *Cryptosporidium* sporozoites recognized by immune sera of infected animals and humans." J Parasitol. **74**(1): 135-43.

Medini, D., Serruto, D., Parkhill, J., Relman, D. A., Donati, C., Moxon, R., Falkow, S., Rappuoli, R. (2008). "Microbiology in the post-genomic era." Nat Rev Microbiol. **6**(6): 419-30.

Meinhardt, P., Casemore, DP, Miller, KB. (1996). "Epidemiologic aspects of human cryptosporidiosis and the role of waterborne transmission." Epidemiol Rev. **18**(2): 118-36.

Meisel, J., Perera, DR, Meligro, C, Rubin, CE. (1976). "Overwhelming watery diarrhea associated with a *Cryptosporidium* in an immunosuppressed patient." Gastroenterology. **70**(6): 1156-60.

Mele, R., Gomez Morales, MA, Tosini, F, Pozio, E. (2003). "Indinavir reduces *Cryptosporidium parvum* infection in both *in vitro* and *in vivo* models." Int J Parasitol. **33**(7): 757-64.

Mellon, R. (1926). "Studies in microbic heredity VIII. The infectivity and virulence

of a filtrable phase in the life history of B. fusiformis and related organisms." J Bacteriol. **12**(4): 279-98.

Meloni, B., Thompson, RC. (1996). "Simplified methods for obtaining purified oocysts from mice and for growing *Cryptosporidium parvum in vitro*." J Parasitol. **82**(5): 757-62.

Meuten, D., Van Kruiningen, HJ, Lein, DH. (1974). "Cryptosporidiosis in a calf." J Am Vet Med Assoc. **165**(10): 914-7.

Miao, Y., Awad-El-Kariem, FM, Franzen, C, Ellis, DS, Müller, A, Counihan, HM, Hayes, PJ, Gazzard, BG. (2000). "Eradication of cryptosporidia and microsporidia following successful antiretroviral therapy." J Acquir Immune Defic Syndr. **25**(2): 124-9.

Millar, B., Finn, M, Xiao, LH, Lowery, CJ, Dooley, JSG, Moore, JE. (2002). "*Cryptosporidium* in foodstuffs—an emerging aetiological route of human foodborne illness " Trends Food Sci Tech. **13**(5): 168-187.

Miller, C. M., Akratos, C., Johnson, A. M., Smith, N. C. (2000). "The production of a 70 kDa heat shock protein by *Toxoplasma gondii* RH strain in immunocompromised mice." Int J Parasitol. **30**(14): 1467-73.

Moore, D., Zeman, DH. (1991). "Cryptosporidiosis in neonatal calves: 277 cases (1986-1987)." J Am Vet Med Assoc. **198**(11): 1969-71.

Mor, S. M., Tzipori, S. (2008). "Cryptosporidiosis in children in Sub-Saharan Africa: a lingering challenge." Clin Infect Dis. **47**(7): 915-21.

Morgan, U., Weber, R., Xiao, L., Sulaiman, I., Thompson, R. C., Ndiritu, W., Lal, A., Moore, A., Deplazes, P. (2000). "Molecular characterization of *Cryptosporidium* isolates obtained from human immunodeficiency virus-infected individuals living in Switzerland, Kenya, and the United States." J Clin Microbiol. **38**(3): 1180-3.

Morgan, U. M., Deplazes, P., Forbes, D. A., Spano, F., Hertzberg, H., Sargent, K. D., Elliot, A., Thompson, R. C. (1999). "Sequence and PCR-RFLP analysis of the internal transcribed spacers of the rDNA repeat unit in isolates of *Cryptosporidium* from different hosts." Parasitology. **118**: 49-58.

Morgan, U., Thompson, RC. (1998). "PCR detection of *cryptosporidium*: the way forward?" Parasitol Today. **14**(6): 241-5.

Morgan, U. M., Constantine, C. C., Forbes, D. A., Thompson, R. C. (1997). "Differentiation between human and animal isolates of *Cryptosporidium parvum* using rDNA sequencing and direct PCR analysis." J Parasitol. **83**(5): 825-30.

Morgan-Ryan, U., Fall, A, Ward, LA, Hijjawi, N, Sulaiman, I, Fayer, R, Thompson, RC, Olson, M, Lal, A, Xiao, L. (2002). "*Cryptosporidium hominis* n. sp. (Apicomplexa: Cryptosporidiidae) from *Homo sapiens*." J Eukaryot Microbiol. **49**(6): 433-40.

Morrison, L. J., Mallon, M. E., Smith, H. V., MacLeod, A., Xiao, L., Tait, A. (2008). "The population structure of the *Cryptosporidium parvum* population in Scotland: a complex picture." Infect Genet Evol. **8**(2): 121-9.

Morrison, D., Ellis, JT. (1997). "Effects of nucleotide sequence alignment on phylogeny estimation: a case study of 18S rDNAs of apicomplexa." Mol Biol Evol. **14**(4): 428-41.

Moxon, E. R., Rainey, P. B., Nowak, M. A.,Lenski, R. E. (1994). "Adaptive evolution of highly mutable loci in pathogenic bacteria." Curr Biol. **4**(1): 24-33.

Mtambo, M., Nash, AS, Blewett, DA, Wright, S. (1992). "Comparison of staining

and concentration techniques for detection of *Cryptosporidium* oocysts in cat faecal specimens." Vet Parasitol. **45**(1-2): 49-57.

Mueller-Doblies, D., Giles, M. Elwin, K., Smith, R. P., Clifton-Hadley, F. A., Chalmers, R. M. (2008). "Distribution of *Cryptosporidium* species in sheep in the UK." Vet Parasitol. **154**(3-4): 214-9.

Nagamune, K., Sibley, LD. (2006). "Comparative genomic and phylogenetic analyses of calcium ATPases and calcium-regulated proteins in the apicomplexa." Mol Biol Evol. **23**(8): 1613-27.

Nelson, J., Cai, YC, Giesler, TL, Farchaus, JW, Sundaram, ST, Ortiz-Rivera, M, Hosta, LP, Hewitt, PL, Mamone, JA, Palaniappan, C, Fuller, CW. (2002). "TempliPhi, phi29 DNA polymerase based rolling circle amplification of templates for DNA sequencing." Biotechniques. **Suppl**: 44-7.

Nesterenko, M. V., Woods, K., Upton, S. J. (1999). "Receptor/ligand interactions between *Cryptosporidium parvum* and the surface of the host cell." Biochim Biophys Acta. **1454**(2): 165-73.

Nesterenko, M. V., Tilley, M., Upton, S. J. (1995). "A metallo-dependent cysteine proteinase of *Cryptosporidium parvum* associated with the surface of sporozoites." Microbios. **83**(335): 77-88.

Nime, F., Burek, JD, Page, DL, Holscher, MA, Yardley, JH. (1976). "Acute enterocolitis in a human being infected with the protozoan *Cryptosporidium*." Gastroenterology. **70**(4): 592-8.

Nimri, L., Batchoun, R. (1994). "Prevalence of *Cryptosporidium* species in elementary school children." J Clin Microbiol. **32**(4): 1040-2.

Nishikawa, Y., Xuan, X., Nagasawa, H., Igarashi, I., Fujisaki, K., Otsuka, H., Mikami, T. (2000). "Monoclonal antibody inhibition of *Neospora caninum* tachyzoite invasion into host cells." Int J Parasitol. **30**(1): 51-8.

O'Brien, E., McInnes, L, Ryan, U. (2007). "*Cryptosporidium* GP60 genotypes from humans and domesticated animals in Australia, North America and Europe." Exp Parasitol. **118**(1): 118-21.

O'Connor, R. M., Burns, P. B., Ha-Ngoc, T., Scarpato, K., Khan, W., Kang, G., Ward, H. (2009). "Polymorphic mucin antigens CpMuc4 and CpMuc5 are integral to *Cryptosporidium parvum* infection *in vitro*." Eukaryot Cell. **8**(4): 461-9.

O'Connor, R., Wanyiri, JW, Wojczyk, BS, Kim, K, Ward, H. (2007 b). "Stable expression of *Cryptosporidium parvum* glycoprotein gp40/15 in *Toxoplasma gondii*." Mol Biochem Parasitol. **152**(2): 149-58.

O'Connor, R., Wanyiri, JW, Cevallos, AM, Priest, JW, Ward, HD. (2007). "*Cryptosporidium parvum* glycoprotein gp40 localizes to the sporozoite surface by association with gp15." Mol Biochem Parasitol. **156**(1): 80-3.

O'Connor, R. M., Kim, K., Khan, F., Ward, H. D. (2003). "Expression of Cpgp40/15 in *Toxoplasma gondii*: a surrogate system for the study of *Cryptosporidium* glycoprotein antigens." Infect Immun. **71**(10): 6027-34.

O'Donoghue, P. (1995). "*Cryptosporidium* and cryptosporidiosis in man and animals." Int J Parasitol. **25**(2): 139-95.

O'Handley, R., Olson, ME. (2006). "*Giardia*sis and cryptosporidiosis in ruminants." Vet Clin North Am Food Anim Pract. **22**(3): 623-43.

O'Hara, S. P., Yu, J. R., Lin, J. J. (2004). "A novel *Cryptosporidium parvum* antigen, CP2, preferentially associates with membranous structures." Parasitol Res. **92**(4): 317-27.

Okafor, J. I., Okunji, P. O. (1996). "Prevalence of *Cryptosporidium* oocysts in faecal

samples of some school children in Enugu State, Nigeria." J Commun Dis. **28**(1): 49-55.

Okhuysen, P. C., Chappell, C. L. (2002). "*Cryptosporidium* virulence determinants--are we there yet?" Int J Parasitol. **32**(5): 517-25.

Okhuysen, P., Chappell, CL, Crabb, JH, Sterling, CR, DuPont, HL. (1999). "Virulence of three distinct *Cryptosporidium parvum* isolates for healthy adults." J Infect Dis. **180**(4): 1275-81.

Okhuysen, P. C., Chappell, C. L., Kettner, C., Sterling, C. R. (1996). "*Cryptosporidium parvum* metalloaminopeptidase inhibitors prevent *in vitro* excystation." Antimicrob Agents Chemother. **40**(12): 2781-4.

Okhuysen, P. C., DuPont, H. L., Sterling, C. R., Chappell, C. L. (1994). "Arginine aminopeptidase, an integral membrane protein of the *Cryptosporidium parvum* sporozoite." Infect Immun. **62**(10): 4667-70.

Ong, C., Eisler, DL, Goh, SH, Tomblin, J, Awad-El-Kariem, FM, Beard, CB, Xiao, L, Sulaiman, I, Lal, A, Fyfe, M, King, A, Bowie, WR, Isaac-Renton, JL. (1999). "Molecular epidemiology of cryptosporidiosis outbreaks and transmission in British Columbia, Canada." Am J Trop Med Hyg. **61**(1): 63-9.

Pain, A., Crossman, L, Parkhill, J (2005). "Comparative Apicomplexan genomics." Nat Rev Microbiol. **3**(6): 454-5.

Panciera, R. J., Thomassen, R. W., Gardner, F. M. (1971). "Cryptosporidial infection in a calf." Vet. Pathol. **8**: 479-484.

Park, J., Beaty, TH, Boyce, P, Scott, AF, McIntosh, I. (2005). "Comparing whole-genome amplification methods and sources of biological samples for single-nucleotide polymorphism genotyping." Clin Chem. **51**(8): 1520-3.

Paunio, T., Reima, I, Syvänen, AC. (1996). "Preimplantation diagnosis by whole-genome amplification, PCR amplification, and solid-phase minisequencing of blastomere DNA." Clin Chem. **42**(9): 1382-90.

Pavlasek, I. (1999). "Cryptosporidia: biology, diagnosis, host spectrum, specificity, and the environment." Remidia Klinicka Mikrobiologie. **3**: 290-301.

Pearson, W. R., Wood, T., Zhang, Z., Miller, W. (1997). "Comparison of DNA sequences with protein sequences." Genomics. **46**(1): 24-36.

Pedraza-Díaz, S., Amar, C, Nichols, GL, McLauchlin, J. (2001). "Nested polymerase chain reaction for amplification of the *Cryptosporidium* oocyst wall protein gene." Emerg Infect Dis. **7**(1): 49-56.

Peng, M. M., Meshnick, S. R., Cunliffe, N. A., Thindwa, B. D., Hart, C. A., Broadhead, R. L., Xiao, L. (2003 b). "Molecular epidemiology of cryptosporidiosis in children in Malawi." J Eukaryot Microbiol. **50 Suppl**: 557-9.

Peng, M. M., Wilson, M. L., Holland, R. E., Meshnick, S. R., Lal, A. A., Xiao, L. (2003). "Genetic diversity of *Cryptosporidium* spp. in cattle in Michigan: implications for understanding the transmission dynamics." Parasitol Res. **90**(3): 175-80.

Peng, M. M., Matos, O., Gatei, W., Das, P., Stantic-Pavlinic, M., Bern, C., Sulaiman, I. M., Glaberman, S., Lal, A. A., Xiao, L. (2001). "A comparison of *Cryptosporidium* subgenotypes from several geographic regions." J Eukaryot Microbiol. **Suppl**: 28S-31S.

Perkins, M. E., Riojas, Y. A., Wu, T. W., Le Blancq, S. M. (1999). "CpABC, a *Cryptosporidium parvum* ATP-binding cassette protein at the host-parasite boundary in intracellular stages." Proc Natl Acad Sci U S A. **96**(10): 5734-9.

Perryman, L. E., Jasmer, D. P., Riggs, M. W., Bohnet, S. G., McGuire, T. C., Arrowood, M. J. (1996). "A cloned gene of *Cryptosporidium parvum* encodes neutralization-sensitive epitopes." Mol Biochem Parasitol. **80**(2): 137-47.

Petersen, C., Barnes, D. A., Gousset, L. (1997). "*Cryptosporidium parvum* GP900, a unique invasion protein." J Eukaryot Microbiol. **44**(6): 89S-90S.

Petersen, C., Gut, J., Doyle, P. S., Crabb, J. H., Nelson, R. G., Leech, J. H. (1992 b). "Characterization of a > 900,000-M(r) *Cryptosporidium parvum* sporozoite glycoprotein recognized by protective hyperimmune bovine colostral immunoglobulin." Infect Immun. **60**(12): 5132-8.

Petersen, C., Gut, J., Leech, J. H., Nelson, R. G. (1992). "Identification and initial characterization of five *Cryptosporidium parvum* sporozoite antigen genes." Infect Immun. **60**(6): 2343-8.

Petry, F., Kneib, I, Harris, JR. (2009). "Morphology and *in vitro* infectivity of sporozoites of *Cryptosporidium parvum*." J Parasitol. **95**(5): 1243-6. .

Pike, R., Mackenzie, GM. (1940). "Virulence of *Salmonella typhimurium*: I. Analysis of experimental infection in mice with strains of high and low virulence." J Bacteriol. **40**(2): 171-95.

Pinchbeck, G., Morrison, LJ, Tait, A, Langford, J, Meehan, L, Jallow, S, Jallow, J, Jallow, A, Christley, RM. (2008). "Trypanosomosis in The Gambia: prevalence in working horses and donkeys detected by whole genome amplification and PCR, and evidence for interactions between trypanosome species." BMC Vet Res. **4**: 7.

Pintar, K. D., Pollari, F., Waltner-Toews, D., Charron, D. F., McEwen, S. A., Fazil, A., Nesbitt, A. (2009). "A modified case-control study of cryptosporidiosis (using non-*Cryptosporidium*-infected enteric cases as controls) in a community setting." Epidemiol Infect. **137**(12): 1789-99.

Poli, G., Sorio, C., Berton, G. (1991). "Protein myristoylation in human mononclear phagocytes: modulation by interferon-gamma and tumor necrosis factor-alpha." J Cell Sci. **100 (Pt 4)**: 833-40.

Pollok, R. C., McDonald, V., Kelly, P., Farthing, M. J. (2003). "The role of *Cryptosporidium parvum*-derived phospholipase in intestinal epithelial cell invasion." Parasitol Res. **90**(3): 181-6.

Poulin, R., Combes, C. (1999). "The concept of virulence: interpretations and implications." Parasitol Today. **15**(12): 474-5.

Power, M. L., Cheung-Kwok-Sang, C., Slade, M., Williamson, S. (2009). "*Cryptosporidium fayeri*: diversity within the GP60 locus of isolates from different marsupial hosts." Exp Parasitol. **121**(3): 219-23.

Power, M., Ryan, UM. (2008). "A new species of *Cryptosporidium* (Apicomplexa: Cryptosporidiidae) from eastern grey kangaroos (Macropus giganteus)." J Parasitol. **94**(5): 1114-7.

Preiser, G., Preiser, L, Madeo, L. (2003). "An outbreak of cryptosporidiosis among veterinary science students who work with calves." J Am Coll Health. **51**(5): 213-5.

Price, H. P., Menon, M. R., Panethymitaki, C., Goulding, D., McKean, P. G., Smith, D. F. (2003). "Myristoyl-CoA:protein N-myristoyltransferase, an essential enzyme and potential drug target in kinetoplastid parasites." J Biol Chem. **278**(9): 7206-14.

Priest, J. W., Xie, L. T., Arrowood, M. J., Lammie, P. J. (2001). "The immunodominant 17-kDa antigen from *Cryptosporidium parvum* is glycosylphosphatidylinositol-anchored." Mol Biochem Parasitol. **113**(1):

117-26.

Priest, J., Kwon, JP, Arrowood, MJ, Lammie, PJ. (2000). "Cloning of the immunodominant 17-kDa antigen from *Cryptosporidium parvum*." Mol Biochem Parasitol. **106**(2): 261-71.

Puiu, D., Enomoto, S, Buck, GA, Abrahamsen, MS, Kissinger, JC. (2004). "CryptoDB: the *Cryptosporidium* genome resource." Nucleic Acids Res. **32**(Database issue): D329-31.

Putignani, L., Possenti, A., Cherchi, S., Pozio, E., Crisanti, A., Spano, F. (2008). "The thrombospondin-related protein CpMIC1 (CpTSP8) belongs to the repertoire of micronemal proteins of *Cryptosporidium parvum*." Mol Biochem Parasitol. **157**(1): 98-101.

Que, X., Reed, S. L. (2000). "Cysteine proteinases and the pathogenesis of amebiasis." Clin Microbiol Rev. **13**(2): 196-206.

Quílez, J., Sánchez-Acedo, C, Clavel, A, del Cacho, E, López-Bernad, F. (1996). "Comparison of an acid-fast stain and a monoclonal antibody-based immunofluorescence reagent for the detection of *Cryptosporidium* oocysts in faecal specimens from cattle and pigs." Vet Parasitol. **67**(1-2): 75-81.

Quintero-Betancourt, W., Peele, ER, Rose, JB. (2002). "*Cryptosporidium parvum* and *Cyclospora cayetanensis*: a review of laboratory methods for detection of these waterborne parasites." J Microbiol Methods. **49**(3): 209-24.

Ramirez, N., Sreevatsan, S. (2006). "Development of a sensitive detection system for *Cryptosporidium* in environmental samples." Vet Parasitol. **136**(3-4): 201-13.

Ramirez, N., Ward, LA, Sreevatsan, S. (2004). "A review of the biology and epidemiology of cryptosporidiosis in humans and animals." Microbes Infect. **6**(8): 773-85.

Rasmussen, K., Larsen, NC, Healey, MC. (1993). "Complete development of *Cryptosporidium parvum* in a human endometrial carcinoma cell line." Infect Immun. **61**(4): 1482-5.

Ratnam, S., Paddock, J., McDonald, E., Whitty, D., Jong, M., Cooper, R. (1985). "Occurrence of *Cryptosporidium* oocysts in fecal samples submitted for routine microbiological examination." J Clin Microbiol. **22**(3): 402-4.

Reduker, D., Speer, CA. (1985). "Factors influencing excystation in *Cryptosporidium* oocysts from cattle." J Parasitol. **71**(1): 112-5.

Reduker, D., Speer, CA, Blixt, JA. (1985). "Ultrastructure of *Cryptosporidium parvum* oocysts and excysting sporozoites as revealed by high resolution scanning electron microscopy." J Protozool. **32**(4): 708-11.

Reperant, J. M., Naciri, M., Iochmann, S., Tilley, M., Bout, D. T. (1994). "Major antigens of *Cryptosporidium parvum* recognised by serum antibodies from different infected animal species and man." Vet Parasitol. **55**(1-2): 1-13.

Rider, S. D., Jr., Zhu, G. (2010). "*Cryptosporidium*: genomic and biochemical features." Exp Parasitol. **124**(1): 2-9.

Riggs, M. W., Stone A. L., Yount, P. A., Langer, R. C., Arrowood, M. J., Bentley, D. L. (1997). "Protective monoclonal antibody defines a circumsporozoite-like glycoprotein exoantigen of *Cryptosporidium parvum* sporozoites and merozoites." J Immunol. **158**(4): 1787-95.

Riggs, M. W., Cama, V. A., Leary, H. L., Jr., Sterling, C. R. (1994). "Bovine antibody against *Cryptosporidium parvum* elicits a circumsporozoite precipitate-like reaction and has immunotherapeutic effect against persistent cryptosporidiosis in SCID mice." Infect Immun. **62**(5): 1927-39.

Riggs, M. W., McGuire, T. C., Mason, P. H., Perryman, L. E. (1989).

"Neutralization-sensitive epitopes are exposed on the surface of infectious *Cryptosporidium parvum* sporozoites." J Immunol. **143**(4): 1340-5.

Riordan, C., Langreth, SG, Sanchez, LB, Kayser, O, Keithly, JS. (1999). "Preliminary evidence for a mitochondrion in *Cryptosporidium parvum*: phylogenetic and therapeutic implications." J Eukaryot Microbiol. **46**(5): 52S-55S.

Robert, B., Ginter, A, Antoine, H, Collard, A, Coppe, P. (1990). "Diagnosis of bovine cryptosporidiosis by an enzyme-linked immunosorbent assay." Vet Parasitol. **37**(1): 1-8.

Robertson, L., Gjerde BK. (2007). "*Cryptosporidium* oocysts: challenging adversaries?" Trends Parasitol. **23**(8): 344-7.

Robertson, B., Sinclair, MI, Forbes, AB, Veitch, M, Kirk, M, Cunliffe, D, Willis, J, Fairley, CK. (2002). "Case-control studies of sporadic cryptosporidiosis in Melbourne and Adelaide, Australia." Epidemiol Infect. **128**(3): 419-31.

Robertson, L., Campbell, AT, Smith, HV. (1993). "*In vitro* excystation of *Cryptosporidium parvum*." Parasitology. **106**(Pt 1): 13-9.

Robinson, G., Chalmers, RM. (2009). "The European rabbit (*Oryctolagus cuniculus*), a source of zoonotic cryptosporidiosis." Zoonoses Public Health. **[Epub ahead of print]**.

Robinson, G. (2005). Investigating the public health significance of *Cryptosporidium* in the environment. College of Medicine. Cardiff, University of Wales. **PhD**.

Rochelle, P. A., Fallar, D., Marshall, M. M., Montelone, B. A., Upton, S. J., Woods, K. (2004). "Irreversible UV inactivation of *Cryptosporidium* spp. despite the presence of UV repair genes." J Eukaryot Microbiol. **51**(5): 553-62.

Rochelle, P., Marshall, MM, Mead, JR, Johnson, AM, Korich, DG, Rosen, JS, De Leon, R. (2002). "Comparison of *in vitro* cell culture and a mouse assay for measuring infectivity of *Cryptosporidium parvum*." Appl Environ Microbiol. **68**(8): 3809-17.

Rochelle, P., De Leon, R, Johnson, A, Stewart, MH, Wolfe, RL. (1999). "Evaluation of immunomagnetic separation for recovery of infectious *Cryptosporidium parvum* oocysts from environmental samples." Appl Environ Microbiol. **65**(2): 841-5. .

Rochelle, P., Ferguson, DM, Handojo, TJ, De Leon, R, Stewart, MH, Wolfe, RL. (1997). "An assay combining cell culture with reverse transcriptase PCR to detect and determine the infectivity of waterborne *Cryptosporidium parvum*." Appl Environ Microbiol. **63**(5): 2029-37.

Rosales, M., Cordón, GP, Moreno, MS, Sánchez, CM. (2005). "Extracellular like-gregarine stages of *Cryptosporidium parvum*." Acta Trop. **95**(1): 74-8.

Rothblatt, J. A., Meyer, D. I. (1986). "Secretion in yeast: reconstitution of the translocation and glycosylation of alpha-factor and invertase in a homologous cell-free system." Cell. **44**(4): 619-28.

Roy, S., DeLong, SM, Stenzel, SA, Shiferaw, B, Roberts, JM, Khalakdina, A, Marcus, R, Segler, SD, Shah, DD, Thomas, S, Vugia, DJ, Zansky, SM, Dietz, V, Beach, MJ; Emerging Infections Program FoodNet Working Group. (2004). "Risk factors for sporadic cryptosporidiosis among immunocompetent persons in the United States from 1999 to 2001." J Clin Microbiol. **42**(7): 2944-51.

Rusnak, J., Hadfield, TL, Rhodes, MM, Gaines, JK. (1989). "Detection of *Cryptosporidium* oocysts in human fecal specimens by an indirect immunofluorescence assay with monoclonal antibodies." J Clin Microbiol.

27(5): 1135-6.

Ryan, U., Power, M, Xiao, L. (2008). "*Cryptosporidium fayeri* n. sp. (Apicomplexa: Cryptosporidiidae) from the Red Kangaroo (*Macropus rufus*)." J Eukaryot Microbiol. **55**(1): 22-6.

Ryan, U., Monis, P, Enemark, HL, Sulaiman, I, Samarasinghe, B, Read, C, Buddle, R, Robertson, I, Zhou, L, Thompson, RC, Xiao, L. (2004). "*Cryptosporidium suis* n. sp. (Apicomplexa: Cryptosporidiidae) in pigs (*Sus scrofa*)." J Parasitol. **90**(4): 769-73.

Schaefer, D. A., Auerbach-Dixon, B. A., Riggs, M. W. (2000). "Characterization and formulation of multiple epitope-specific neutralizing monoclonal antibodies for passive immunization against cryptosporidiosis." Infect Immun. **68**(5): 2608-16.

Schmidt, W., Wahnschaffe, U, Schäfer, M, Zippel, T, Arvand, M, Meyerhans, A, Riecken, EO, Ullrich, R. (2001). "Rapid increase of mucosal CD4 T cells followed by clearance of intestinal cryptosporidiosis in an AIDS patient receiving highly active antiretroviral therapy." Gastroenterology. **120**(4): 984-7.

Schupp, D. G., Erlandsen, S L (1987). "A new method to determine *Giardia* cyst viability: correlation of fluorescein diacetate and propidium iodide staining with animal infectivity." Appl Environ Microbiol. **53**(4): 704–707.

Sestak, K., Ward, LA, Sheoran, A, Feng, X, Akiyoshi, DE, Ward, HD, Tzipori, S. (2002). "Variability among *Cryptosporidium parvum* genotype 1 and 2 immunodominant surface glycoproteins." Parasite Immunol. **24**(4): 213-9.

Shenai, B. R., Sijwali, P. S., Singh, A., Rosenthal, P. J. (2000). "Characterization of native and recombinant falcipain-2, a principal trophozoite cysteine protease and essential hemoglobinase of Plasmodium falciparum." J Biol Chem. **275**(37): 29000-10.

Shin, M. H., Kita, H., Park, H. Y., Seoh, J. Y. (2001). "Cysteine protease secreted by Paragonimus westermani attenuates effector functions of human eosinophils stimulated with immunoglobulin G." Infect Immun. **69**(3): 1599-604.

Shirafuji, H., Xuan, X., Kimata, I., Takashima, Y., Fukumoto, S., Otsuka, H., Nagasawa, H., Suzuki, H. (2005). "Expression of P23 of *Cryptosporidium parvum* in *Toxoplasma gondii* and evaluation of its protective effects." J Parasitol. **91**(2): 476-9.

Shoaib, M., Baconnais, S, Mechold, U, Le Cam, E, Lipinski, M, Ogryzko, V. (2008). "Multiple displacement amplification for complex mixtures of DNA fragments." BMC Genomics. **9**: 415.

Singer, V., Jones, LJ, Yue, ST, Haugland, RP. (1997). "Characterization of PicoGreen reagent and development of a fluorescence-based solution assay for double-stranded DNA quantitation." Anal Biochem. **249**(2): 228-38.

Siński, E., Behnke, JM. (2004). "Apicomplexan parasites: environmental contamination and transmission." Pol J Microbiol. **53**(Suppl): 67-73. .

Slavin, D. (1955). "*Cryptosporidium* meleagridis (sp. nov.)." J Comp Pathol. **65**(3): 262-6.

Slifko, T., Friedman, D., Rose, J. B, Jakubowski, W. (1997). "An *in vitro* method for detecting infectious *Cryptosporidium* oocysts with cell culture." Appl Environ Microbiol. **63**(9): 3669-75.

Smerdon, W., Nichols, T, Chalmers, RM, Heine, H, Reacher, MH. (2003). "Foot and mouth disease in livestock and reduced cryptosporidiosis in humans, England and Wales." Emerg Infect Dis. **9**(1): 22-8.

Smiley, H., Pearse, H. (1926). "Determining the virulence of positive diphtheria cultures." Am J Public Health (N Y). **16**(10): 998-1002.

Smith, H., Cacciò, SM, Cook, N, Nichols, RA, Tait, A. (2007). "*Cryptosporidium* and *Giardia* as foodborne zoonoses." Vet Parasitol. **149**(1-2): 29-40.

Smith, H., Cacciò, SM, Tait, A, McLauchlin, J, Thompson, RC. (2006). "Tools for investigating the environmental transmission of *Cryptosporidium* and *Giardia* infections in humans." Trends Parasitol. **22**(4): 160-7.

Smith, H., Nichols, RA, Grimason, AM. (2005). "*Cryptosporidium* excystation and invasion: getting to the guts of the matter." Trends Parasitol. **21**(3): 133-42.

Smith, H., Corcoran, GD. (2004). "New drugs and treatment for cryptosporidiosis." Curr Opin Infect Dis. **17**(6): 557-64.

Smith, K., Stenzel, SA., Bender, JB, Wagstrom, E, Soderlund, D, Leano, FT, Taylor, CM, Belle-Isle, PA, Danila, R. (2004). "Outbreaks of enteric infections caused by multiple pathogens associated with calves at a farm day camp." Pediatr Infect Dis J. **23**(12): 1098-104.

Smith, H., Rose, JB. (1998). "Waterborne cryptosporidiosis: current status." Parasitol Today. **14**(1): 14-22.

Snabes, M., Chong, SS, Subramanian, SB, Kristjansson, K, DiSepio, D, Hughes, MR. (1994). "Preimplantation single-cell analysis of multiple genetic loci by whole-genome amplification." Proc Natl Acad Sci U S A. **91**(13): 6181-5.

Soave, R., Danner, RL, Honig, CL, Ma, P, Hart, CC, Nash, T, Roberts, RB. (1984). "Cryptosporidiosis in homosexual men." Ann Intern Med. **100**(4): 504-11.

Soldati, D., Dubremetz, J. F., Lebrun, M. (2001). "Microneme proteins: structural and functional requirements to promote adhesion and invasion by the apicomplexan parasite *Toxoplasma gondii*." Int J Parasitol. **31**(12): 1293-302.

Spano, F., Putignani, L., Naitza, S., Puri, C., Wright, S., Crisanti, A. (1998 b). "Molecular cloning and expression analysis of a *Cryptosporidium parvum* gene encoding a new member of the thrombospondin family." Mol Biochem Parasitol. **92**(1): 147-62.

Spano, F., Putignani, L, Guida, S, Crisanti, A. (1998). "*Cryptosporidium parvum*: PCR-RFLP analysis of the TRAP-C1 (thrombospondin-related adhesive protein of *Cryptosporidium*-1) gene discriminates between two alleles differentially associated with parasite isolates of animal and human origin." Exp Parasitol. **90**: 195-8.

Spano, F., Putignani, L, McLauchlin, J, Casemore, DP, Crisanti, A. (1997). "PCR-RFLP analysis of the *Cryptosporidium* oocyst wall protein (COWP) gene discriminates between C. wrairi and *C. parvum*, and between *C. parvum* isolates of human and animal origin." FEMS Microbiol Lett. **150**(2): 209-17.

Steele, M. I., Kuhls, T. L., Nida, K., Meka, C. S., Halabi, I. M., Mosier, D. A., Elliott, W., Crawford, D. L., Greenfield, R. A. (1995). "A *Cryptosporidium parvum* genomic region encoding hemolytic activity." Infect Immun. **63**(10): 3840-5.

Stibbs, H., Ongerth, JE. (1986). "Immunofluorescence detection of *Cryptosporidium* oocysts in fecal smears." J Clin Microbiol. **24**(4): 517-21.

Stinear, T., Matusan, A, Hines, K, Sandery, M. (1996). "Detection of a single viable *Cryptosporidium parvum* oocyst in environmental water concentrates by reverse transcription-PCR." Appl Environ Microbiol. **62**(9): 3385-90.

Striepen, B., Kissinger, JC. (2004). "Genomics meets transgenics in search of the elusive *Cryptosporidium* drug target." Trends Parasitol. **20**(8): 355-8.

Strong, W., Gut, J, Nelson, RG. (2000). "Cloning and sequence analysis of a highly

polymorphic *Cryptosporidium parvum* gene encoding a 60-kilodalton glycoprotein and characterization of its 15- and 45-kilodalton zoite surface antigen products." Infect Immun. **68**(7): 4117-34.

Sturbaum, G., Reed, C, Hoover, PJ, Jost, BH, Marshall, MM, Sterling, CR. (2001). "Species-specific, nested PCR-restriction fragment length polymorphism detection of single *Cryptosporidium parvum* oocysts." Appl Environ Microbiol. **67**(6): 2665-8.

Sulaiman, I., Hira, PR, Zhou, L, Al-Ali, FM, Al-Shelahi, FA, Shweiki, HM, Iqbal, J, Khalid, N, Xiao, L. (2005). "Unique endemicity of cryptosporidiosis in children in Kuwait." J Clin Microbiol. **43**(6): 2805-9.

Sulaiman, I., Lal, AA, Xiao, L. (2001). "A population genetic study of the *Cryptosporidium parvum* human genotype parasites." J Eukaryot Microbiol. (Suppl): 24S-27S.

Sulaiman, I., Morgan, UM, Thompson, RC, Lal, AA, Xiao, L. (2000). "Phylogenetic relationships of *Cryptosporidium* parasites based on the 70-kilodalton heat shock protein (HSP70) gene." Appl Environ Microbiol. **66**(6): 2385-91.

Sulaiman, I., Xiao, L, Lal, AA. (1999). "Evaluation of *Cryptosporidium parvum* genotyping techniques." Appl Environ Microbiol. **65**(10): 4431-5.

Sulaiman, I., Xiao, L, Yang, C, Escalante, L, Moore, A, Beard, CB, Arrowood, MJ, Lal, AA. (1998). "Differentiating human from animal isolates of *Cryptosporidium parvum*." Emerg Infect Dis. **4**(4): 681-5.

Sundermann, C., Lindsay, DS, Blagburn, BL. (1987). "*In vitro* excystation of *Cryptosporidium baileyi* from chickens." J Protozool. **34**(1): 28-30.

Sunnotel, O., Lowery, CJ, Moore, JE, Dooley, JS, Xiao, L, Millar, BC, Rooney, PJ, Snelling, WJ. (2006). "*Cryptosporidium*." Lett Appl Microbiol. **43**(1): 7-16.

Tait., A., Wastling, JM, Smith, H, MacLeod, A, Mallon, ME (2004). "Response to: population genetics of *Cryptosporidium parvum* " Trends Parasitol. **20**(1): 6.

Tamura, K., Dudley, J, Nei, M, Kumar, S. (2007). "MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0." Mol Biol Evol. **24**(8): 1596-9. .

Tanriverdi, S., Grinberg, A., Chalmers, R. M., Hunter, P. R., Petrovic, Z., Akiyoshi, D. E., London, E., Zhang, L., Tzipori, S., Tumwine, J. K., Widmer, G. (2008). "Inferences about the global population structures of *Cryptosporidium parvum* and *Cryptosporidium hominis*." Appl Environ Microbiol. **74**(23): 7227-34.

Tanriverdi, S., Widmer, G. (2006). "Differential evolution of repetitive sequences in *Cryptosporidium parvum* and *Cryptosporidium hominis*." Infect Genet Evol. **6**(2): 113-22.

Tanriverdi, S., Arslan, MO, Akiyoshi, DE, Tzipori, S, Widmer, G. (2003). "Identification of genotypically mixed *Cryptosporidium parvum* populations in humans and calves." Mol Biochem Parasitol. **130**(1): 13-22.

Telenius, H., Carter, NP, Bebb, CE, Nordenskjöld, M, Ponder, BA, Tunnacliffe, A. (1992). "Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer." Genomics. **13**(3): 718-25.

Templeton, T., Iyer, LM, Anantharaman, V, Enomoto, S, Abrahante, JE, Subramanian, GM, Hoffman, SL, Abrahamsen, MS, Aravind, L. (2004). "Comparative analysis of apicomplexa and genomic diversity in eukaryotes." Genome Res. **14**(9): 1686-95.

Tetley, L., Brown, S. M., McDonald, V., Coombs, G. H. (1998). "Ultrastructural analysis of the sporozoite of *Cryptosporidium parvum*." Microbiology. **144**

(12): 3249-55.

Thompson, H., Dooley, JS, Kenny, J, McCoy, M, Lowery, CJ, Moore, JE, Xiao, L. (2007). "Genotypes and subtypes of *Cryptosporidium* spp. in neonatal calves in Northern Ireland." Parasitol Res. **100**(3): 619-24.

Thorstenson, Y., Hunicke-Smith, SP, Oefner, PJ, Davis, RW. (1998). "An automated hydrodynamic process for controlled, unbiased DNA shearing." Genome Research. **8**(8): 848-55.

Tilley, M., Eggleston, M. T., Upton, S. J. (1993). "Multiple oral inoculations with *Cryptosporidium parvum* as a means of immunization for production of monoclonal antibodies." FEMS Microbiol Lett. **113**(2): 235-40.

Tomley, F. (2009). "Apicomplexan biology in the post-genomic era: perspectives from the European COST Action 857." Int J Parasitol. **39**(2): 133-4.

Tomley, F. M., Soldati, D. S. (2001). "Mix and match modules: structure and function of microneme proteins in apicomplexan parasites." Trends Parasitol. **17**(2): 81-8.

Tosini, F., Agnoli, A., Mele, R., Gomez Morales, M. A., Pozio, E. (2004). "A new modular protein of *Cryptosporidium parvum*, with ricin B and LCCL domains, expressed in the sporozoite invasive stage." Mol Biochem Parasitol. **134**(1): 137-47.

Towler, D. A., Gordon, J. I., Adams, S. P., Glaser, L. (1988). "The biology and enzymology of eukaryotic protein acylation." Annu Rev Biochem. **57**: 69-99.

Trotz-Williams, L., Martin, DS, Gatei, W, Cama, V, Peregrine, AS, Martin, SW, Nydam, DV, Jamieson, F, Xiao, L. (2006). "Genotype and subtype analyses of *Cryptosporidium* isolates from dairy calves and humans in Ontario." Parasitol Res. **99**(4): 346-52.

Tyzzer, E. E. (1929). "Coccidiosis in gallinaceous birds." Am. J. Hyg. **10**: 269-383.

Tyzzer, E. E. (1912). "*Cryptosporidium parvum* (sp. nov.), a coccidium found in the small intestine of the common mouse." Arch. Protistenkd. **26**: 394-412.

Tyzzer, E. E. (1910). "An extracellular coccidium, *Cryptosporidium* muris (gen et sp nov.) of the gastric glands of the common mouse." J. Med. Res. **23**: 487-516.

Tyzzer, E. E. (1907). "A sporozoan found in the peptic glands of the common mouse." Proc. Soc. Exp. Biol. Med. **5**: 12-13.

Tzipori, S., Widmer, G. (2008). "A hundred-year retrospective on cryptosporidiosis." Trends Parasitol. **24**(4): 184-9.

Tzipori, S., Ward, H (2002). "Cryptosporidiosis: biology, pathogenesis and disease." Microbes Infect. **4**(10): 1047-58.

Tzipori, S., Griffiths, JK. (1998). "Natural history and biology of *Cryptosporidium parvum*." Adv Parasitol. **40**: 5-36.

Tzipori, S. (1998). "Cryptosporidiosis: laboratory investigations and chemotherapy." Adv Parasitol. **40**: 187-221.

Tzipori, S., Smith, M, Birch, C, Barnes, G, Bishop, R. (1983). "Cryptosporidiosis in hospital patients with gastroenteritis." Am J Trop Med Hyg. **32**(5): 931-4.

Tzipori, S., Campbell, I, Sherwood, D, Snodgrass, DR, Whitelaw, A. (1980). "An outbreak of calf diarrhoea attributed to cryptosporidial infection." Vet Rec. **107**(25-26): 579-80.

Ungar, B. (1990). "Enzyme-linked immunoassay for detection of *Cryptosporidium* antigens in fecal specimens." J Clin Microbiol. **28**(11): 2491-5.

Upton, S., Tilley, M, Brillhart, DB. (1995). "Effects of select medium supplements on *in vitro* development of *Cryptosporidium parvum* in HCT-8 cells." J Clin Microbiol. **33**(2): 371-5.

Upton, S., Tilley, M, Nesterenko, MV, Brillhart, DB. (1994 b). "A simple and reliable method of producing *in vitro* infections of *Cryptosporidium parvum* (Apicomplexa)." FEMS Microbiol Lett. **118**(1-2): 45-9.

Upton, S., Tilley, M, Brillhart, DB. (1994). "Comparative development of *Cryptosporidium parvum* (Apicomplexa) in 11 continuous host cell lines." FEMS Microbiol Lett. **118**(3): 233-6.

Valderrama, A. L., Hlavsa, M. C., Cronquist, A., Cosgrove, S., Johnston, S. P., Roberts, J. M., Stock, M. L., Xiao, L., Xavier, K., Beach, M. J. (2009). "Multiple risk factors associated with a large statewide increase in cryptosporidiosis." Epidemiol Infect. **137**(12): 1781-8.

Van Valen, L. (1973). "A new evolutionary law." Evol. Theory **1**: 1-30.

Venditti, C., Meade, A., Pagel, M. (2010). "Phylogenies reveal new interpretation of speciation and the Red Queen." Nature. **463**(7279): 349-52.

Vesey, G., Ashbolt, N, Fricker, EJ, Deere, D, Williams, KL, Veal, DA, Dorsch, M. (1998). "The use of a ribosomal RNA targeted oligonucleotide probe for fluorescent labelling of viable *Cryptosporidium parvum* oocysts." J Appl Microbiol. **85**(3): 429-40.

Vetterling, J., Jervis, HR, Merrill, TG, Sprinz, H. (1971). "*Cryptosporidium wrairi* sp. n. from the guinea pig *Cavia porcellus*, with an emendation of the genus." J Protozool. **18**(2): 243-7.

Vincze, T., Posfai, J, Roberts, RJ. (2003). "NEBcutter: A program to cleave DNA with restriction enzymes." Nucleic Acids Res. **31**(13): 3688-91.

Wagner, S., Klepsch, MM, Schlegel, S, Appel, A, Draheim, R, Tarry, M, Högbom, M, van Wijk, KJ, Slotboom, DJ, Persson, JO, de Gier, JW. (2008). "Tuning *Escherichia coli* for membrane protein overexpression." Proc Natl Acad Sci U S A. **105**(38): 14371-6.

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**(8): 2789-97.

Wanyiri, J. W., Techasintana, P., O'Connor, R. M., Blackman, M. J., Kim, K., Ward, H. D. (2009). "Role of CpSUB1, a subtilisin-like protease, in *Cryptosporidium parvum* infection *in vitro*." Eukaryot Cell. **8**(4): 470-7.

Wanyiri, J. W., O'Connor, R., Allison, G., Kim, K., Kane, A., Qiu, J., Plaut, A. G., Ward, H. D. (2007). "Proteolytic processing of the *Cryptosporidium* glycoprotein gp40/15 by human furin and by a parasite-derived furin-like protease activity." Infect Immun. **75**(1): 184-92.

Wanyiri, J., Ward, H. (2006). "Molecular basis of *Cryptosporidium*-host cell interactions: recent advances and future prospects." Future Microbiol. **1**: 201-8.

Ward, P., Deplazes, P, Regli, W, Rinder, H, Mathis, A. (2002). "Detection of eight *Cryptosporidium* genotypes in surface and waste waters in Europe." Parasitology. **124**(Pt 4): 359-68.

Wastling, J. M., Xia, D., Sohal, A., Chaussepied, M., Pain, A., Langsley, G. (2009). "Proteomes and transcriptomes of the Apicomplexa--where's the message?" Int J Parasitol. **39**(2): 135-43.

Weidner, M., Taupp, M., Hallam, S. J. (2010). "Expression of recombinant proteins in the methylotrophic yeast *Pichia pastoris*." J Vis Exp. (36).

Wetzel, D. M., Schmidt, J., Kuhlenschmidt, M. S., Dubey, J. P., Sibley, L. D. (2005). "Gliding motility leads to active cellular invasion by *Cryptosporidium*

parvum sporozoites." Infect Immun. **73**(9): 5379-87.

Widmer, G. (2009). "Meta-analysis of a polymorphic surface glycoprotein of the parasitic protozoa *Cryptosporidium parvum* and *Cryptosporidium hominis*." Epidemiol Infect. **137**(12): 1800-8.

Widmer, G., Carlton, JM, Silva, JC, London, E. (2007). The *Cryptosporidium* muris genome project: a progress report. . II International *Giardia* and *Cryptosporidium* Conference. Centro Cultural Universitario, Morelia, Michoacan, Mexico.May 13-18. **64**.

Widmer, G., Lin, L., Kapur, V., Feng, X., Abrahamsen, M. S. (2002). "Genomics and genetics of *Cryptosporidium parvum*: the key to understanding cryptosporidiosis." Microbes Infect. **4**(10): 1081-90.

Widmer, G., Akiyoshi, D, Buckholt, MA, Feng, X, Rich, SM, Deary, KM, Bowman, CA, Xu, P, Wang, Y, Wang, X, Buck, GA, Tzipori, S. (2000). "Animal propagation and genomic survey of a genotype 1 isolate of *Cryptosporidium parvum*." Mol Biochem Parasitol. **108**(2): 187-97.

Widmer, G., Orbacz, EA, Tzipori, S. (1999). "Beta-tubulin mRNA as a marker of *Cryptosporidium parvum* oocyst viability." Appl Environ Microbiol. **65**(4): 1584-8.

Widmer, G., Tchack, L., Chappell, C. L., Tzipori, S. (1998). "Sequence polymorphism in the beta-tubulin gene reveals heterogeneous and variable population structures in *Cryptosporidium parvum*." Appl Environ Microbiol. **64**(11): 4477-81.

Wiedenmann, A., Kruger, P., Botzenhart, K. (1998). "PCR detection of *Cryptosporidium parvum* in environmental samples - a review of published protocols and current developments." J. Indust. Microbiol. Biotechnol. **21**(3): 150-66.

Willemot, P., Klein, MB. (2004). "Prevention of HIV-associated opportunistic infections and diseases in the age of highly active antiretroviral therapy." Expert Rev Anti Infect Ther. **2**(4): 521-32.

Winter, G., Gooley, A. A., Williams, K. L., Slade, M. B. (2000). "Characterization of a major sporozoite surface glycoprotein of *Cryptosporidium parvum*." Funct Integr Genomics. **1**(3): 207-17.

Woods, K., Upton, SJ. (2007). "*In vitro* development of *Cryptosporidium parvum* in serum-free media." Lett Appl Microbiol. **44**(5): 520-3.

Woods, K. M., Tilley, M., Iseli, A., Upton, S. J., Montelone, B. A., Khramtsov, N. V. (1999). "Sequence of the gene encoding hsp90e from *Cryptosporidium parvum*." DNA Seq. **10**(4-5): 339-42.

Woods, K., Nesterenko, MV, Upton, SJ. (1996). "Efficacy of 101 antimicrobials and other agents on the development of *Cryptosporidium parvum in vitro*." Ann Trop Med Parasitol. **90**(6): 603-15.

Woolhouse, M. E., Webster, J. P., Domingo, E., Charlesworth, B., Levin, B. R. (2002). "Biological and biomedical implications of the co-evolution of pathogens and their hosts." Nat Genet. **32**(4): 569-77.

Wu, Z., Nagano, I., Boonmars, T., Nakada, T., Takahashi, Y. (2003). "Intraspecies polymorphism of *Cryptosporidium parvum* revealed by PCR-restriction fragment length polymorphism (RFLP) and RFLP-single-strand conformational polymorphism analyses." Appl Environ Microbiol. **69**(8): 4720-6.

Xiao, L. (2010). "Molecular epidemiology of cryptosporidiosis: an update." Exp Parasitol. **124**(1): 80-9.

Xiao, L., Hlavsa, M. C., Yoder, J., Ewers, C., Dearen, T., Yang, W., Nett, R., Harris, S., Brend, S. M., Harris, M., Onischuk, L., Valderrama, A. L., Cosgrove, S., Xavier, K., Hall, N., Romero, S., Young, S., Johnston, S. P., Arrowood, M., Roy, S., Beach, M. J. (2009). "Subtype analysis of *Cryptosporidium* specimens from sporadic cases in Colorado, Idaho, New Mexico, and Iowa in 2007: widespread occurrence of one *Cryptosporidium hominis* subtype and case history of an infection with the *Cryptosporidium* horse genotype." J Clin Microbiol. **47**(9): 3017-20.

Xiao, L., Fayer, R. (2008). "Molecular characterisation of species and genotypes of *Cryptosporidium* and *Giardia* and assessment of zoonotic transmission." Int J Parasitol. **38** (11): 1239-55.

Xiao, L., Feng, Y. (2008). "Zoonotic cryptosporidiosis." FEMS Immunol Med Microbiol. **52**(3): 309-23.

Xiao, L., Ryan, UM. (2004). "Cryptosporidiosis: an update in molecular epidemiology." Curr Opin Infect Dis. **17**(5): 483-90.

Xiao, L., Fayer, R, Ryan, U, Upton, SJ. (2004). "*Cryptosporidium* taxonomy: recent advances and implications for public health." Clin Microbiol Rev. **17**(1): 72-97.

Xiao, L., Bern, C, Limor, J, Sulaiman, I, Roberts, J, Checkley, W, Cabrera, L, Gilman, RH, Lal, AA. (2001 b). "Identification of 5 types of *Cryptosporidium* parasites in children in Lima, Peru." J Infect Dis. **183**(3): 492-7.

Xiao, L., Singh, A, Limor, J, Graczyk, TK, Gradus, S, Lal, A. (2001). "Molecular characterization of *Cryptosporidium* oocysts in samples of raw surface water and wastewater." Appl Environ Microbiol. **67**(3): 1097-101.

Xiao, L., Alderisio, K, Limor, J, Royer, M, Lal, AA. (2000). "Identification of species and sources of *Cryptosporidium* oocysts in storm waters with a small-subunit rRNA-based diagnostic and genotyping tool." Appl Environ Microbiol. **66**(12): 5492-8.

Xiao, L., Escalante, L, Yang, C, Sulaiman, I, Escalante, AA, Montali, RJ, Fayer, R, Lal, AA. (1999). "Phylogenetic analysis of *Cryptosporidium* parasites based on the small-subunit rRNA gene locus." Appl Environ Microbiol. **65**(4): 1578-83.

Xiao, L., Herd, RP, Rings, DM. (1993). "Diagnosis of *Cryptosporidium* on a sheep farm with neonatal diarrhea by immunofluorescence assays." Vet Parasitol. **47**(1-2): 17-23.

Xu, P., Widmer, G, Wang, Y, Ozaki, LS, Alves, JM, Serrano, MG, Puiu, D, Manque, P, Akiyoshi, D, Mackey, AJ, Pearson, WR, Dear, PH, Bankier, AT, Peterson, DL, Abrahamsen, MS, Kapur, V, Tzipori, S, Buck ,GA. (2004). "The genome of *Cryptosporidium hominis*." Nature. **431**(7012): 1107-12.

Yang, X., Figueiredo, L. M., Espinal, A., Okubo, E., Li, B. (2009). "RAP1 is essential for silencing telomeric variant surface glycoprotein genes in *Trypanosoma brucei*." Cell. **137**(1): 99-109.

Yang, S., Healey, MC, Du, C, Zhang, J. (1996). "Complete development of *Cryptosporidium parvum* in bovine fallopian tube epithelial cells." Infect Immun. **64**(1): 349-54.

Yoder, J. S., Beach, M. J. (2010). "*Cryptosporidium* surveillance and risk factors in the United States." Exp Parasitol. **124**(1): 31-9.

Zarlenga, D. S., Gasbarre, L. C. (2009). "From parasite genomes to one healthy world: are we having fun yet?" Vet Parasitol. **163**(3): 235-49.

Zhang, L., Cui, X, Schmitt, K, Hubert, R, Navidi, W, Arnheim, N. (1992). "Whole

genome amplification from a single cell: implications for genetic analysis." Proc Natl Acad Sci U S A. **89**(13): 5847-51.

Zhou, X. (2009). "Proteomics and parasite-host interactions " Amino Acids **37:** 122.

Zhou, L., Singh, A., Jiang, J., Xiao, L. (2003). "Molecular surveillance of *Cryptosporidium* spp. in raw wastewater in Milwaukee: implications for understanding outbreak occurrence and transmission dynamics." J Clin Microbiol. **41**(11): 5254-7.

Zhu, G., LaGier, M. J., Stejskal, F., Millership, J. J., Cai, X., Keithly, J. S. (2002). "*Cryptosporidium parvum*: the first protist known to encode a putative polyketide synthase." Gene. **298**(1): 79-89.

Zhu, G., Keithly, JS, Philippe, H. (2000 a). "What is the phylogenetic position of *Cryptosporidium*?" Int J Syst Evol Microbiol. **50**: 1673-81.

Zhu, G., Marchewka, MJ, Keithly, JS. (2000 b). "*Cryptosporidium parvum* appears to lack a plastid genome." Microbiology. **146**(2): 315-21.

Zhu, G., Marchewka, MJ, Ennis, JG, Keithly, JS. (1998). "Direct isolation of DNA from patient stools for polymerase chain reaction detection of *Cryptosporidium parvum*." J Infect Dis. **177**(5): 1443-6.

Zintl, A., Proctor, A. F., Read, C., Dewaal, T., Shanaghy, N., Fanning, S., Mulcahy, G. (2009). "The prevalence of *Cryptosporidium* species and subtypes in human faecal samples in Ireland." Epidemiol Infect. **137**(2): 270-7.