Genome Analyses of Filamentous Pathogen-Plant Interactions

LILIANA MARIA CANO MOGROVEJO

Thesis submitted to the University of East Anglia for the degree of Doctor of Philosophy

September 2011

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Plant pathogens encode effectors with N-terminal signal peptides that are secreted to reprogram the host and enable parasitic infection. Here, I used the Signal Sequence Trap (SST) genetic assay to functionally validate the signal peptides (SP) of four representative cytoplasmic RXLR effector genes of the Irish famine pathogen *Phytophthora infestans* that are induced *in planta* and that can trigger or suppress defenses. I found that the SP of these RXLRs are functional in yeast and confirm previous observations that predictions obtained with signalPv2.0 are highly accurate. Protease inhibitors belong to another class of effectors that are secreted in the apoplast, and that were firstly identified in *P. infestans.* I annotated the protease inhibitor effector repertoires of recently sequenced oomycete genomes. The results confirmed previous observations that these effectors are common features of oomycetes pathogens, probably because they can serve as a powerful counterdefense mechanism. *P. infestans* and other three closely related *Phytophthora* species (clade 1c) evolve by host jumps followed by specialization on plants belonging to four different botanical families. Comparative genome analyses of the *Phytophthora* clade 1c revealed that dynamic gene spare repeat-rich genome compartments (GSR) are enriched in genes with accelerated gene evolution. GSRs are also enriched in induced-*in planta* genes, implicating host adaption in genome evolution. Within the *P. infestans* lineage, a new emerging clone 13_A2 that overcome previously effective forms of plant host resistance has been identified. Genome analyses of a 13_A2 isolate 06_3928A revealed significant genetic and expression polymorphisms in effector genes, including known *Avrs.* Importantly, some *Avrs* were still induced *in planta,* intact and recognized by their cognate *R* genes. These conserved *Avrs* can be used as a genetic strategy for mitigating the impact of 13_A2 epidemics. Finally, I investigated the transcriptional changes occurring in *Boechera stricta* plant during the formation of pseudoflowers by the rust fungus pathogen *Puccinia monoica.* The results suggest that several biological processes are significantly differentially regulated in pseudoflowers. This study is the first step towards understanding at a molecular level how this rust fungus pathogen manipulates its host plant.

ABBREVIATIONS

aa amino acids
RP Rank Products
FDR False Discovery Rate
°C Degrees Centigrade
µg micro gram
µl micro litre
µM micro molar
bp base pairs
DNA Deoxy ribonucleic acid
BSA Bovine Serum Albumin
mg milli gram
ml milli litre
mM milli molar
PCR Polymerase Chain Reaction
R Resistance protein
RNA Ribonucleic acid
SST Signal Sequence Trap System
CWD minimum media minus tryptophan
YPRAA yeast peptone raffinose antimycin media
TTC 2,3,5-Triphenyltetrazolium Chloride
ORF Open Reading Frame
EPIC extracellular cystatin-like cysteine protease inhibitors
EPI extracellular Kazal-like serine protease inhibitors
GSR gene sparse regions
GDR gene dense regions
AVR avirulence protein
R resistance protein
Pf Pseudoflowers
F Host Flowers
SL Host Stem and Leaves
SAM Shoot Apical Meristem
PGP1 P-GLYCOPROTEIN2
PGP9 P-GLYCOPROTEIN9
ICU4 INCURVATA4
AMP1 ALTERED MERISTEM PROGRAM1
PHV PHAVOLUTA
NGA3 NGATHA3
TAA1 TRYPTOPHAN AMINOTRANSFERASE OF ARABIDOPSIS1
WAG2 KINASE PROTEIN SERINE/THREONINE KINASE ACTVITY
TCP2 TEOSINTE BRANCHED, CYCLOIDEA, and PCF1
TCP3 TEOSINTE BRANCHED, CYCLOIDEA, and PCF2
PKS1 PHYTOCHROME KINASE SUBSTRATE1
PMI2 PLASTID MOVEMENT IMPAIRED2
RTFL2 ROTUNDIFOLIA-LIKE
CYP78A5 CYTOCHROME P450 MONOOXYGENASE
MAX1 MORE AXILLARY GROWTH1
FRA8 FRAGILE FIBER8

IRX3 IRREGULAR XYLEM3
IRX8 IRREGULAR XYLEM8
IRX9 IRREGULAR XYLEM9
IRX10 IRREGULAR XYLEM10
IRX12 IRREGULAR XYLEM12
IRX14 IRREGULAR XYLEM14
IRX14-L IRREGULAR XYLEM14-LIKE
PGSIP1/ GUX1 PLANT GLYCOGENIN-LIKE STARCH INITIATION PROTEIN1
PGSIP3/ GUX2 PLANT GLYCOGENIN-LIKE STARCH INITIATION PROTEIN3
GATL1 GALACTURONOSYLTRANSFERASE-LIKE1
GATL-like GALACTURONOSYLTRANSFERASE-LIKE
NST1 NAC (NO APICAL MERISTEM) SECONDARY WALL THICKENING PROMOTING FACTOR 1
NST3 NAC (NO APICAL MERISTEM) SECONDARY WALL THICKENING PROMOTING FACTOR3
TBL3 TRICHOME BIREFRINGENCE-LIKE3
CESA8 CELLULOSE SYNTHASE 8
KCS8 3-KETOACYL-COA SYNTHASE8
WSD7 WAX ESTER SYNTHASE/ACYLCOA: DIACYLGLYCEROL ACETYLTRANSFERASE7
DCR CUTICULAR RIDGES
ABCG13 ATP-BINDING-CASSETTE (ABC) TRANSPORTERS SUPERFAMILY G 13
IDD14 INDETERMINANT DOMAIN14
SUS1 SUCROSE SYNTHASE1
SUS4 SUCROSE SYNTHASE4
FT FLOWERING LOCUS T
QRT2 QUARTER2
AFO ABNORMAL FLORAL ORGANS1
KNAT1 KNOTTED-LIKE1
PNF POUND-FOOLISH
SEP4/ AGL3 SEPATALLA4
DRF DIHYDROFLAVONOL 4-REDUCTASE
LDOX LEUCOANTHOCYANIDIN DIOXYGENASE
SWEET1 SUGAR TRANSPORTER1
SWEET15 SUGAR TRANSPORTER15
cwINV1 CELL WALL INVERTASE1
TPS10 TERPENE SYNTHASE10
TPS21 TERPENE SYNTHASE21
GH3.2 IAA AMINO ACID SYNTHASE, AUXIN-RESPONSIVE GH3 FAMILY PROTEIN
GH3.4 IAA AMINO ACID SYNTHASE, AUXIN-RESPONSIVE GH3 FAMILY PROTEIN
ARF18 AUXIN RESPONSE FACTOR18
ATGSTU1 GLUTATHIONE S-TRANSFERASE TAU1
ATGSTU2 GLUTATHIONE S-TRANSFERASE TAU2
ATGSTU4 GLUTATHIONE S-TRANSFERASE TAU4
ATGSTU17 GLUTATHIONE S-TRANSFERASE TAU17
ATGSTU26 GLUTATHIONE S-TRANSFERASE TAU26
RAP2.6L RELATED TO AP2 6L

Remorin REMORIN FAMILY PROTEIN
LEA4 LATE EMBRYOGENESIS ABUNDANT PROTEIN4
LEA family LATE EMBRYOGENESIS ABUNDANT FAMILY PROTEIN
LEA family LATE EMBRYOGENESIS ABUNDANT FAMILY PROTEIN

ACKNOWLEDGMENTS

I would like to thank my supervisor, Professor Sophien Kamoun, for giving me the opportunity to study at the Sainsbury Laboratory and for his support with my scientific work. He has been a fantastic leader and a great motivator throughout my studies. I would also like to thank Sylvain Raffaele who has been an incredible mentor and friend. All of the members of the Kamoun Lab have provided me with their support and help during the past four years and I am very grateful for this. All members of my family back in Colombia have given me their endless sharing and support, especially to my father Libardo Cano who had inspired me to pursue adventures abroad. Finally to my fiancée, Iain Sanderson, for his love and encouragement.

PUBLICATIONS ARISING FROM THIS THESIS

Part of this publication arise from the work presented in chapter 1
Schornack, S., Huitema, E., **Cano, L.M.,** Bozkurt, T.O., Oliva, R., van Damme, M., Schwizer, S., Raffaele, S., Chaparro-Garcia, A., Farrer, R., Segretin, M.E., Bos, J., Haas, B.J., Zody, M.C., Nusbaum, C., Win, J., Thines, M., and Kamoun, S. 2009. Ten things to know about oomycete effectors. Molecular Plant Pathology, 10:795-803.

Part of this publication arise from the work presented in chapter 1
van Damme, M., **Cano, L.M**., Oliva, R., Schornack, S., Segretin, M.E., Kamoun, S., and Raffaele, S. EvolutionaryandFunctionalDynamics of Oomycete Effector Genes. Effectors in Plant–Microbe Interactions, First edition. (Accepted).

Part of this publication arise from the work presented in chapter 3
Oh, S.-K., Young, C., Lee, M., Oliva, R., Bozkurt, T., **Cano, L.M.,** Win, J., Bos, J.I.B., Liu, H.,-Y., van Damme, M., Morgan, W., Choi, D., van der Vossen, E.A.G., Vleeshouwers, V., and Kamoun, S. (2009). In planta expression screens of *Phytophthora infestans* RXLR effectors reveal diverse phenotypes, including activation of the Solanum bulbocastanum disease resistance protein Rpi-blb2. Plant Cell, 21:2028-2947.

Part of this publication arise from the work presented in chapter 4
Levesque, C.A., Brouwer, H., **Cano, L.,** Hamilton, J.P., Holt, C., Huitema, E., Raffaele, S., Robideau, G.P., Thines, M., Win, J., Zerillo, M.M., Beakes, G.W., Boore, J.L., Busam, D., Dumas, B., Ferriera, S., Fuerstenberg, S.I., Gachon, C.M., Gaulin, E., Govers, F., Grenville-Briggs, L., Horner, N., Hostetler, J., Jiang, R.H., Johnson, J., Krajaejun, T., Lin, H., Meijer, H.J., Moore, B., Morris, P., Phuntmart, V., Puiu, D., Shetty, J., Stajich, J.E., Tripathy, S., Wawra, S., van West, P., Whitty, B.R., Coutinho, P.M., Henrissat, B., Martin, F., Thomas, P.D., Tyler, B.M., De Vries, R.P., Kamoun, S., Yandell, M., Tisserat, N., Buell, C.R. (2010). Genome sequence of the necrotrophic plant pathogen, *Pythium ultimum*, reveals original pathogenicity mechanisms and effector repertoire. Genome Biology, 11:R73

Part of this publication arise from the work presented chapter 5 and 6
Haas, B.J., Kamoun, S., Zody, M.C., Jiang, R.H.Y., Handsaker, R.E., **Cano, L.M.,** et al. (2009). Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. Nature 461:393-398.

Part of this publication arise from the work presented in chapter 5
Raffaele*, S., Farrer*, R.A., **Cano*, L.M.,** Studholme, D.J., MacLean, D., Thines, M., Jiang, R.H.Y., Zody, M.C., Kunjeti, S.G., Donofrio, N.M., Meyers, B.C., Nusbaum, C., and Kamoun, S. (2010). Genome evolution following host jumps in the Irish potato famine pathogen lineage. Science, 330:1540-1543.

Part of this publication arise from the work presented in chapter 6
Gilroy, E.M., Breen, S., Whisson, S.C., Squires, J., Hein, I., Kaczmarek, M., Turnbull, D., Boevink, P.C., Lokossou, A., **Cano, L.M.,** Morales, J., Avrova, A.O., Pritchard, L., Randall, E., Lees, A., Govers, F., van West, P., Kamoun, S., Vleeshouwers, V.G.A.A., Cooke, D.E.L. and Birch, P.R.J. 2011. Presence/absence, differential expression and sequence polymorphisms between PiAVR2 and PiAVR2-like in *Phytophthora infestans* determine virulence on R2 plants. New Phytologist, 191:763-776

CHAPTER 1: General Introduction

1.1. Filamentous pathogens

1.1.1. Pathogenic oomycetes

1.1.1.1. Introduction

Oomycetes plant pathogens cause great economic losses of important crop species such as potato and tomato (Haas et al., 2009). These fungus-like eukaryotic microorganisms represent a distinct lineage (Kamoun, 2003), which are related to photosynthetic algae such as brown algae and diatoms (Baldauf, 2003, 2008). Among these, members of the genus *Phytophthora* and other well-known plant pathogens, such as downy mildews and *Pythium*, cause enormous economic losses on crop species (Haas et al., 2009). Some species, including the potato and tomato late blight agent *Phytophthora infestans* and the soybean root and stem rot agent *Phytophthora sojae* have caused long-standing problems for agriculture (Fry, 2008; Schmitthenner, 1985). More recent problems in agriculture are due to the epidemic outbreaks of oomycete pathogens like the fish pathogenic *Saprolegnia* species associated with salmonid saprolegniosis in Japan (Hussein and Hatai, 2002; Phillips et al., 2008; van West, 2006). Another example of epidemic diseases in potato and tomato are attributed to the emerging *P. infestans* genotypes 13_A2 and US22 that have caused high economic losses to farmers in the UK and in both USA and Canada, respectively (Chapman et al., 2010; Fry et al., 2009; Seidl et al., 2010; Vleeshouwers et al., 2011). Other significant oomycetes include the downy mildews, a heterogeneous and diverse group of obligate parasites (Agrios, 2005). Some downy mildews infect economically important hosts such as grapevines and sunflowers by *Plasmopara viticola and Plasmopara halstedii,* respectively (Hall, 1989; Hewitt and Pearson, 1988). *Hyaloperonospora arabidopsidis* is a natural pathogen of *Arabidopsis thaliana* and widely used in research on disease mechanisms in this model plant (Slusarenko and Schlaich, 2003).

Oomycetes can associate in different manners with their host plants. For example the *Arabidopsis thaliana* downy mildew *Hyaloperonospora arabidopsidis*, the white rust pathogen *Albugo laibachii* also pathogen of *A. thaliana* and the sunflower downy mildew *P. halstedii* are obligate biotrophs and rely on living plant tissue for growth and reproduction (Hall, 1989; Holub and Beynon, 1997; Slusarenko and Schlaich, 2003). In contrast, *Phytophthora* species are hemibiotrophic pathogens, which means that their life cycle alternates between two-step infection process: a “biotrophic” phase of infection followed and an extensive necrosis of host tissue associated with additional growth and sporulation (Fig. 1.1A) (Erwin and Ribeiro, 1996; Kamoun and Smart, 2005; Lee et al., 2006; Tyler, 2007). Within the genus *Pythium* spp. there is a diversity of life styles. Some *Pythium* spp. can behave as hemibiotrophs, similar to *Phytophthora* spp. or as necrotrophs causing rapid tissue damage and death (Bouwmeester et al., 2009).

A typical infection cycle for most plant parasitic oomycetes begins when zoospores encyst (although sporangia also initiate infections) and germinate on the plant surface. Subsequently, the germ tubes form an appressorium and a penetration peg that perforates the cuticle leading to the formation of haustoria. Haustoria site is of importance in host-pathogen interactions studies, as it plays a role in the delivery of effector proteins inside the host cell but it may also function as the site for nutrients uptake (Birch et al., 2008; Whisson et al., 2007). As the infection progresses, the plant tissue necrotizes and sporangiophores and sporangia develop usually through the stomata of leaves or the root surface to complete the life cycle (Fig. 1.1A).

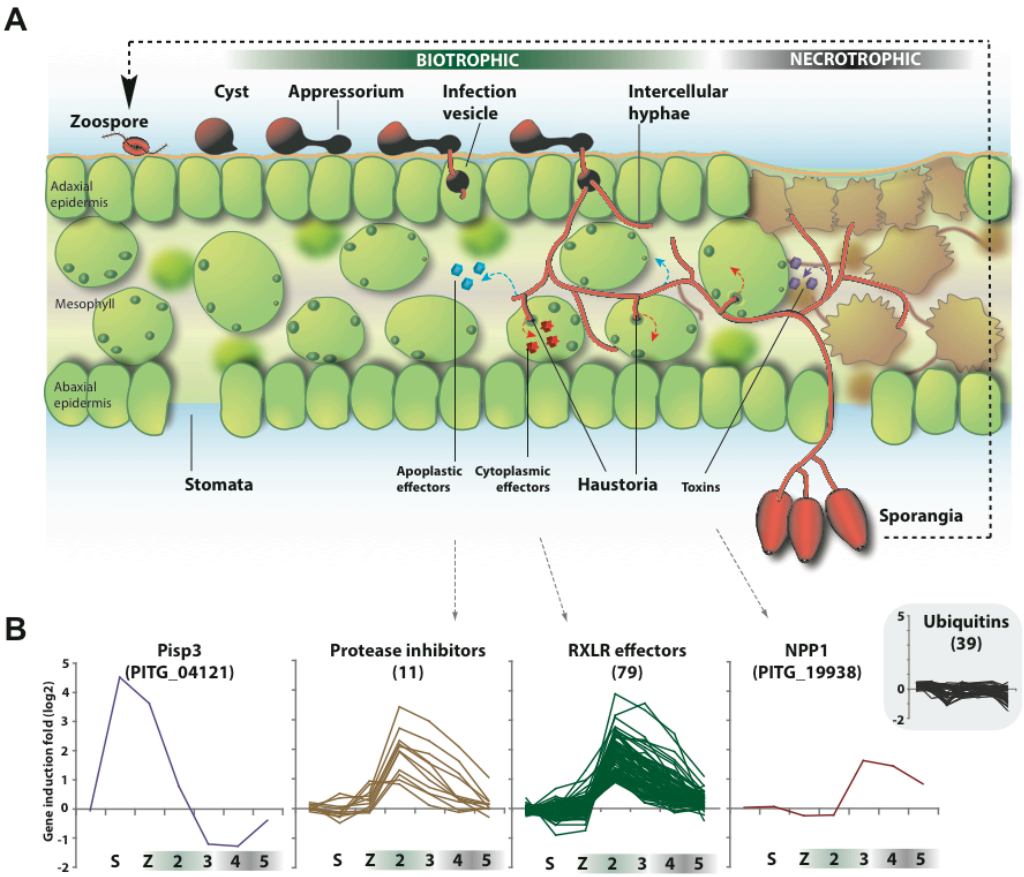


Fig. 1.1 The infection cycle of *Phytophthora infestans* stage-specific gene expression during hemibiotrophy
(A) The hemibiotrophic infection cycle of *P. infestans*. (B) Dynamic gene expression patterns in developmental and infection stages of *P. infestans*. The data are based on Haas et al. (Haas et al., 2009). Gene identifiers or description are shown with the number of genes indicated in parenthesis. S, sporangia; Z, zoospores; 2, 3, 4, and 5 are the days post inoculation with *P. infestans* strain T30–4 on potato. Mycelia were used as a baseline time point (first time point in the most left corner of the line graph) in each of the gene expression values of S, Z, and potato 2, 3, 4 and 5 dpi shown in the line graph (see microarray analysis in chapter 2, section 2.5.1). Figure by Sylvain Raffaele and Liliana Cano.

To established successful colonization, oomycete plant pathogens secrete an arsenal of molecules known as effectors (Birch et al., 2006; Hogenhout et al., 2009; Kamoun, 2007; Schornack et al., 2009; Stassen and Van den Ackerveken, 2011) and their main function is to perturb the host physiology and to repress plant immunity (Kamoun, 2007).

Nevertheless, effectors can be recognized in some plant genotypes by resistance (R) proteins, which are intracellular immune receptors of the nucleotide-binding

leucine-rich repeat (NB-LRR) family. The recognized effectors are called AVR proteins as they render the pathogen avirulent on plants that carry the cognate receptor (Morgan and Kamoun, 2007; Oh et al., 2009; van der Lee et al., 2001; Vleeshouwers et al., 2008; Whisson et al., 2001). Notable, all known AVR effectors belong to the RXLR effector class, whose RXLR motif is associated with translocation of the effectors inside the host cell (Allen et al., 2004; Armstrong et al., 2005; Champouret, 2010; Dou et al., 2008a; Halterman et al., 2010; Oh et al., 2009; Rehmany et al., 2005; Vleeshouwers et al., 2011; Vleeshouwers et al., 2008).

Several oomycete genomes are now available (three *Phytophthora* species, *Pythium ultimum*, *H. arabidopsidis* and *A. laibachii*) (Table 1.1) (Baxter et al., 2010; Haas et al., 2009; Kemen et al., 2011; Levesque et al., 2010; Tyler et al., 2006). Moreover, the coding sequences of five genomes representing four *Phytophthora* clade1c species (*P. infestans* PIC99189, *P. infestans* 90128, *Phytophthora ipomoeae*, *Phytophthora mirabilis* and *Phytophthora phaseoli*) are also available (Table 1.1) (see this thesis, chapter 5) (Raffaele et al., 2010a). The genome of the fish pathogen *Saprolegnia parasitica* has not yet been reported, but the sequences of coding genes are available at the Broad Institute website (http://www.broadinstitute.org/) (Torto-Alalibo et al., 2005). In addition, transcriptome sequences for the legume pathogen *A. euteiches* and sunflower downy mildew *P. halstedii* have been reported (Table 1.1) (Bouzidi et al., 2007; Gaulin et al., 2008).

This rich available dataset presents an excellent opportunity/tool for mining of novel effector candidates that carry conserved motifs like in RXLRs or other motifs like LFLAK from another class of host translocated effectors named CRNs (Haas et al., 2009). In addition, the above mentioned dataset could be use in comparative genomics studies that will lead to better understanding of genome structure and evolution of effector genes and the study of complex processes such as host adaptation or pathogenicity (Haas et al., 2009; Raffaele et al., 2010a; Raffaele et al., 2010b; Tyler et al., 2006).

Table 1.1. Genomic and transcriptomics resources of pathogenic oomycetes

Oomycete species	Host	Disease	Lifestyle	Genome size	No.
Genome resources					
Phytophthora infestans T30-4[a]	*Solanum* species (e.g. potato, tomato)	Late blight	Hemiobiotrophic	240Mb	18155 genes
Phytophthora infestans 90128[b]					
Phytophthora infestans PIC99189[b]					
Phytophthora infestans 06_3928A[c]					
Phytophthora ipomoeae PIC99167[b]	*Ipomoea logipedunculata*	Leaf blight			
Phytophthora mirabilis PIC99114[b]	*Mirabilis jalapa*	Leaf blight			
Phytophthora phaseoli F18[b]	*Phaseolus lunatus*	Downy Mildew			
Phytophthora sojae P649[d]	Soybean	Damping-off and root rot		95Mb	19027 genes
Phytophthora ramorum Pr-102[d]	Several trees and bushes (e.g. oak, rhododendron)	Sudden oak death, canopy dieback		65Mb	15743 genes
Pythium ultimum DAOM BR144[e]	Multiple dicots (e.g. potato) and monocots (e.g. turf grass)	Damping-off	Necrotrophic	42.8Mb	15290 genes
Hyaloperonospora arabidopsidis Emoy2[f]	Several brassicaceous plants including *Arabidopsis thaliana*	Downey mildew	Obligate biotrophic	100Mb	14543 genes
Albugo laibachii NC14[g]	Several brassicaceous plants including *Arabidopsis thaliana*	White rust	Obligate biotrophic	37Mb	14619 genes
Saprolegnia parasitica CBS223.65[h]	Fish (e.g. salmon, trout)	Saprolegniosis	Opportunistic, saprophytic and necrotrophic	53Mb	20113 genes
Transcriptome resources					-
Aphanomyces euteiches ATCC201684[i]	Several legumes, including peas, alfalfa, *Medicago truncatula* and clover	Root rot	Necrotrophic	-	7,977 uni-genes
Plasmopara halstedii race 300[j]	Asteraceae, including sunflower	Downy mildew	Obligate biotrophic	-	145 ESTs

[a] Reported by Haas et al., (Haas et al., 2009).
[b] Reported by Raffaele et al., (Raffaele et al., 2010a), and in this thesis (see chapter 5).
[c] Reported in this thesis (see chapter 6).
[d] Reported by Tyler et al., (Tyler et al., 2006); Jiang et al., (Jiang et al., 2008).
[e] Reported by Levesque et al., (Levesque et al., 2010).
[f] Reported by Baxter et al., (Baxter et al., 2010).
[g] Reported by Kemen et al., (Kemen et al., 2011).
[h] The genome data is available at http://www.broadinstitute.org/annotation/genome/Saprolegnia_parasitica/MultiHome.html and cDNA data was reported by Torto-Alalibo et al., (Torto-Alalibo et al., 2005)
[i] Reported by Gaulin et al., (Gaulin et al., 2008).
[j] Reported by Bouzidi et al., (Bouzidi et al., 2007).

1.1.1.2. Oomycete effectors target different sites in host plant tissue

Based on the plant compartment that oomycete effector proteins target, they can be classified into apoplastic effectors, which are present in the extracellular space, and cytoplasmic effectors, which are translocated into the cytoplasm of the plant cell where they can target different subcellular compartments (Kamoun, 2006, 2007). Seven classes of apoplastic effector and two classes of cytoplasmic effectors along with their distribution within the sequenced oomycete genomes are shown in Table 1.2.

Table 1.2. Major known classes of oomycete effectors

Effector class	Number of genes in the genomes of					
	Phytophthora infestans[a]	*Phytophthora sojae*[a]	*Pythophthora ramorum*[a]	*Pythium ultimum*[b]	*Hyaloperonospora arabidopsidis*[c]	*Albugo laibachii*[d]
Apoplastic effectors						
PcF/ SCRs	16	8	1	3	ND	ND
Protease inhibitors (serine and cystatin protease inhibitors)	41[e]	19	16	21[e]	5[e]	7[e]
NLPs	27	39	59	7	10	0
Elicitins	40	57	50	24	15	3
Proteases (Aspartyl, cysteine and serine proteases)	69	63	68	156	18	58
Cell wall degrading enzymes	198	241	216	209	>69	47
Lipases and phospholipases	55	58	45	51	ND	25
Cytoplasmic effectors						
RXLRs	563	335	309	0	134	49
CRNs	196	100	19	26	20	3

[a] Annotations reported by Tyler et al., (Tyler et al., 2006); Jiang et al., (Jiang et al., 2008); Haas et al., (Haas et al., 2009).
[b] Annotations reported by Levesque et al., (Levesque et al., 2010).
[c] Annotations reported by Baxter et al., (Baxter et al., 2010); ND, not documented.
[d] Annotations reported by Kemen et al., (Kemen et al., 2011); ND, not documented. Note that the method to predict RXLR effectors in *A. laibachii* is different to the method used in the RXLRs of the other oomycete genomes listed in Table 1.2 (Kemen et al., 2011).
[e] Annotations reported in this thesis (see chapter 4).

The first class of apoplastic effectors listed in Table 1.2 includes secreted small cysteine rich (SCR) proteins with similarity to a phytotoxin. For example, in *P. infestans* a member of this class is the effector gene *Scr74* that encodes a predicted 74-amino acid secreted cysteine rich protein with similarity to the *Phytophthora cactorum* phytotoxin PcF (Liu et al., 2005). Another class of apoplastic effectors listed in Table 1.2 includes protease inhibitors that function

as inhibitors of serine and cysteine proteases neutralizing plant defense (Tian et al., 2005; Tian et al., 2004; Tian and Kamoun, 2005; Tian et al., 2007). Some apoplastic effectors induce cell death *in planta* like Nep1-like proteins (NLPs), which contain a characteristic NPP domain, and have been identified in bacteria, fungi, and oomycetes (Gijzen and Nurnberger, 2006). A NLP from *P. infestans*, NPP1.1, triggers cell death in *Nicotiana benthamiana* and in tomato, probably functioning as a toxin during the necrotrophic phase of infection (Kanneganti et al., 2006). Another example is one of the elicitins, INF1 from *P. infestans*, a 10-KDa protein that also triggers cell death and defence response in plants (Chaparro-Garcia et al., 2011; Hann and Rathjen, 2007; Heese et al., 2007; Kamoun et al., 1997; Kawamura et al., 2009; Vleeshouwers et al., 2006).

There are two classes of host-translocated cytoplasmic effectors (shown in Table 1.2) that are classified based on conserved motifs in their N-termini. The RXLR effector family is characterized by an RXLR amino acid motif (arginine, any amino acid, leucine, arginine) (Birch et al., 2006; Morgan and Kamoun, 2007; Rehmany et al., 2005; Tyler et al., 2006). The CRN effector family contains a conserved LFLAK amino acid motif (leucine, phenylalanine, alanine, lysine) and induces a crinkling and necrosis phenotype when ectopic expression of the proteins in plants occurs, hence the name (Haas et al., 2009; Torto et al., 2003). It has been shown that both motifs are required for translocation of the effectors inside the cytoplasm of the host cell (Bhattacharjee et al., 2006; Dou et al., 2008b; Kamoun, 2007; Schornack et al., 2010; Whisson et al., 2007). More new motifs in the genomes of recently sequenced oomycete pathogens are being discovered, such as YxSL[RK] in candidate effectors from *P. ultimum* (Levesque et al., 2010). However, the functions of these motifs are still unknown and need to be experimentally determined.

1.1.1.3. Oomycete effectors have a modular architecture

Oomycete effectors have a modular architecture (Kamoun, 2006). Apoplastic effectors have an N-terminal signal peptide for secretion, followed by a C-terminal effector domain(s), but it is unknown whether they have an additional host targeting signal (Tian et al., 2005; Tian et al., 2004; Tian and Kamoun, 2005; Tian et al., 2007). The Kazal-like serine protease inhibitors occur in *Phytophthora, Pythium, Aphanomyces* and downy mildews (see detail annotation of protease inhibitors effectors in chapter 4) (Bouzidi et al., 2007; Gaulin et al., 2008; Haas et al., 2009; Levesque et al., 2010; Tyler et al., 2006). In *P. infestans*, Kazal-like serine protease inhibitors such as EPI1 and EPI10 inhibit the tomato subtilisin-like serine protease P69B (Tian et al., 2004). The inhibitory activity is restricted to the Kazal-like domain 1 (out of 2) in EPI1 and the Kazal-like domain 2 (out of 3) in EPI10 (Tian et al., 2005; Tian et al., 2004) (Fig. 1.2). This suggests that the different domains or modules within the effector may have different levels of specificity towards proteases. The targets or function of the other domains of EPI1 and EPI10 are still unknown.

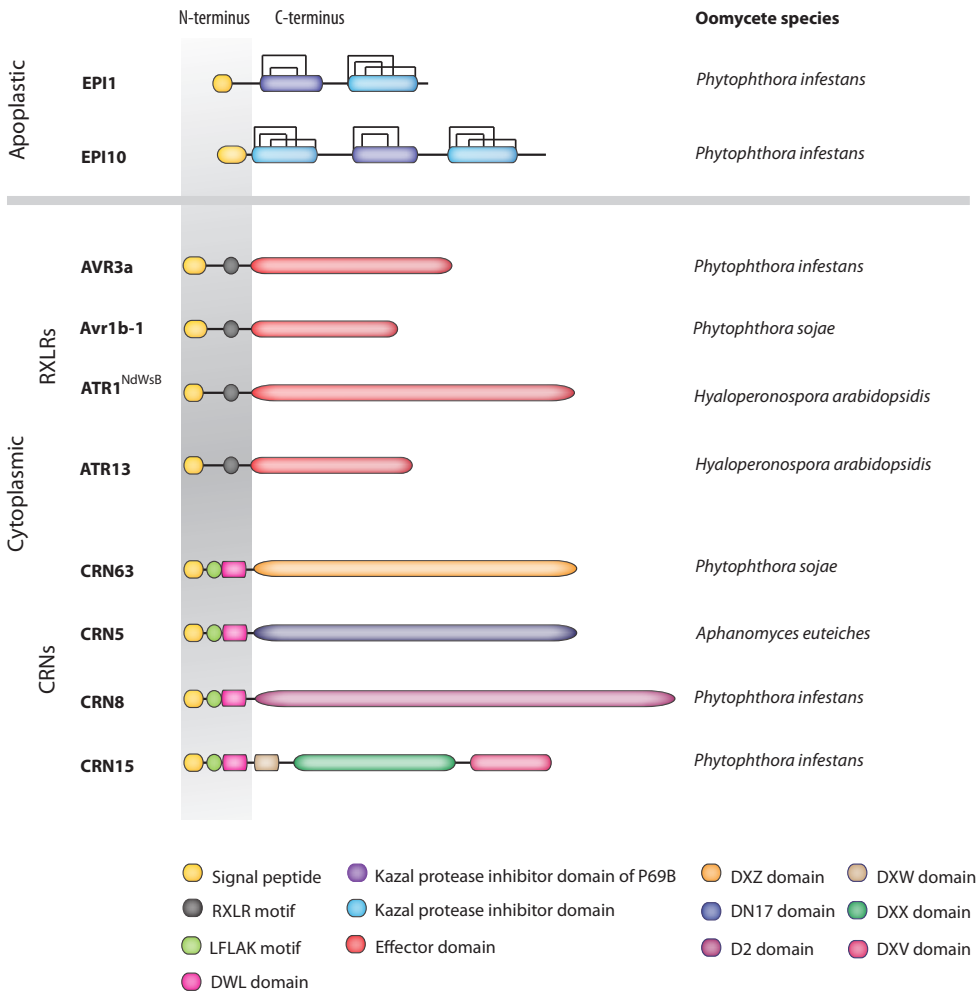


Fig. 1.2. Oomycete effectors are modular proteins
Illustration of the various functional modules forming some of the best-characterized classes of oomycete effectors. Two apoplastic effectors (EPI1 and EPI10) and eight cytoplasmic effectors (4 RXLRs; AVR3a, AVR1b-1, ATR1, ATR13 and 4 CRNs; CRN63, CRN5, CRN8, CRN15) from different oomycetes are illustrated. All modules are depicted by various patterns and the six different CRN C-terminal domains in white and named as identified and described by Haas et al. (Haas et al., 2009). Behind each protein the oomycete of origin is indicated. All four CRN structures also included a predicted NLS. Figure by Sylvain Raffaele, Mireille Van Damme and Liliana Cano.

The N-terminal domain of cytoplasmic effectors is associated with the translocation of the effector whereas the C-terminal domain is where the effector biochemical activity resides (Bos et al., 2006; Dou et al., 2008a; Dou et al., 2008b; Kamoun, 2006; Liu et al., 2011; Morgan and Kamoun, 2007; Oh et al., 2009; Schornack et al., 2010; Whisson et al., 2007). Additionally, some cytoplasmic effectors contain extra signals to target them to specific cellular compartments, for instance nuclear localization signals (Schornack et al., 2010).

A typical N-terminal domain of cytoplasmic effectors carries an RXLR motif after the signal peptide and this motif is very conserved and analogous to the PEXEL translocation motif of *Plasmodium* spp. (Bhattacharjee et al., 2006; Dou et al., 2008b; Grouffaud et al., 2008). In contrast, the C-terminal domain of RXLR effectors is highly polymorphic and shows signatures of positive selection, supporting the idea that this is the functional domain of the effector and that it is probably co-evolving with the host proteins (Fig. 1.3A, see also chapter 5 Fig. 5.2C, chapter 6 Fig. 6.3 and Fig. 6.4) (Allen et al., 2004; Jiang et al., 2008; Rehmany et al., 2005; Win et al., 2007). Examples of modular RXLR effectors are shown in Fig. 1.2. The other family of cytoplasmic effector described above, the CRN proteins also show a modular organization, including a signal peptide followed by the conserved LFLAK motif, and a diverse C-terminal domain (Haas et al., 2009). Interestingly, the LFLAK motif is also involved in translocation of the effector inside the host cell (Schornack et al., 2010). An DWL domain that ends with the HVLVXXP motif in most CRN proteins follows the LFLAK motif. The high degree of variability in the C-terminal domains of CRNs in the family is markedly found to be after the HVLVXXP motif that suggests a putative recombination point (Fig. 1.2) (Haas et al., 2009). Remarkably *in planta* expression of some CRN C-terminal domains can induce cell death (Haas et al., 2009; Schornack et al., 2010; Torto et al., 2003; Torto-Alalibo et al., 2007) or suppress it (Liu et al., 2011).

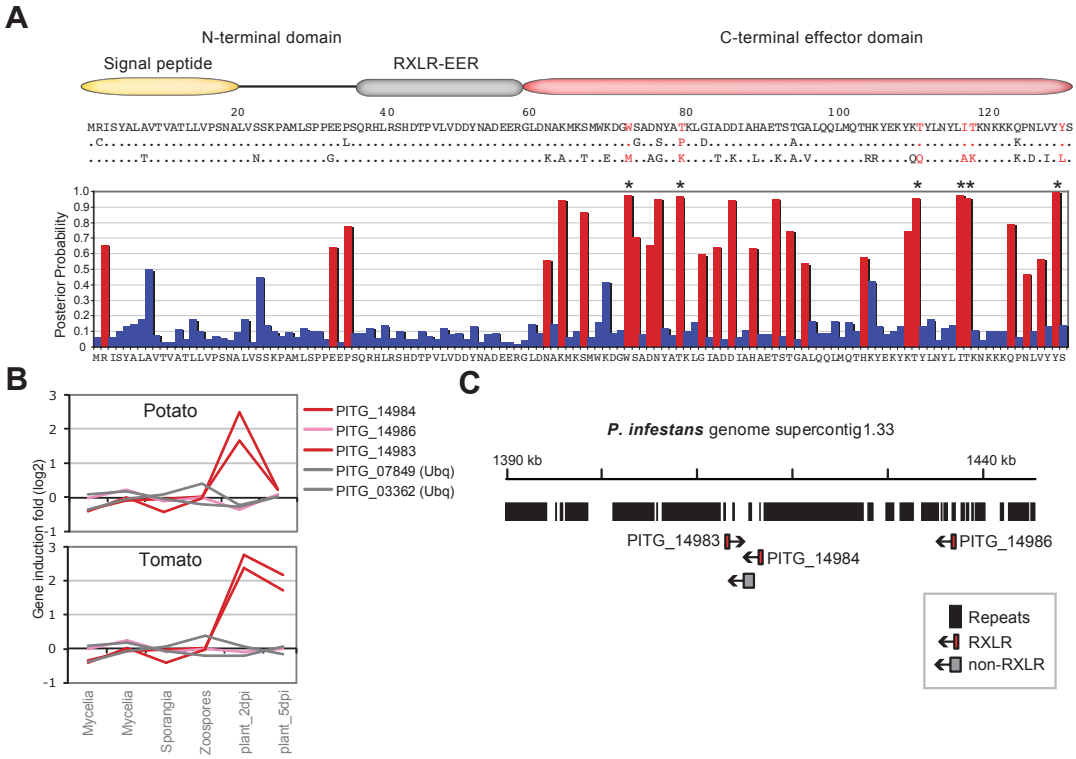


Fig. 1.3. RXLR effector genes typically show adaptive selection in their C-termini, are *in planta* induced and occur in the gene sparse repeat-rich regions
The figure depicts the features of a representative RXLR gene cluster (RXLR family 6) of *Phytophthora infestans* (Haas et al., 2009). (A) Domain structure and sequence variability of three paralogues RXLR effectors of *P. infestans* (PITG_14983, PITG_14984 and PITG_14986, top to bottom). Residues with evidence of positive selection are highlighted in red. Dots in the alignment represent identical amino acid residues. Positive selection analyses based on the methods described in Win et al. (Win et al., 2007) (see chapter 2 section 2.4.8). Posterior probabilities (blue, red) for the site class with expected ω value >1 (ω = 21.07706) and P = 0.16379 estimated under the model M8 in the PAML program (http://abacus.gene.ucl.ac.uk/software/paml.html). Positively selected sites are shown in red. Asterisks label residues with P > 95%. (B) Gene induction fold (log2) at different developmental stages during infection of potato and tomato plants 2 and 5 days post-inoculation (dpi) using mycelia as baseline (see microarray analysis in chapter 2 section 2.5.1). Two RXLR genes are induced *in planta* (red lines) and one is not (pink line). Two constitutive ubiquitin genes (Ubq) are shown as controls (grey lines). (C) Genome browser SybilLite view of ~55 Kbp region of the *P. infestans* genome (supercontig 1.33) containing the cluster of related RXLR genes locate in the gene sparse region (see chapter 2 section 2.3). The high content of repetitive sequences is evidenced by the presence of several black bars (repeats) (see chapter 2 section 2.3). Modified figure published in Schornack et al (Schornack et al., 2009). Figure by Sebastian Schornack and Liliana Cano.

1.1.1.4. Oomycete effector genes show distinct patterns of expression during plant colonization

The study of *P. infestans* gene expression during a time course of infection (potato, see *P. infestans* T30-4 Nimblegen data analysis in chapter 2 section 2.5.1) using a NimbleGen microarray (Haas et al., 2009) revealed distinct patterns of gene induction as the infection developed (Fig. 1.1B). The expression of most RXLR effector genes, including effectors with known avirulence activity (*Avr1, Avr2, Avr3a, Avr4, Avrblb1*, and *Avrblb2*) peaks during the biotrophic phase at 2 days post inoculation (dpi) (Fig. 1.1B, Fig. 1.3B, also see gene expression of 79 RXLR effectors of *P. infestans* T30-4 in chapter 3 Fig. 3.3 and appendix 1.1), and declines during the necrotrophic phase (4-5 dpi) (Haas et al., 2009; Vleeshouwers et al., 2011). Effector genes that belong to other families like protease inhibitors and cysteine-rich secreted (SCR) proteins exhibit similar induction peaks during biotrophy (Fig. 1.1B, also see gene induction patterns of 41 protease inhibitors in *P. infestans* T30-4 in chapter 4 Table 4.1). Interestingly, *PiNPP1* a gene encoding for a Nep1-like (NLP) cytolytic toxin, is up regulated during the transition from biotrophic to necrotrophic growth and remains induced during necrotrophy (Fig. 1.1B) (Haas et al., 2009; Kanneganti et al., 2006). This is consistent with the view that NLPs might be involved in the transition to the necrotrophic phase. In contrast to NLPs, RXLR effectors are mainly needed during the biotrophic phase and can function in the suppression of plant immunity (Gijzen and Nurnberger, 2006; Haas et al., 2009; Ottmann et al., 2009).

1.1.1.5. Effector genes populate plastic regions of oomycete genomes

P. infestans (Haas et al., 2009), *P. sojae* and *P. ramorum* (Tyler et al., 2006) represent three of the ten major phylogenetic clades of *Phytophthora* (Blair et al., 2008; Kroon et al., 2004). These species differ in a number of biological, genetic, and genomic features (Table 1.1) (Haas et al., 2009). The genome size diverges dramatically among them, ranging from 65 megabases (Mb) for *P. ramorum* to 240 Mb for *P. infestans* (Fig. 1.4A). Some effector families are expanded in *P. infestans* (Table 1.2, CRN effectors shown in Fig. 1.4A) but the dramatic genome

size difference cannot be explained by changes in gene content (Table 1.1). Instead, the expansion of the *P. infestans* genome has occurred through a proliferation of non-coding repeats as this species contains ~74% repeats versus <40% in the other two *Phytophthora* species (Gijzen, 2009).

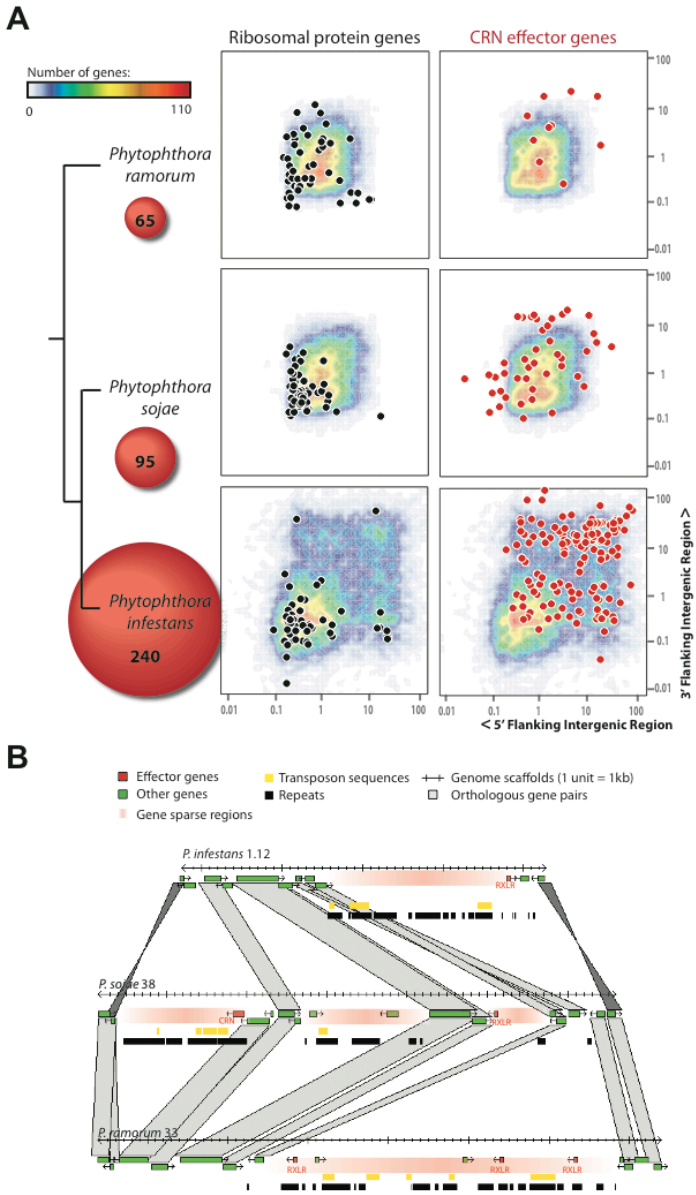


Fig. 1.4. Effector genes populate the repeat-rich expanded regions of *Phytophthora* genomes
(A) Genome organization and the distribution of core function genes (ribosomal protein genes) compared to effector genes (of the CRN family) in three *Phytophthora* species. Genome size for these three *Phytophthora* species is indicated under their name (as Mbp) and shown by a red circle of proportional diameter. Heat map diagrams show the distribution of genes according to the length of their flanking intergenic regions (in Kbp) as described in Haas et al., (Haas et al., 2009). Individual ribosomal protein and CRN

effector genes are shown over the heat map as dots. (B) *Phytophthora* genomes are formed of collinear blocks interrupted by repeat-rich regions. A 60 kb alignment window of the genomes of *Phytophthora infestans*, *Phytophthora sojae*, and *Phytophthora ramorum* showing collinear blocks separated by species-specific gene-sparse regions (GSR). Alignment window of *P. infestans*, *P. sojae*, and *P. ramorum* correspond to a snapshot from the genome browser SybilLite (see chapter 2 section 2.3). The GSRs contain the majority of the effector genes.

The three *Phytophthora* sequenced genomes share a core set of around 7000 genes that show 1:1:1 orthology among them (Haas et al., 2009; Tyler et al., 2006). These core orthologs are mainly housekeeping genes, including those involved in basic cellular processes like DNA replication, transcription, and protein translation (Haas et al., 2009). Nevertheless, the three genomes display a unique and conserved gene order in which regions that do not show such order separate the core genes. Interestingly, gene density is high in the conserved regions whereas the content of repeat and transposable elements (TE) content is low. In non-conserved regions, transposable elements are abundant, forming the so-called gene sparse regions (GSR) (Fig. 1.4B) (Raffaele et al., 2010b). The genome of *P. infestans* shows a more dramatic discontinuous distribution of gene density compared to the other genomes (Fig. 1.4A) (Raffaele et al., 2010b). Delimiting the GSR based on the length of intergenic DNA flanking genes in *P. infestans* showed that only 5% of known effector families are contained within gene-dense regions. This is in accordance with the fact that in eukaryote genomes genes encoding highly variable traits are hosted in plastic regions of the genome (Bustamante et al., 2005; Kosiol et al., 2008; Pain et al., 2008; van de Lagemaat et al., 2003; Volkman et al., 2007). The same holds true for virulence plasmids of *Yersinia pestis*, which are known to be in regions of high genome plasticity (Cornelis et al., 1998). Also, this is observed in *Phytophthora*, in which the RXLR, CRN and apoplastic effectors are predominantly in the GSR (Haas et al., 2009) (Fig. 1.4, Fig 1.3C). The distribution pattern of effector genes residing mainly in the GSR in *Phytophthora* is also described in other oomycete like *H. arabidopsidis* (Baxter et al., 2010). Without a doubt, this fact is a valuable tool in the search and identification of novel candidate effectors (see chapter 5) (Levesque et al., 2010; Raffaele et al., 2010b).

Darby and colleagues suggested that intracellular pathogens are often favored by reductions in their genome size since they are very well adapted to their stable

niche (Darby et al., 2007). Hence, a genome expansion do not ocurr frequently due to metabolic and replication costs (Cavalier-Smith, 2005). A great exception is *Phytophthora* with multiple genome expansions, driven perhaps by adaptations to a more changeable environment, for example the ever changing ability of host plants to develop resistance or become susceptible. This observation is consistent with previous comparative genomics analyses that revealed that *Phytophthora* effector genes have undergone accelerated patterns of birth and death evolution with evidence of extensive gene duplication and gene loss in the genomes of *P. infestans, P. sojae* and *P. ramorum* (Van Damme, unpublished) (Jiang et al., 2008; Jiang et al., 2006a; Qutob et al., 2009; Win et al., 2007) (see chapter 5 and chapter 6). In *P. infestans*, the RXLR and CRN effector gene families are among the most expanded relative to *P. sojae* and *P. ramorum* (Table 1.2) (Haas et al., 2009). Also, effector genes show patterns of positive selection with extensive nonsynonymous sequence substitutions, leading to high rates of amino acid polymorphisms (Fig. 1.3A, see chapter 5 and chapter 6) (Jiang et al., 2008; Liu et al., 2005; Oh et al., 2009; Qutob et al., 2009; Win et al., 2007).

1.1.1.6. Evolution of *Phytophthora infestans* genome and effector genes following host jumps

Host jumps followed by adaptation and specialization on distinct plant species play a major role in pathogen species evolution. This model of evolution has been reported notably for rust fungi (Roy, 2001), the anther smut fungi, *Microbotryum* spp. (Giraud et al., 2008), and in *Phytophthora* clade 1c species, which includes *P. infestans*, *P. mirabilis*, *P. ipomoeae*, and *P. phaseoli* (Blair et al., 2008). In Toluca Valley (Mexico), *P. infestans* naturally co-occurs with two very closely related species, *P. ipomoeae* and *P. mirabilis*, that specifically infect plants as diverse as morning glory (*Ipomoea longipedunculata*) and four-o'clock (*Mirabilis jalapa*), respectively. A fourth clade 1c species, *P. phaseoli*, infects lima beans (*Phaseolus lunatus*) in North America. There is no congruence between the phylogeny of the clade 1c species and their host plants indicating that these *Phytophthora* species evolved by host jump. Host jumps require the ability to

rapidly adapt to a change of environment (the host) and therefore are expected to have important consequences on the evolution of the genome and more specifically on the repertoire of effectors. Comparative genomic analysis of the four species from *Phytophthora* clade 1c demonstrated that faster evolution in the GSRs compared to the rest of the genome is a general feature within this lineage (see chapter 5) (Raffaele et al., 2010a).

1.1.1.7. Exploiting effectors in deployment of resistance

The identification of effector repertoire of plant pathogenic oomycetes is highly valuable for deployment of disease resistance against those oomycetes pathogens (Ellis et al., 2009; Vleeshouwers et al., 2008). Nowadays, having established *in planta* transient assays such as agroinfiltration (Van der Hoorn et al., 2000; Vleeshouwers et al., 2011; Vleeshouwers and Rietman, 2009), a set of effector proteins with unknown functions can be screened for avirulence activity on wild *Solanum* species (Vleeshouwers et al., 2011; Vleeshouwers et al., 2008). Also, effectors can be used in the screening of predicted resistance loci for a more efficient and less time consuming identification and cloning of the functional *R* gene from those loci (Vleeshouwers et al., 2011; Vleeshouwers et al., 2008).

1.1.2. Pathogenic rust fungi

Besides plant pathogenic oomycetes, rust fungi are also plant pathogens of economically agricultural crops. There are more than 120 genera and 6,000 rust fungi species that cause plant diseases in crops including coffee (e.g. *Hemileia vastatrix*), linola (e.g. *Melampsora lini*), wheat (e.g. *Puccinia graminis*), cowpea (e.g. *Uromyces vignae*), and beans (e.g. *Uromyces fabae*) (Aime et al., 2006; Cummins and Hiratsuka, 2004).

Rust fungi (Basidiomycetes of the order Uredinales) are obligate parasites of plants from which they obtain nutrients, live and reproduce in their host tissues. The majority of rust fungi are heteroecious which means that they require two phylogenetically distinct hosts to reproduce and complete their life cycles. During

infection in the host plant rust fungi form haustoria, which are specialized feeding structures within the host cell that can function in the acquisition of nutrients (Dodds et al., 2009). In addition to its role in nutrient acquisition, haustoria were proposed to function in delivery of effector molecules in the host cytoplasm (Dodds et al., 2009; Hogenhout et al., 2009; Oliva et al., 2010).

Rust fungi can cause a diversity of symptoms on their host plants and some fungal species exhibit extraordinary mimicry of plant flowers (Kaiser, 2006; Ngugi and Scherm, 2006; Roy, 1993a; Roy et al., 1998). *Puccinia* and *Uromyces* are two genera of rust fungi that modify host organs to produce flower-mimicking structures (pseudoflowers) to attract pollinators to enable gamete transfer and fertilization (Naef et al., 2002; Pfunder and Roy, 2000; Roy, 1993a).

1.1.2.1. Pseudoflower-forming rust fungus *Uromyces pisi*

The rust fungus *Uromyces pisi* (Pers.) species presents a heteroecious life cycle, which means that they need to alternate from *Euphorbia cyparissias* to another host member of the *Fabaceae* to complete their life cycle. Systemic infection of *E. cyparissias* by *U. pisi* inhibits flowering and results in pseudoflowers that mimic true plant flowers (Pfunder and Roy, 2000). *U. pisi* pseudoflowers are composed by arrangements of yellow leaves covered with gametes in a sweet-smelling fungal nectar that attract pollinators that bring together the two mating types and cross-fertilize the fungus (Pfunder and Roy, 2000). Pseudoflowers that occur together with true flowers exhibit shorter insect visits than those that occur alone, suggesting that true flowers might be competitors for pseudoflowers (Fig. 1.5A). The similarity of pseudoflowers to true flowers is proposed to be an adaptation in this system but further experiments are needed to evaluate this hypothesis (Pfunder and Roy, 2000).

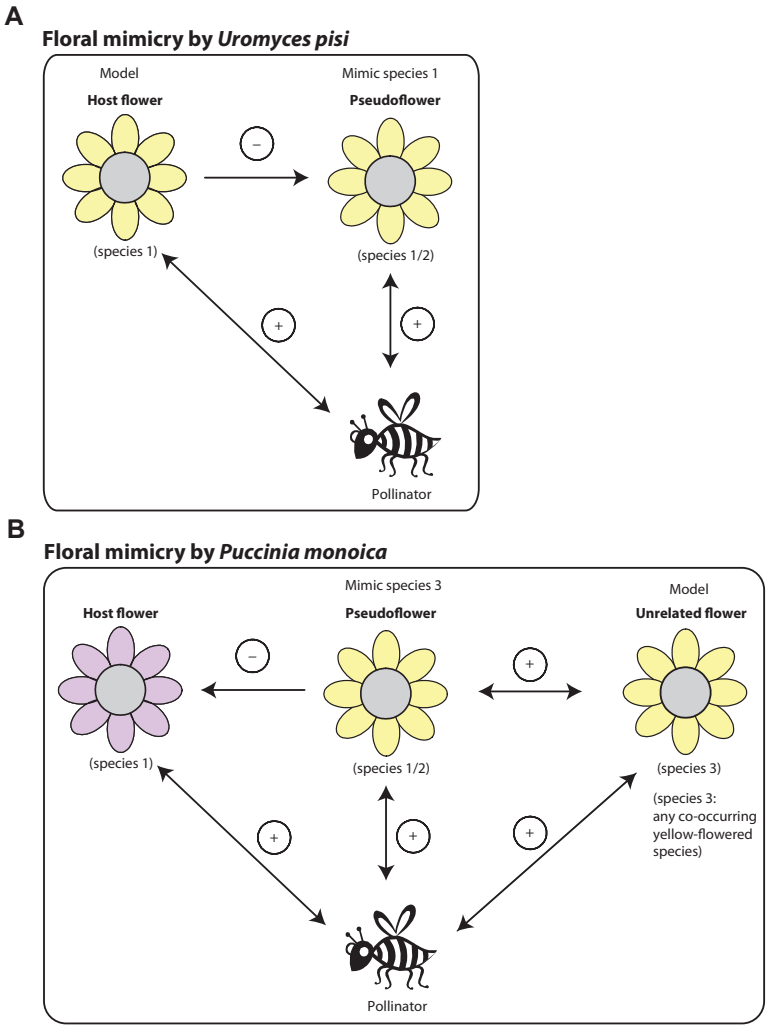


Fig. 1.5. Schematic illustration of floral mimicry in plant pathogenic rust fungi
(A) Floral mimicry by *Uromyces pisi* inhibits flowering in the host *Euphorbia cyparissias* (species 1) and mediates changes in the host morphology producing yellow pseudoflowers (species 1/2) that resemble the host flowers (species 1). Both true flowers and pseudoflowers are able to attract pollinators. However, shorter visits are observed on pseudoflowers in mixtures than monocultures, suggesting that true flowers might be competitors for pseudoflowers, this is indicated with a '-' sign on top of a unidirectional arrow. (B) Floral mimicry by *Puccinia monoica* inhibits the formation of flowers in the host *Boechera stricta* (species 1) that greatly affect host reproduction, this is indicated with a '-' sign on top of a unidirectional arrow and mediates changes in the host morphology producing pseudoflowers (species 1/2) that resemble unrelated flowers (species 3). Pseudoflowers can attract pollinators by itself and when they are together with other unrelated flowered plants have a positive effect receiving greater insect visitations, this indicated with a '+' sign on top of a bidirectional arrow. Open circles with '+' and '-' signs indicate positive and negative relationships, respectively. Graph modified from Ngugi and Scherm (Ngugi and Scherm, 2006).

1.1.2.2. Pseudoflower-forming rust fungus *Puccinia monoica*

Puccinia monoica (*Pucciniales, Basidiomycota*) is a rust fungus that possesses a heteroecious life cycle, alternating from *Boechera stricta* to an unknown host grass. *P. monoica* is a remarkable obligate biotroph pathogen that inhibits flowering in its host plant *B. stricta* and radically transforms host morphology to produce pseudoflowers which are flower-like leaves that mimic true flowers in shape, size, color and nectar production from unrelated plant species like the buttercup (*Ranunculus inamoenus* Greene) (Fig. 1.6) (Roy, 1993a, 1994). *P. monoica* pseudoflowers are efficient in attracting pollinators as they (i) contain as much or more sugar than co-occurring flowers; (ii) have a bright yellow color that functions as a visual cue to attract pollinators; (iii) exhibit long period of flying insect visits due to the sugary fluid rich in spermatia (spores of a single mating type) that attract pollinators; and (iv) release a distinct fragrance composed of aromatic alcohols, aldehydes and esters, which function as olfactory cues that can attract pollinators by itself, particularly flies (Roy, 1993a, 1994; Roy and Raguso, 1997). *P. monoica* pseudoflowers can negatively affect host reproduction and survival as they prevent the formation of true flowers (Fig. 1.5B) (Roy, 1993a, b, 1994).

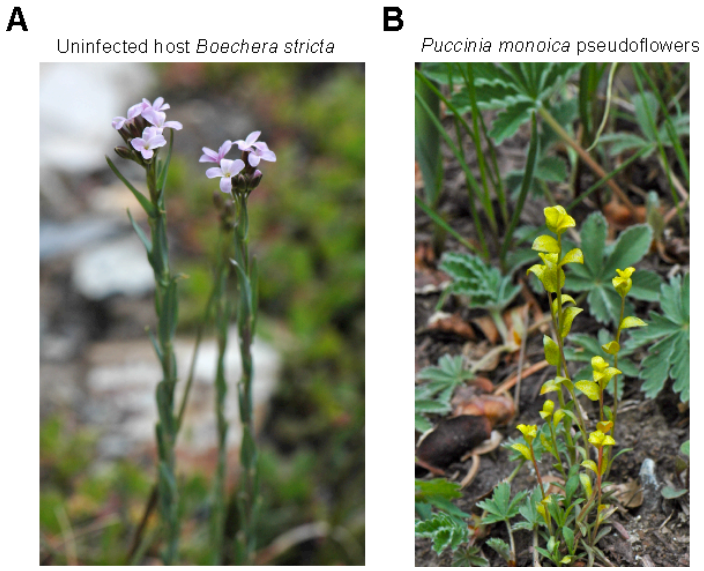


Fig. 1.6. Illustration of uninfected plant and infected plant with pseudoflowers
(A) Picture of uninfected flowering *Boechera stricta* plant. (B) Picture of infected *B. stricta* plant that produces pseudoflowers upon infection with *Puccinia monoica*.

The floral mimicry observed in *P. monoica*-induced pseudoflowers has similarities with those produced by the rust fungi *Uromyces pisi* in that they both are covered by sugary fluid rich in spermatia (spores) and release scents that attract pollinators (Pfunder and Roy, 2000; Roy and Raguso, 1997). Despite the similarities, floral mimicry in *P. monoica* differs from other rust fungi pathogens such as *U. pisi*. This is because *P. monoica* changes host morphology to produce pseudoflowers, which do not resemble the color of the flowers of its host like in *U. pisi* (Fig. 1.5 and Fig. 1.6) (Pfunder and Roy, 2000; Roy, 1993a).

1.1.2.2.1. *Boechera stricta*, the host of *Puccinia monoica*

The genus *Boechera* contains an array of morphologically and ecologically diverse taxa with highest diversity in western North America (Dobes et al., 2004). *B. stricta* Gray is the best-defined *Boechera* species; it is predominantly diploid with 7 chromosomes, sexual, highly self-fertilizing and most accessions form of a monophyletic group, making it a good system for ecological genomic studies (Dobes et al., 2004; Kantama et al., 2007; Schranz et al., 2005). Comparative analyses of *B. stricta* and its close relative *Arabidopsis thaliana* revealed that at least 9000 non-redundant sequences of *B. stricta* have highly significant similarity to annotated coding sequenced of *A. thaliana* (Windsor et al., 2006). The conservation among coding genes between *B. stricta* and *A. thaliana* suggests that *A. thaliana* can be used in genetic studies of *B. stricta*, for example gene expression profiling using existing *A. thaliana* arrays (Schranz et al., 2007; Windsor et al., 2006).

1.2. Aims of this thesis

The main objectives of this thesis were 1) to provide insights of the evolution of filamentous plant pathogen effectors using comparative genomics and 2) to understand the molecular changes produced by these effectors in the host plant using microarray analysis of the host-pathogen interaction.

Filamentous plant pathogens such as oomycetes secrete an arsenal of effector proteins to modulate host innate immunity and enable parasitic infection (Birch et al., 2006; Chisholm et al., 2006; de Jonge et al., 2011; Kamoun, 2006, 2009; O'Connell and Panstruga, 2006; Oliva et al., 2010; Schornack et al., 2009). In the oomycete pathogen *Phytophthora infestans* hundreds of RXLR effectors can be predicted to be secreted using SignalPv2.0 program (Haas et al., 2009; Raffaele et al., 2010b). In addition, it is known that all Avirulence proteins (AVRs) of *P. infestans* carry secretion signals prior the RXLR motifs, therefore the detection and characterization of these secretion signals is a pre-requisite for the discovery of putative candidate AVR effectors (Kamoun, 2007, 2009; Schornack et al., 2009). In Chapter 3, my objective was to demonstrate that predicted secretion signals of *P. infestans* RXLR effectors are functional and to highlight the importance of these secretion signals in the identification of candidate effectors using high-throughput computational methods. In Chapter 3, I used a genetic assay called Signal Sequence Trap (SST) to validate these computationally predicted pathogen secretion signal peptides, based on the requirement of yeast cells for invertase secretion to grow on sucrose or raffinose media. I showed that signal peptides of four representatives *in planta*-induced RXLR effector genes of *P. infestans* are functional and that the predictions obtained with the SignalPv2.0 program are accurate (Jacobs et al., 1997; Klein et al., 1996; Lee et al., 2006; Menne et al., 2000; Oh et al., 2009; Schneider and Fechner, 2004).

In Chapter 4, my objective was to provide clues of the evolution of two apoplastic protease inhibitors effectors families in pathogenic oomycetes using comparative genomics. In Chapter 4, I annotated the protease inhibitor effector repertoires in various recently sequenced oomycete pathogens (*P. infestans*, *P. ultimum*, *S. parasitica, H. arabidopsidis* and *A. laibachii)* and confirmed that protease inhibitors Kazal-like and cystatin domains are conserved in various oomycete pathogens. I also exploited the information from the microarray time course of *P. infestans* during infection on potato and tomato to investigate the gene expression profiles in the two protease inhibitor gene families: Kazal-like and cystatin-like. I found that many of protease inhibitor genes in *P. infestans* are induced *in planta* implicating them in virulence.

In Chapter 5, my objective was to provide insights in how host adaptation affects genome evolution of closely related filamentous plant pathogens, particularly oomycetes. To do this I studied the patterns and selective forces that shape sequence variation in the *P. infestans* clade1c species that have adapted to unrelated host plants (Grunwald and Flier, 2005). I showed three main findings in Chapter 5: 1) The *P. infestans* genome exhibits a “two-speed” pattern of organization, with gene-sparse repeat rich regions (GSR) experiencing accelerate rates of evolution; 2) gene sparse regions are also highly enriched in *in planta*-induced genes; 3) within the gene sparse regions there are at least 65 fast-evolving protein families, including effectors (Raffaele et al., 2010a). All together, these findings suggest that the 2–speed genome organization of the *P. infestans* clade1c species complex favors genome plasticity that is driven by selective forces in order to adapt to the new host. Moreover the 2–speed genome organization also favors the plasticity of effectors genes contained in the repeat-rich regions.

In Chapter 6, my objective was to identify which effectors molecules and which alterations of these effectors (in structure, sequence and expression) are important determinants of the aggressiveness and virulence reported in emerging plant pathogen strains. In Chapter 6, I studied an aggressive clonal lineage of the oomycete pathogen *P. infestans* termed 13_A2 that had emerged in the UK in 2007 and has since it had dominated and displaced other populations of the pathogen. Importantly, 13_A2 isolates have the ability to infect a wider spectrum of resistant potato cultivars than other *P. infestans* isolates (David Cooke, unpublished). To determine the effector gene repertoire and unravel other genetic features of 13_A2, in Chapter 6, I performed genome analyses (genome sequencing and microarray expression profiling) of a representative isolate *P. infestans* 06_3928A from 13_A2. I found that 06_3928A exhibits significant genetic and expression polymorphisms in effectors genes. 06_3928A also carries intact *Avrblb1, Avrblb2* and *Avrvnt1* effector genes that are induced *in planta*. Consistent with these results, 06_3928A cannot infect potato lines that carry the corresponding *R* immune receptor genes *Rpi-blb1*, *Rpi-blb2*, *Rpi-vnt1.1*. These findings point to a genetic strategy for mitigating the impact of 13_A2 epidemics

and illustrate how pathogen genome analysis can benefit the management of a devastating plant disease epidemic.

In Chapter 7, my initial objective was provide clues in the understanding of how *P. monoica* rust fungus can inhibits host flowering in its host *Boechera stricta* and how can modify host plant leaves to produce instead “pseudoflowers” to promote its own reproduction (Roy, 1993a). Initially, I aimed to discover pathogen effectors of the remarkable rust fungus using genomics. However, due to limitations of DNA material of this obligate biotroph in early times of Illumina sequencing, I have focused my study in Chapter 7 in the molecular changes produced in the host *B. stricta* during the interaction with *P. monoica*. To do this, in Chapter 7, I used a whole-genome microarray expression profiling to study pseudoflowers (pathogen-host interaction) and identified biological processes that are significantly perturbed (differentially regulated) in infected *B. stricta* plants by *P. monoica*. These results suggest that formation of pseudoflowers involves extensive reprogramming of the host including alteration of flower, shoot and leaf development, cell wall and cell surface modifications, and volatiles synthesis.

CHAPTER 2: Materials and Methods

2.1. Yeast Signal Sequence Trap System (SST)

2.1.1. Fusion of signal peptides to invertase in pSUC2 vector

Signal peptides of RXLR effectors were predicted using SignalPv2.0 program (Nielsen et al., 1997) with a HHM signal peptide probability of 0.9 or higher (Torto et al., 2003). In addition to SignalPv2.0 predictions, sequences that contained putative transmembrane domains (TM) predicted with TMHMM program (Krogh et al., 2001) were filtered out. Then I used the yeast Signal Sequence Trap (SST) system based on vector pSUC2T7M13ORI (pSUC2), which carries a truncated invertase gene, SUC2, lacking both the initiation methionine (Met) and signal peptide (SP) (Fig. 2.1) (Jacobs et al., 1997). DNA fragments coding for the signal peptides and the following two amino acids of PexRD6/IpiO, PexRD8, PexRD39, and PexRD40 were codon optimized for expression in yeast using OPTIMIZER program (Puigbo et al., 2007) and synthesized by GenScript and introduced into pSUC2 using EcoRI and XhoI restriction sites to create in-frame fusions to the invertase (Fig. 2.1, Table 2.1, see appendix 1.2).

Table 2.1. PexRD signal peptide sequences fused to invertase in the pSUC2 vector

PexRD protein	Signal peptide probability HMM model*	S-mean probability NN model*	SP length (aa)*	Signal peptide fused to invertase
PexRD6, AVRblb1	1.000	0.968	21	MRSLLLTVLLNLVVLLATTGAVSSNL…
PexRD8	0.990	0.860	22	MRLSCVYLVVATVTTIIASANAAAEAS…
PexRD39, AVRblb2	1.000	0.864	23	MRSFLYGVLAFAVLARSSAVAAFPIPD…
PexRD40, AVRblb2	1.000	0.864	22	MRSCLYGILAFAVLARSSAVAAFPIPD…

* Probability values were predicted using SignalPv2.0 http://www.cbs.dtu.dk/services/SignalP-2.0/.

2.1.1.1. Transformation of yeast cells

The invertase negative yeast *Saccharomyces cerevisiae* strain YTK12 (Jacobs et al., 1997) was transformed with 20 ng of each one of the pSUC2-derived

plasmids individually using a modified Lithium Acetate (LiAc/TE) method (Fig. 2.1) (Gietz et al., 1995).

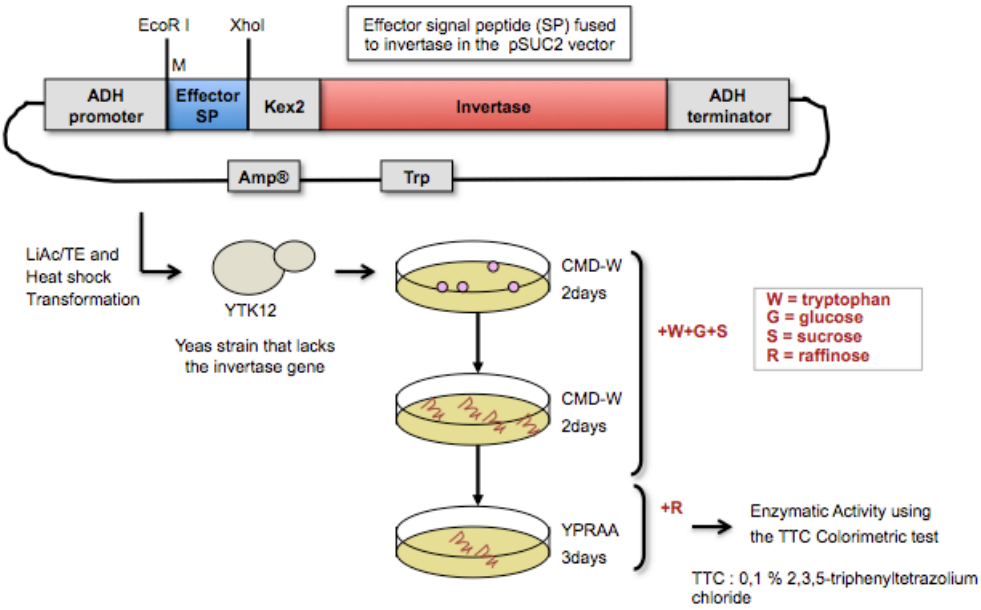


Fig. 2.1. Schematic diagram showing the Yeast Signal Sequence Trap System (SST)
Transformation with LiAc/TE method was modified from the previously established LiAc/SS-DNA/PEG method (see chapter 2 section 2.1.1.1). *Saccharomyces cerevisiae* YTK12 is a negative invertase yeast strain and pSUC2 vector carries a truncated invertase gene, SUC2 that lacks both the initiation methionine (Met) and signal peptide (Jacobs et al., 1997). In the SST method you will use methionine and signal peptide sequence of your gene of interest to fused to pSUC2 and transformed in yeast cells in order to test secretion of invertase in yeast. Only yeast transformed cells that carry the pSUC2 vector will grow in plates containing CMD-W media as these cells will carry pSUC2 selective marker tryptophan (Trp, W) (see media preparation in chapter 2 section 2.1.1.1.2.1). After transformed yeast cells are re-streaked in plates containing YPRAA, only transformed cells carrying the signal peptide of your gene of interest will grow as these cells will need to secrete the invertase to degrade the complex sugar raffinose (formed by glucose, fructose and galactose), produce glucose and grow from this media (see media preparation in chapter 2 section 2.1.1.1.2.1).

2.1.1.1.1. Preparation of competent yeast cells

1. Early in the morning inoculate 100 ml pre-warmed YPDA (see media preparation in section 2.1.1.1.2.1) with the pre-culture to an OD600 0.08-0.1.
2. Incubate at 28ºC and 180-200 rpm until OD600 0.5-0.6.

Note: Minimal incubation time is at least the time necessary for 2-3 duplications.

3. Place 25 ml cell culture in sterile 4 tubes (50 ml falcon tubes). Then centrifuged at 2500 rpm for 5 min at 20ºC.
4. Remove the medium without disturbing the cell pellet, re-suspend cells in 5 ml sterile distilled water each and re-centrifuged again as described above.
5. Remove the distilled water from the tube, resuspend cells in 2.5 ml LiAc/TE (see media preparation in chapter 2 section 2.1.1.1.2.1), finally pool the suspensions together and mix carefully.
6. Centrifuged at 2500 rpm for 5 min at 20ºC and remove the supernatant. NOTE: To freeze the competent cells you can add 0.5 volumes of Freeze Solution (see media preparation in chapter 2 section 2.1.1.1.2.1) and centrifuged at 2500 rpm for 5 min at 20ºC and remove the supernatant. Resuspend cell pellet in 0.2 ml of Freeze Solution (see media preparation in chapter 2 section 2.1.1.1.2.1) and slowly freeze in liquid N_2 and stored at -80ºC.
7. Resuspend the cell pellet carefully on 0.8-1 ml LiAc/TE (see media preparation in chapter 2 section 2.1.1.1.2.1) and transfer to a sterile 1.5 ml eppendorf tube. Incubate suspension for 30 min at room temperature.
Note 1: If competent cells are going to be used in a period of time longer than 30 min but less than 2 hours, it is recommended to keep the tube(s) at 4ºC.
Note 2: A starting volume was 100 ml of cell culture would regularly produce a final volume of 1000 µl of competent cell culture. Because each transformation reaction requires 200 µl of competent cells, the total product of 1000 µl competent cell culture will serve for approximately for 20 transformation reactions.

2.1.1.1.2. Yeast transformation protocol

1. Before starting boil the single strand DNA (ssDNA) (Sigma Cat No. 31149) for 3 min and chill it on ice immediately.
2. For each transformation add reagents indicated below in order:
20 µl carrier-ssDNA (2 mg/ml),
20 µl of DNA mix (2 µl of plasmid plus insert (construct that contains methionine and signal peptide at 10 ng/ µl) and 18 µl of 1x TE buffer and mix well)

4.5 µl of 1 M LiAc (see media preparation in chapter 2 section 2.1.1.1.2.1) and mix well,
50 µl competent cells and mix well,
300 µl of PEG/LiAc (see media preparation in chapter 2 section 2.1.1.1.2.1) and mix well.
3. Incubate shaking for 20 min at 30ºC (thermo-mixer).
4. Heat shock in 42ºC water-bath for 20 min.
5. Centrifuged at 6000-8000 rpm for 1 min and carefully remove the supernatant with a pipette.
6. Carefully resuspend the cell pellet on 100 µl sterile distilled water (or sterile 1x TE buffer) with the pipette.
7. Streak the transformation on selective media selective media CMD-W (media minus tryptophan (W)) and grow for 3 days at 30ºC on (see media preparation in chapter 2 section 2.1.1.1.2.1).
Note 1. If there is condensation water in media in petri dishes that contain the selective media, it is recommended to let plates dry for 5 min in the flood hood in advance.
Note 2: Be aware that the yeast transformation efficiency tends to be high so make sure to resuspend cells (step 6 above) in at least 1 ml of sterile distilled water and streak out no more than 100 µl in each plate. Colonies are often visible after 2 days at 30ºC.
8. Transfer by streaking at least 5 colonies separately to plates with the same CMD-W media and incubate plates at 30ºC for at least 2 days.
9. Transfer by streaking growing cells from CMD-W to raffinose media YPRAA (yeast peptone raffinose with antimycin media) (see media preparation in chapter 2 section 2.1.1.1.2.1) for another 2-3 days at 30ºC.

2.1.1.1.2.1. Solutions used for yeast transformation

LiAc/TE

1 ml 10x TE

1 ml 1M LiAc

8 ml ddH20

Final volume of 10 ml

PEG/LiAc mix

0.5 ml 10x TE

0.5 ml 1M LiAc

4 ml 50% PEG

Final volume of 5 ml

1 M LiAc

LiAc in ddH20, pH not adjustable, sterilize by filtering Millipore filter units, 0.22 µM

10x TE

100 mM Tris HCl

10 mM EDTA, pH 7.5, adjust with NaOH

50% PEG 3350 or PEG 4000

Dissolve PEG in small volume of distilled water and mix carefully, heat in the microwave for 1-2 min to homogenize the mixture, sterilize by autoclaving.

Note: This solution should be prepared right before it is needed and the remaining solution should be discard.

5 mg/ml salmon sperm DNA (ssDNA) in 1x TE

Prepare several aliquots and keep at -20ºC

YPDA media (final volume of 400 ml), for general yeast culture

20 g YPD

8 mg Adenosine hemisulphate salt

(For plates add 8 g agar, 2 %)

Selective media CMD-W (final volume of 400 ml)

0.67% (w/v) yeast nitrogen base without amino acids 2.68 g

0.075% (w/v) -Trp dropout supplement 0.3 g

2.0% (w/v) agar 8 g

Add after autoclaved

0.1% (w/v) glucose 2 ml from the 20% stock solution

2.0% (w/v) saccharose/sucrose 20 ml from the 40% stock solution

Stock of sugar solutions

16 g sucrose in 40 ml distilled H_20 (40%) filtered using filter unit Millipore (0.45 µm)

8 g sucrose in 40 ml distilled H_20 (20%) filtered using filter unit Millipore (0.45 µm)

Raffinose media YPRAA (final Volume of 400 ml)

1% (w/v) yeast extract 2 g

2% (w/v) peptone 8 g

2% (w/v) raffinose 8 g

2% (w/v) agar 8 g

50 mg antimycin A (Sigma Cat No. A8674) in 1 ml (stock at 50 mg/ml)

Freezing Competent cells using freeze solution

To prepare 40 ml of Freeze Solution

1 M Sorbitol 20 ml from 2 M Sorbitol stock solution

10 mM Bicine 0.8 ml from 0.5 M Bicine-HCl stock solution

3% ethylenglicol 12 ml 12 ml from 10% solution

5% DMSO 2 ml

Stock solutions for the preparation of the freeze solution

2 M Sorbitol 7.28 g in 20 ml dH_20, autoclaved

0.5 M Bicine-HCl 1.63 g in 20 ml distilled H_20 (adjust pH at 8.35), autoclaved

10% ethylenglicol 2 g in 20 ml distilled H_20, autoclaved

2.1.2. Screening for positive yeast transformant colonies using selective media

After transformation, yeast was plated on CMD-W (minus Trp, W) plates (0.67% yeast N base without amino acids, 0.075% Trp (W) dropout supplement, 2% sucrose, 0.1% glucose, and 2% agar) (see media preparation in section 2.1.1.1.2.1). Transformed colonies were transferred to fresh CMD-W plates and incubated at 30°C for 2 days. To assay for invertase secretion, colonies were replica plated on YPRAA plates (1% yeast extract, 2% peptone, 2% raffinose, and 2 mg/ml antimicyn A) (see media preparation in chapter 2 section 2.1.1.1.2.1) containing raffinose and lacking glucose. Also, invertase enzymatic activity was detected by the reduction of 2,3,5-Triphenyltetrazolium Chloride (TTC) to insoluble red colored 1,3,5-Triphenylformazan (TPF) as follows (Klotz, 2004). Five milliliter of sucrose media were inoculated with the transformed yeast cells and incubated for 24 h at 30°C. Then, cell pellet was collected, washed, and resuspended in distilled sterile water, and an aliquot was incubated at 35°C for 35 min with 0.1% of the colorless dye 0,1% 2,3,5-Triphenyltetrazolium Chloride (TTC) (BD Difco™, Cat. No. 231121). Colorimetric change from TTC to TPF was checked after 5 min incubation at room temperature.

2.1.3. PCR validation of yeast transformant colonies

Transformed colonies were picked and resuspended on 50 µl distilled sterile water in a 0.6 ml Eppendorf tube, then cell solution was lysed by boiling for 3 min at 94°C. After spin down for 30 s, an aliquot of 5 µl from the supernatant was used to confirm the transformation status by PCR with pSUC2 vector-specific primers (pSUC2-F: GGTGTGAAGTGGACCAAAGGTCTA and pSUC2-R: CCTCGTCATTGTTCTCGTTCCCTT) (Jacobs et al., 1997). PCR reactions were carried out on 50 µl reaction volume using a Primus 96plus Thermalcycler

(MWG-Biotech, Ebersberg, Germany). Each reaction contained 1 × GoTaq® Flexi buffer, 1.5 mM MgCl2, 100 µM dNTPs, 0.5 unit of Taq polymerase (GoTaq® DNA polymerase, Promega), 0.4 µM of primers and 5 µl of template yeast cell lysate. Amplification conditions consisted of one cycle of 94°C for 4 min, 30 cycles of 94°C for 20 s, 60°C for 20 s, 72°C for 45 s and a final cycle of 72°C for 5 min. PCR product aliquots of 10 µl were loaded in 1% agarose gels with 100 bp DNA ladder (Fermentas Cat No. SM0243) and pictures were taken under UV light with digital imaging system gel doc (BioRad).

2.2. Sequence analysis of protease inhibitors

Signal peptides of protease inhibitor effector families were predicted using SignalP v2.0 program (Nielsen et al., 1997) with a HHM signal peptide probability of 0.9 or higher (Torto et al., 2003). In addition to signalP predictions, sequences that contained putative transmembrane domains (TM) predicted with TMHMM program (Krogh et al., 2001) were filtered out. Sequence analysis was done using NCBI BLAST sequence similarity search (blastall) programs, with low complexity filter on. Protease inhibitor domains were predicted with interproscan (http://www.ebi.ac.uk/Tools/pfa/iprscan/). DNA Strider was used for ORF sequence search. ClustalX (1.83.1) (Thompson et al., 2002) was used for multiple sequence alignments of nucleotide and protein sequences. Phylogenetic analysis were conducted using the neighbor-joining method in MEGA5 (Tamura et al., 2011). Bootstrap values equal or greater than 50% from 1000 replicate trees are shown at the nodes. Horizontal branch lengths and scale bar correspond to evolutionary distances assigned by MEGA5. The evolutionary distances are measured as the proportion of nucleotide substitutions between sequences (Tamura et al., 2011).

2.3. Web genome browser resources used in this study

To visualize specific regions of the genome of *Phytophthora infestans* that show are syntenic with genomes of other two *Phytophthora* species *(P. infestans, Phytophthora sojae* and *Phytophthora ramorum)* I used a customized genome

web browser based on SybilLite that includes constructed by Brian Haas. *P. infestans* SybilLite custom genome browser is not freely available on the web but could be made available to others upon email request to Brian Haas at the Broad Institute (http://www.broadinstitute.org). SybilLite is based on Sybil a web-based software package for comparative genomics that can be downloaded at http://sybil.sourceforge.net/. Screen shots images of genomic regions from *P. infestans* SybilLite genome browser were generated using the Mac OS X 10.5 application Grab.

2.4. Genome analyses of *Phytophthora* species

2.4.1. List of *Phytophthora* species used in this study

I generated Illumina reads from the genomic DNA of three *Phytophthora infestans* isolates: PIC99189, 90128, 06_3928A, the reference strain T30-4. Also, I generated genomic DNA sequence data from *Phytophthora ipomoeae* PIC99167, *Phytophthora mirabilis* PIC99114 and *Phytophthora phaseoli* F18 to complement data obtained from collaborator laboratories from Broad Institute (Table 2.2)

Table 2.2. Illumina genome sequenced *Phytophthora* species and used in this study

Phytophthora spp.	Strain	Host	Country of Origin	Year of isolation	Reference	Estimated genome coverage
Phytophthora infestans	T30-4	*Solanum spp.*	The Netherlands	1995	(Drenth et al., 1995)	10.5x
Phytophthora infestans	PIC99189	*Solanum spp.*	Mexico	1999	(Flier et al., 2002)	10.4x
Phytophthora infestans	90128	*Solanum spp.*	The Netherlands	1990	(Vleeshouwers et al., 1999)	17.1x
Phytophthora infestans	06_3928A	*Solanum spp.*	United Kingdom	2006	(Cooke and Cano, unpublished)	57.9x
Phytophthora ipomoeae	PIC99167	*Ipomoea purpurea, Ipomoea longipedunculata*	Mexico	1999	(Flier et al., 2002)	12.5x
Phytophthora mirabilis	PIC99114	*Mirabilis jalapa*	Mexico	1999	(Flier et al., 2002)	11.0x
Phytophthora phaseoli	F18, Race F	*Phaseolus lunatus*	United States	2000	(Evans et al., 2002)	9.0x

Phytophthora infestans reference genome strain T30-4 (Haas et al., 2009) was re-sequenced in this study to validate SNP calling in other *Phytophthora* species (see section 2.4.7).

A set *P. infestans* strains, characterized for their Multilocus Genotype (MLG), available at The Hutton Institute (Scotland) was used for PCR validation of assembled RXLRs events (data provided by David Cooke, Hutton Institute, Scotland). The evaluated set contains a group of 19 MLGs with a total of 44 strains including the sequenced strain 06_3928A and the reference genome strain T30-4 (MLG set contains: strains 2006_3928A, 2008_7038_A, 2008_6250A, 2008_6430A, 2008_6194A, 2006_4132B, 2008_6530C, 2008_6102A, 2006_3964A, 2008_6082F, 07_39, 07_5242A, 08_6422C, 2005_15094, 2006_4144C, 2006_3884B, 2005_14473, 2006_3924C from 13_A2 MLG; strains 2006_4440C and 2006_3936C2 from 10_A2 MLG; strain 2006_4012F from 3_A2 MLG; strain 2006_4244E from 3b_A2 MLG; strains 2006_3984C and 2006_4304A from 1_A1 MLG; strains 2006_3888A, 2006_3960A and 2006_4068B from 2_A1 MLG; strain 2006_4352E from 4_A1 MLG; strain 1996_9_5_1_C4 from 5_A1; strain 07_5866C from 5g_A1; strains 2006_4100A and 2006_3920A from 6_A1; strains 2006_4168B and 2006_4168C from 7_A1; strain 2006_4232C from 8_2a_A1; strain 2006_4256B from 8a_A1; strain 2006_4320F from 12_A1; strain 2004_7804B from 15_A2; strain 2006_3992G from 16_A2; strain 2006_4388D from 17_A2; strains 2003_25_1_3 and 2003_25_3_1 from 22_A2; strain 07_sp12_3A and T30-4 from Misc).

2.4.2. Library preparation and sequencing

For the genomic DNA extraction, *Phytophthora infestans* strains T30-4, PIC99189, 90128 and 06_3928A, *Phytophthora ipomoeae* PIC99167, *Phytophthora mirabilis* PIC99114 and *Phytophthora phaseoli* F18 were cultured in Rye Sucrose Agar (RSA) plates at 18°C for 12 days. Plugs with mycelium were transferred to modified Plich medium, grown for another two weeks at 18°C and then harvested for genomic DNA isolation using OmniPrep™ kit (G-Biosciences, Cat No. 786-136) with minor modifications.

For sequencing, the flow cells were prepared according to the manufacturer's instructions using Illumina single end cluster generation kit FC-103-2001 or Illumina pair read cluster generation kit PE-203-4001. Sequencing reactions were performed mostly on 2G GAs (Illumina Inc.).

The reference genome sequence of *P. infestans* strain T30-4, annotation and gene/exon locations was downloaded from www.broad.mit.edu (GenBank project accession number AATU01000000).

2.4.3. Alignment of reads to the reference genome

I generated Illumina single-end read sequence data for *Phytophthora infestans* T30-4, PIC99189 and 90128, *Phytophthora ipomoeae* PIC99167, *Phytophthora mirabilis* PIC99114 and *Phytophthora phaseoli* F18 and Illumina pair-end read sequence data for *P. infestans* 06_3928A.

The generated single-end reads were aligned to the genome using Mapping and assemblies with qualities (MAQ) software package v0.6.8 (Li et al., 2008b) using default parameters. Lanes with >0.06 error rates based on the assigned MAQ mapping quality statistics were excluded from the datasets.

The generated raw reads with abnormal lengths and reads containing Ns were removed from the datasets. Filtered reads were used to align to the reference genome strain T30-4. The filtered pair-end read data was aligned using the Burrows-Wheeler Transform alignment (BWA) software package v0.5.7 (Li and Durbin, 2009) using as parameters: seed length (l) of 38 and a maximum of mismatches (M) allowed of 3.

2.4.4. *De novo* assembly of unmapped reads

I extracted 8,722,383 unmapped pair reads of isolate 06_3928A using a homemade script (Table 6.1). Unmapped reads were assembled using velvet package v1.0.18 with a Kmer of 53, a minimum contig length of 200 bp

nucleotides and an insertion length of 300 bp as parameters. I obtained 15,654 contigs with a N50 of 359 bp, a mean size of 367 bp and a median size of 278 bp. The smaller contig in the assembly have a size 201 bp and the largest contig a size of 6,286 bp. The obtained contigs are equivalent to 5.4 Mb in size. Then, all 15,654 contigs were mapped back to the reference genome strain T30-4 using NUCmer program, included in MUMmer 3.2 package (Kurtz et al., 2004). A total of 9,838 contigs equivalent to 2.77 Mb of the assembly showed hits to T30-4 and were filtered out of the assembly. The remaining 5,816 contigs were kept for the next steps of the analysis of the genes encoding proteins, which included prediction of secretion signals and RXLR motifs.

2.4.5. Prediction of secreted proteins and RXLR motif from assembled contigs

A total of 5,816 assembled contigs were translated to amino acids using a homemade script and to predicted signal peptides with SignalP v2.0 program (Nielsen et al., 1997) and putative transmembrane domains with TMHMM (Krogh et al., 2001) program. Secreted proteins were selected when showing no transmembrane domains and a SignalP HMM score cutoff of > 0.9 and NN cleavage site within 10 and 40 amino acids. Secreted proteins were predicted to contain RXLR motifs when: RXLR position was present within 30 and 60 amino acids, RXLR position was higher than NN cleavage site and signal peptide length was <= 30 (Torto et al., 2003; Win et al., 2007). The RXLR prediction resulted in the identification of six candidate RXLR effectors in the isolate 06_3928A.

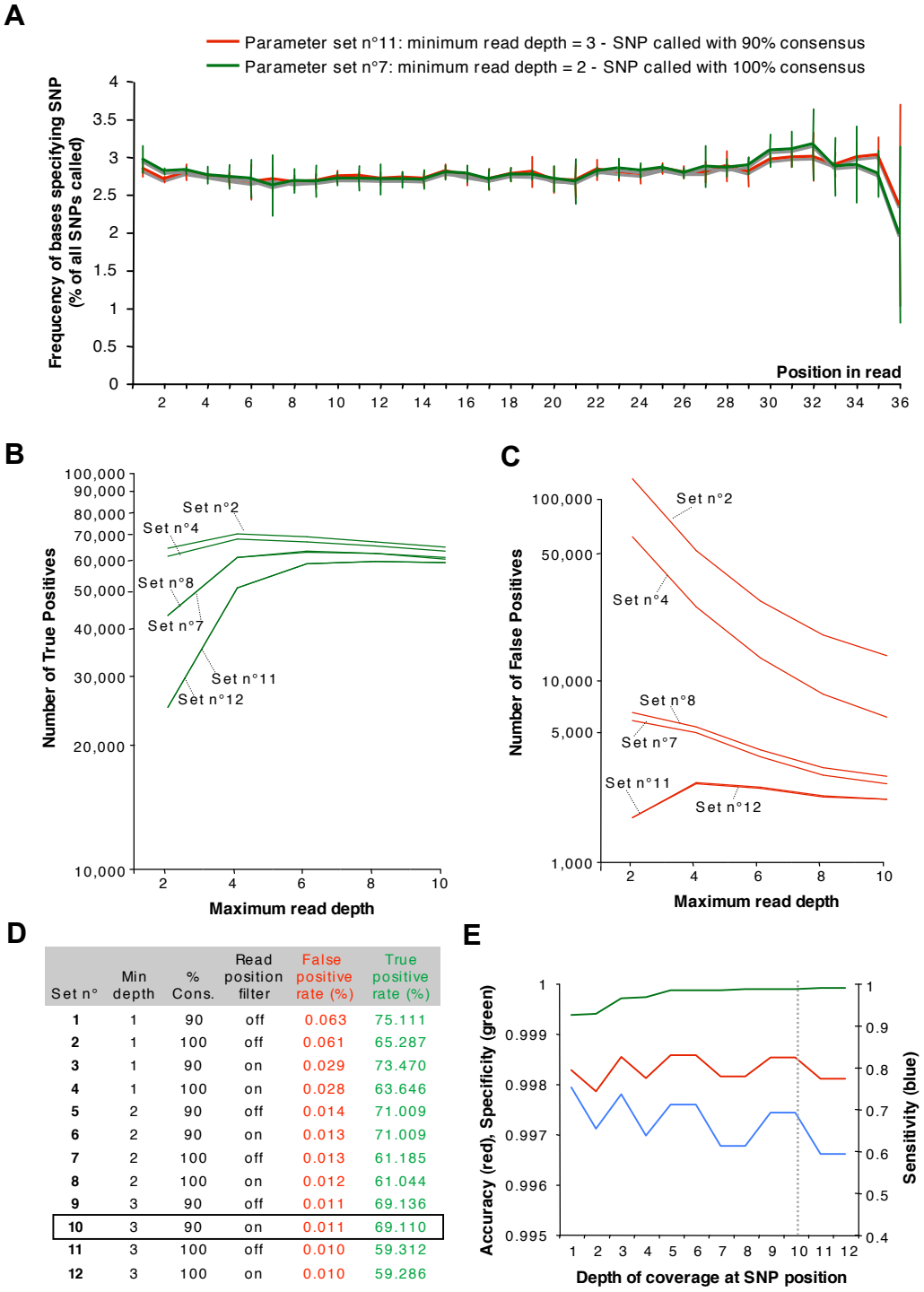
2.4.6. PCR validation of candidate assembled RXLR genes

Assembled RXLRs were validated by PCR amplification of genomic DNA on 20 µl reaction volume using a Primus 96 plus Thermalcycler (MWG-Biotech, Ebersberg, Germany) (data provided by David Cooke, Hutton Institute, Scotland). Specific primers were used for the amplification of six candidate assembled RXLRs genes: *Pex644, Pex50259, Pex30588, Pex46622, Pex15083 and Pex14182* with an expected size of 514, 258, 257, 365, 472 and 859 bp

respectively (*Pex644_F*: TGAGTGGAATCGCATCAGTAGT, *Pex644_R*: ATCCTCTGCCTTTTTAATCTGAC, *Pex50259_F*: TGGCAAGGTAAACGCTCTCT, *Pex50259_R*: GAGGCCGATAAGTCGTCAAC, *Pex30588_F*: TTTCTGTGATGCTGCCTCTG, *Pex30588_R*: CGTCAAACTTGTTAAGGTTTTGC, *Pex46622_F*: ATGCGTATCTCGCAAGCT, *Pex46622_R*: TCATACGTGATCATCGGAGA, *Pex15083_F*: ACGCTTCTATCCGACAACGA, *Pex15083_R*: ATTGGTGGTAATGCCTGCG, *Pex14182_F*: ATGCGTGGCGTCGAAACTA, *Pex14182_R*: CCATTGGCTGATACGGTATTT). Each reaction contained 1 × GoTaq® Flexi buffer, 20 µg BSA, 1.5 mM MgCl2, 100 µM dNTPs, 0.8 unit of Taq polymerase (GoTaq® DNA polymerase, Promega), 0.2 µM of primers and 20 ng of template DNA. Amplification conditions consisted of one cycle of 94°C for 4 min, 30 cycles of 94°C for 20 s, 60°C for 20 s, 72°C for 45 s and a final cycle of 72°C for 5 min.

2.4.7. Optimization of SNP calling parameters

The frequency of bases specifying SNPs (as % of all SNPs detected) for position 1 to 36 along Illumina reads were determined with 2 sets of parameters, a SNP being called when (i) position is covered at least twice and 100% of reads specify a SNP (green) or (ii) position is covered at least 3 times and 90% of reads specify a SNP (red) (Fig. 2.2A). SNPs are uniformly called from all positions on the reads with the 2 sets of parameters, except for a bias for SNPs being called by lower quality bases at the last position of the reads. In the following analysis, the above bias was eliminated by considering SNPs called by at least 1 read on a position < 33 (called “read position filter” hereafter).



Set n°	Min depth	% Cons.	Read position filter	False positive rate (%)	True positive rate (%)
1	1	90	off	0.063	75.111
2	1	100	off	0.061	65.287
3	1	90	on	0.029	73.470
4	1	100	on	0.028	63.646
5	2	90	off	0.014	71.009
6	2	90	on	0.013	71.009
7	2	100	off	0.013	61.185
8	2	100	on	0.012	61.044
9	3	90	off	0.011	69.136
10	3	90	on	0.011	69.110
11	3	100	off	0.010	59.312
12	3	100	on	0.010	59.286

Fig. 2.2. Optimization of Single Nucleotide Polymorphism (SNP) calling method by False Discovery Rate (FDR) analysis in the re-sequenced *P. infestans* T30-4
(A) Frequency of bases specifying a SNP as a function of position in reads. Values shown are average for the 6 re-sequenced genomes using Illumina single end reads and expressed as a % of all SNPs called in a genome. Error bars show standard deviation between genomes. Parameter sets assayed refer to (D). (B and C) Number of SNP correctly detected (True Positives, B) and called by error (False Positives, C) after 100,000 SNP were computationally introduced in the *P. infestans* T30-4 reference

genome. The SNP calls are shown as a function of maximum read depth. Six parameter sets refer to D. (D) Summary table of false and true positive rates obtained with 12 parameter sets in a 10x deep genome sequence of *P. infestans* T30-4. (E) Accuracy, specificity and sensitivity of FDR test obtained with the 12-parameter sets tested in 10x-covered genome. Dotted line highlights parameter set chosen for subsequent analyses. Min., minimum; % cons, consensus percentage.

A False Discovery Rate (FDR) analysis was used to determine the performances for SNP calling in *P. infestans* T30-4 genome relative to the amount of data generated. FDR for the SNP calling methods were calculated by randomly introducing 100,000 SNPs into the coding sequences of the reference genome, aligning re-sequenced *P. infestans* T30-4 singe end reads to the 'modified' reference. Performances were evaluated with the number of introduced SNPs called back (true positives, Fig. 2.2B) and the numbers of SNPs called that were not introduced (false positives, Fig. 2.2C). Six parameter sets (Fig. 2.2D) were tested as a function of depth of coverage at SNP position, artificially limited to 2, 4, 6, 8 or 10. A FDR analysis in *P. infestans* T30-4 re-sequenced genome at a depth of coverage of 10x was used with 12 different parameters sets to optimize the detection of SNPs (Fig. 2.2E). Accuracy was defined as (TP + TN)/(TP + FP + FN + TN), specificity as TN/(TN + FP) and sensitivity as TP/(TP + FN) where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives and FN is the number of false negatives. A 90% consensus among reads calling a SNP with a minimum of 3x coverage is the final set of parameters selected for the following analyses (n°10, highlighted in Fig. 2.2D and 2.2E). A total of 746,744 SNPs were detected in the re-sequenced species (Fig. 5.1).

A False Discovery Rate (FDR) analysis was used to determine the performances for SNP calling in 90% identical genome regions of *P. infestans* 06_3928A isolate. FDR for the SNP calling methods were calculated by randomly introducing 100,000 SNPs into the coding sequences of the 90% identical genome regions of 06_3928A genome, aligning re-sequenced 06_3928A pair end reads to the 'modified' reference.

A total of 54 parameter sets (Fig. 2.3A) were tested as a function of (i) a minimum read depth of coverage at SNP position, artificially limited to 3, 4, 5, 6,

7, 8, 9, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30 and (ii) percentage of reads specify a SNP, also artificially limited to (80, 90, 95). For FDR analysis I measured: accuracy, specificity and sensitivity. A 90% consensus among reads calling a SNP with a minimum of 10x coverage is the final set of parameters selected for the following analyses (final parameter set highlighted with an arrow in Fig. 2.3A and marked with a dashed line in Fig. 2.3B). A total of 22,523 SNPs were detected in the re-sequenced species (Table 6.2).

A

Min Depth	% Cons.	False Positives (%)	True Positives (%)
3	80	0.01309	89.40497
4	80	0.01306	89.08481
5	80	0.01301	88.71160
6	80	0.01297	88.35726
7	80	0.01295	88.03096
8	80	0.01293	87.71908
9	80	0.01291	87.39771
10	80	0.01290	87.04936
12	80	0.01285	86.41784
14	80	0.01279	85.79365
16	80	0.01275	85.24296
18	80	0.01265	84.62325
20	80	0.01261	84.03207
22	80	0.01256	83.46555
24	80	0.01252	82.93276
26	80	0.01248	82.36322
28	80	0.01242	81.77808
30	80	0.01237	81.25368
3	90	0.01255	87.76172
4	90	0.01251	87.44224
5	90	0.01249	87.14974
6	90	0.01247	86.86975
7	90	0.01245	86.62821
8	90	0.01242	86.38521
9	90	0.01241	86.14026
10	90	0.01240	85.82179
12	90	0.01236	85.25586
14	90	0.01231	84.68314
16	90	0.01229	84.18760
18	90	0.01221	83.63598
20	90	0.01217	83.09963
22	90	0.01214	82.56850
24	90	0.01211	82.05986
26	90	0.01207	81.51310
28	90	0.01202	80.96014
30	90	0.01197	80.46071
3	95	0.01192	82.66312
4	95	0.01189	82.34517
5	95	0.01186	82.05394
6	95	0.01184	81.77511
7	95	0.01183	81.53478
8	95	0.01180	81.29285
9	95	0.01179	81.04901
10	95	0.01178	80.79072
12	95	0.01174	80.34777
14	95	0.01170	79.91448
16	95	0.01169	79.56508
18	95	0.01162	79.15060
20	95	0.01159	78.73451
22	95	0.01156	78.26415
24	95	0.01153	77.80932
26	95	0.01149	77.31357
28	95	0.01144	76.80926
30	95	0.01140	76.36571

B

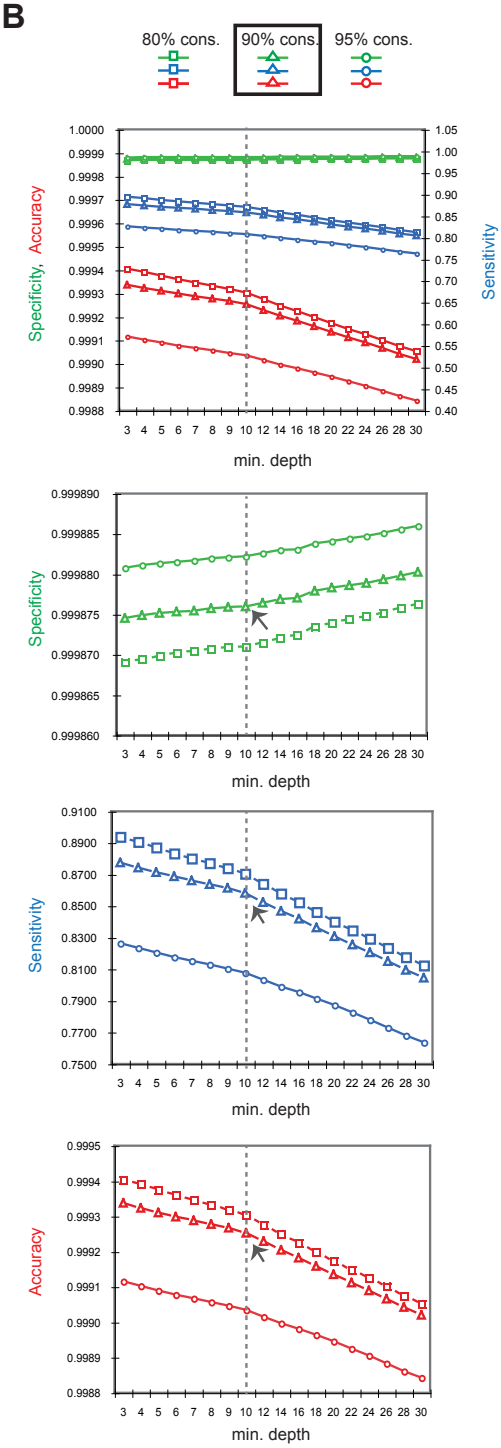


Fig. 2.3. Optimization of Single Nucleotide Polymorphism (SNP) calling method by False Discovery Rate (FDR) analysis in the re-sequenced *P. infestans* 06_3928A isolate

(A) Summary of false and true positive rates obtained with 54 parameter sets in the sequenced genome of *P. infestans* isolate 06_3928A. Percentage of consensus bases chosen is presented with the black square box. Arrow point the minimum read depth chosen. (B) Accuracy, Specificity and Sensitivity of FDR test obtained with 54 parameter

sets tested in 58x coverage genome. Dotted line highlights parameter set chosen for subsequent analyses. Arrow indicates the point that defines the minimum read depth that allows detecting SNPs with 99.92% of accuracy and 85.82% of sensitivity.

2.4.8. Substitution rates and dN/dS ratio

In comparisons of only two gene sequences (gene 1 of specie 1 *vs.* gene 1 from reference specie 2, e.g. *Phytophthora mirabilis* gene 1 *vs*. *Phytophthora infestans* T30-4 gene 1), I estimated the rates of synonymous substitution (dS), nonsynonymous substitution (dN) and omega (dN/dS) using the yn00 program of PAML (Yang, 2007) by implementing the Yang and Nielson method (Yang and Nielsen, 2000). In other instances, where I performed comparison of more tan two gene sequences, I calculated the dN/dS and the posterior probabilities for the amino acid sites being under positive selection under the model M8 as reported by Win et al, (Win et al., 2007) using the codeml program of PAML (Yang, 2007).

2.4.9. Synonymous and nonsynonymous SNP distribution along the N and C-terminal domains of RXLR effector genes

Differences in frequencies of nonsynonymous minus synonymous SNPs were counted per 15 bp-long windows and sliding by 3 bp. Frequencies were calculated as the number of SNPs per bp per gene and averaged over 20 consecutive windows. A total of 118, 3,077 and 2,442 genes were considered in the analysis by having at least one SNP in the RXLRs, core orthologs and GDR gene groups respectively. Numbers of SNPs in RXLR gene domains were counted per 15 bp-long windows and sliding by 3 bp. The 20 windows adjacent to the RXLR motif were considered for each of the domains. In total, 118 RXLR genes having at least one SNP were analyzed.

2.4.10. Breadth of coverage and presence/absence polymorphisms

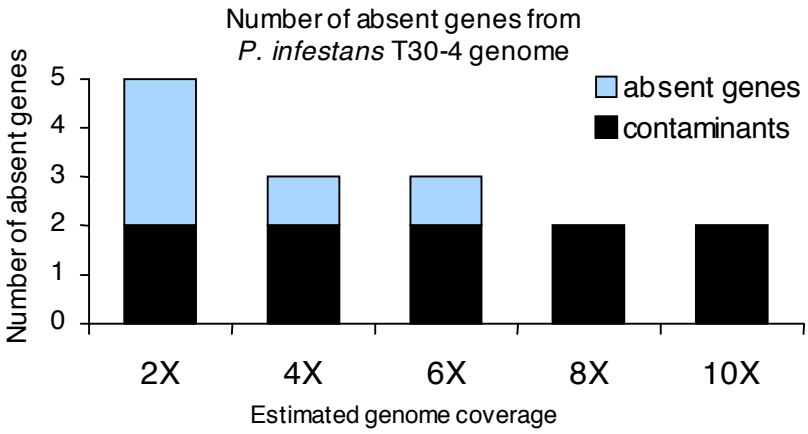
Breadth of coverage was calculated for each of the 18,155 genes as the percentage of nucleotides with at least one read aligned. Genes were considered

absent conservatively when breadth equals 0. We simulated the number of absent genes (Fig. 2.4B) and the average breadth of coverage per gene over the genome (Fig. 2.4C) in *P. infestans* T30-4 re-sequenced strain for average genome coverage 2x, 4x, 6x, 8x and 10x. To avoid any possible flow cell biases, we used random subsets of the full dataset yielding the average genome coverage. Increasing the number of reads increases the breadth of coverage over genes. By 8x coverage, genes were covered >99% in average (Fig. 2.4B). All genes were highly covered and only contaminant genes were identified as absent (Fig. 2.4C). Genes were called absent when breadth of coverage equalled 0 in re-sequenced strain. Two genes (independently identified as bacterial contaminants) were absent in *P. infestans* T30-4 and 13, 25, 210, 194 and 616 genes were absent in *P. infestans* PIC99189, *P. infestans* 90128, *P. ipomoeae* PIC99167, *P. mirabilis* PIC99114 and *P. phaseoli* F18, respectively.

A

Strain	Gb sequenced	Estimated genome coverage
P. infestans T30-4	2.52	10.5x
P. infestans PIC99189	2.48	10.4x
P. infestans 90128	4.09	17.1x
P. ipomoeae PIC99167	2.87	12.5x
P. mirabilis PIC99114	3.06	11.0x
P. phaseoli F18	1.96	9.0x

B



C

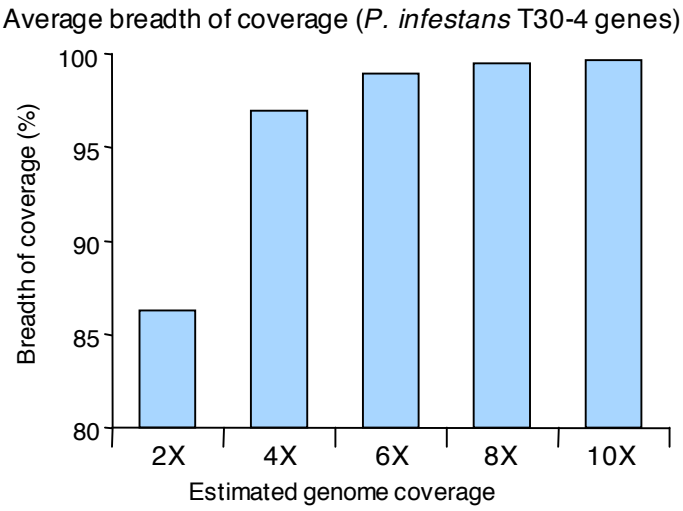


Fig. 2.4. Re-sequencing data: Genomes coverage and gene breadth of coverage for *P. infestans* T30-4

(A) Summary table showing the amount of sequence generated (Gb, gigabases) and estimated genome coverage for the 6 strains used in this study. (B) Number of genes found missing (breadth of coverage = 0) among the 18,155 *P. infestans* T30-4 genes as a function of the estimated genome coverage. Two genes from the reference gene set were identified as contaminants and were independently assigned as bacterial contaminants based on similarity searches to bacterial genomes. (C) Average breadth of coverage (number of CDS bases covered per gene, as a % of full length CDS) for the 18,155 *P. infestans* T30-4 genes as a function of estimated genome coverage.

In re-sequenced *P. infestans* isolate 06_3928A, in which pair end sequence data was generated, breadth of coverage for each of the 18,155 genes was also calculated as described above in the single en read data, by using the percentage of nucleotides with at least one read aligned. Genes were considered absent conservatively when breadth equals 0.

2.4.11. Estimation of copy number from average read depth

Average Read Depth for the CDS of a gene 'g' ARD(g) was calculated and adjusted using GC content in similar manner to a previous reported method (Yoon et al., 2009). Adjusted ARD for a gene 'g' belonging the ith GC content percentile was obtained by the formula AARD(g)= ARD(g).mARD/mARDGC where mARD is the overall mean depth in strain and mARDGC is the mean depth for genes in the ith GC content percentile. Distribution of read depth as a function of GC content scaled by percentile of genes sequenced shows a typical reverse-U shape (Bentley et al., 2008) (Fig. 2.5A). Adjusted read depth frequency is close to random with a distribution variance being 1.45 times that of a Poisson distribution in *P. infestans* T30-4 (Fig. 2.5B). The accuracy of gene copy number prediction based on ARD was tested by comparing members in paralog groups in *P. infestans* T30-4 reference genome (as true copy number) to estimated gene copy number based on ARD in the re-sequenced *P. infestans* T30-4 genome.
A total of 249 paralog groups were identified in *P. infestans* T30-4 reference genome that share 100% amino acid identity along 100% of their predicted CDS (Fig. 2.6A-B). 4,641 *P. infestans* T30-4 genes that lack imperfect paralogs were selected (with blastp e-value < 10E-05) to serve as single copy gene set. ARD was adjusted using GC content and filtered out for outliers. A scatter plot of cumulated depth of coverage as a function of paralog number is shown in (Fig. 2.6C). The "expected" Copy Number (red line) corresponds to True Copy Number x Average Read depth over the whole genome. The regression of cumulated read depth values in paralog groups predicts accurately true copy numbers (members of paralog group) in *P. infestans* T30-4 genome (Fig. 2.6C). ARD provides a good estimate of copy number, although it underestimates copy number for highly duplicated genes. This underestimation is likely due to

imperfect copies, notably truncated copies, or copies containing deletions (see example shown in Fig. 2.6B). Copy Number for a gene ‘g’ CN(g) was calculated as AARD(g).mARD. Copy Number Variation for a gene ‘g’ in a strain ‘s’ is then given by: CNV(g,s) = CN(g,s) - CN(g,T30-4); where CN(g,s) is the estimated copy number of ‘g’ in a strain ‘s’. As a result, an absent gene will have a CNV value of -1, a single copy gene a CNV value of 0, a two-copy gene a CNV value of 1 and so on.

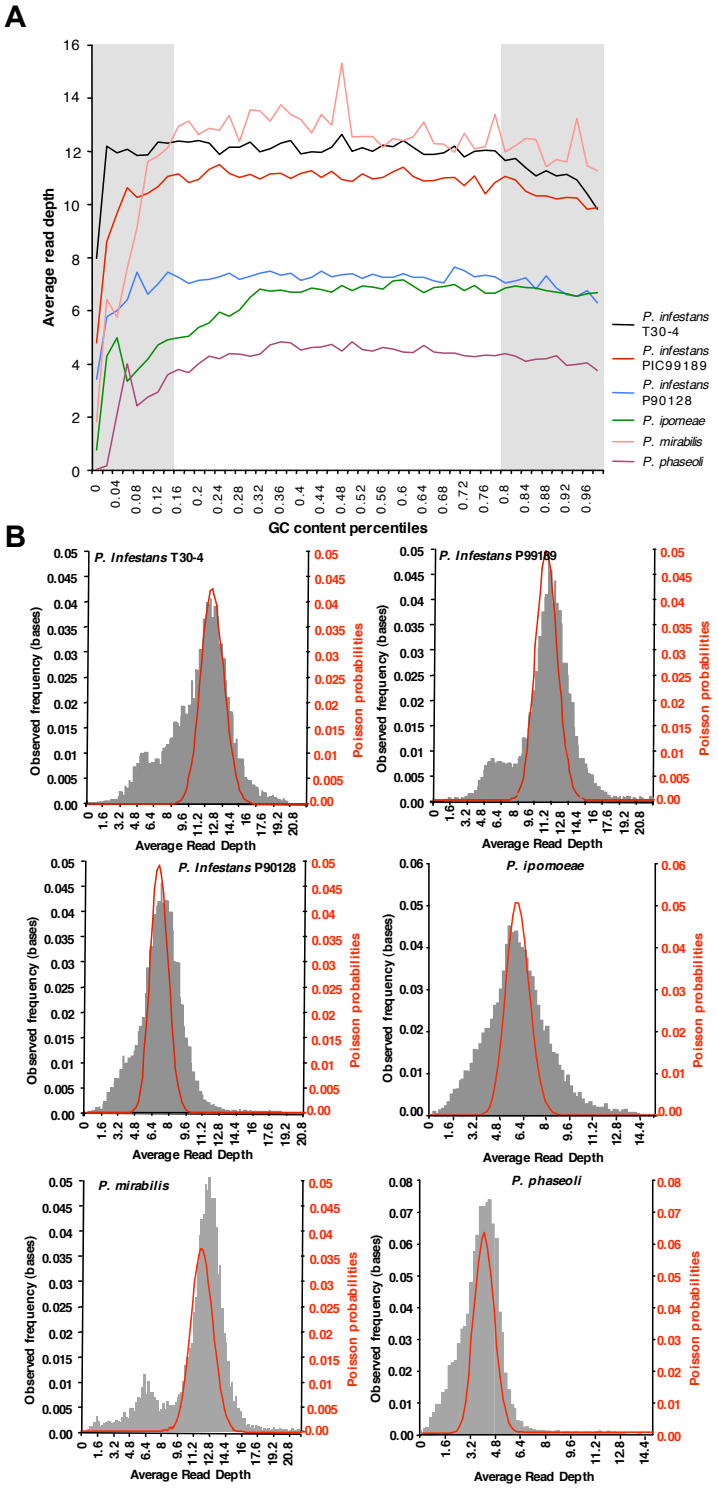


Fig. 2.5. Extreme GC content bias and nearly random distribution of average read depth in re-sequenced *Phytophthora* strains

(A) Distribution of average read depth per gene as a function of GC content percentiles in the re-sequenced strains. The 1 to 4 % lowest and highest GC content genes show lower average read depth. A correction was applied prior to calculation of gene copy numbers to compensate this bias. (B) Distribution of mapped read depth in re-sequenced genomes shown as a histogram. Solid lines represent a Poisson distribution with the same mean.

A

Number of paralog groups identified in *P. infestans* T30-4 genome	Size of paralog group (N° of genes)
1	10
1	6
2	5
6	4
12	3
227	2
4641	1 (no BlastP hit with e-value<10^{-05})

B

```
PITG_13948    1 MAEVELECAVYGEGTVFPVKIASNAKVSALQKAIVNEKKDVNDRFKVDPARLTLYLARKK
PITG_13953    1 MAEVELECAVYGEGTVFPVKIASNAKVSALQKAIVNEKKDVNDRFKVDPARLTLYLARKK
PITG_13950    1 MAEVELECAVYGEGTVFPVKIASNAKVSALQKAIVNEKKDVNDRFKVDPARLTLYLARKK
PITG_13955    1 MAEVELECAVYGEGTVFPVKIASNAKVSALQKAIVNEKKDVNDRFKVDPARLTLYLARKK
PITG_17541    1 MAEVELECAVYGEGTVFPVKIASNAKVSALQKAIVNEKKDVNDRFKVDPARLTLYLARKK
PITG_13952    1 MAEVELECAVYGEGTVFPVKIASNAKVSALQKAIVNEKKDVNDRFKVDPARLTLYLARKK

PITG_13948   61 GEATWMKHDHTVKGFL
PITG_13953   61 GEATWMKHDHTVKGFL
PITG_13950   61 GEATWMKHDHTVKGFL
PITG_13955   61 GEATWMKHDHTVKGFL
PITG_17541   61 GEATWMKHDHTVKGFL
PITG_13952   61 GEATWMKHDHTVKGFL

PITG_20336    1 MRVLCLALMATATVLVPSPASGLTTTVADTAQTATSILTPVLAGEPNKHVATRSLRTHPI
PITG_23230    1 MRVLCLALMATATVLVPSPASGLTTTVADTAQTATSILTPVLAGEPNKHVATRSLRTHPI
PITG_20936    1 MRVLCLALMATATVLVPSPASGLTTTVADTAQTATSILTPVLAGEPNKHVATRSLRTHPI
PITG_20934    1 MRVLCLALMATATVLVPSPASGLTTTVADTAQTATSILTPVLAGEPNKHVATRSLRTHPI

PITG_20336   61 DDSDDGEERLLNGMTDFFKYHAGK---------------------------------
PITG_23230   61 DDSDDGEERLL-------------------------------------------
PITG_20936   61 DDSDDGEERLLNGMTDFFKYHAGKMSPEQLYKYLNLKGLGQEAYKHKNYASYIKKSKKWW
PITG_20934   61 DDSDDGEERLLNGMTDFFKYHAGKMSPEQLYKYLNLKGLGQEAYKHKNYASYIKKSKKWW
```

C

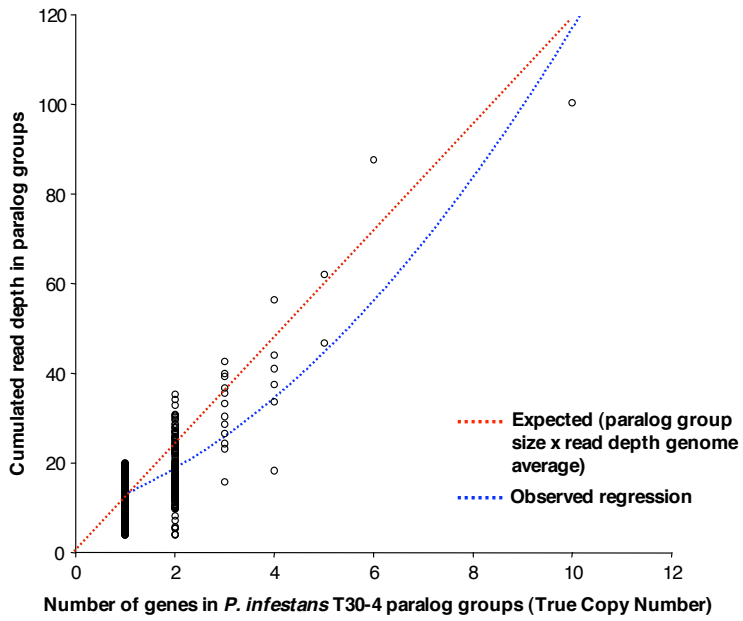


Fig. 2.6. Validation of the estimation of gene copy number from average read depth using paralog groups in *P. infestans* T30-4 reference strain
(A) Summary table of paralog groups identified in *P. infestans* T30-4 reference genome showing number of genes from group and number of groups found. Paralogs were defined as sequences with 100% amino acid identity over 100% of the aligned sequence length. (B) Examples of paralog group alignments. The second example illustrates a possible source of deviation of observed cumulated depth from expected value. (C) Cumulated read depth in paralog groups as a function of the number of genes from group. 'Expected' line corresponds to the number of genes in a group multiplied by the average read depth per gene.

2.4.12. Analysis of polymorphism and gene expression in GDR and GSR

Gene-Dense Region (GDR) genes were considered those with both 5' and 3'FIRs ≤ 1.5Kb (6,689 genes, 36.8% of all predicted genes), and as Gene Sparse Region (GSR) genes those with both 5' and 3'FIRs ≥ 1.5Kb (4,030 genes, 22.1% of all predicted genes). Tuckey Box and Whisker plots were used as a compact way to represent dispersion of CNV, SNPkb, and ω data in GDRs and GSRs (Fig. 5.3). In these plots, the central circle represents the median of the distribution, and the box corresponds to 1st and 3rd quartiles. Top and bottom whiskers values correspond to the first measurement outside of 1.5 times the interquartile range. Outliers were omitted for clarity.

An unpaired Fisher's exact test assuming unequal variance in R software was used to test the significance of differences in the distribution of CNV between GDR and GSR genes. A Mann-Whitney U-test on CNV and gene induction data was also performed. A Fisher's exact test assuming equal or unequal variances was used to test the significance of differences in the distribution of SNP_{kb}, ω and gene induction data between GDR and GSR. Finally, a hypergeometric test was used to test the significance of differences in the distribution of presence/absence polymorphisms between GDR and GSR (Table 2.3 and Table 2.4). The following thresholds for significance of p-values were considered: <10-E04 (***); <0.001 (**); <0.01 (*); <0.1 (.) (Fig. 5.3A).

Table 2.3. Statistical tests supporting differences between GDR and GSR genes regarding gene evolution

	CNV *Pi* PIC99189	CNV *Pi* 90128	CNV *P. ipomoeae*	CNV *P. mirabilis*	CNV *P. phaseoli*
Average	GDR = 0.02424 GSR = 0.03705	GDR = 0.02894 GSR = 0.10086	GDR = 0.11034 GSR = 0.17018	GDR = -0.00190 GSR = 0.17085	GDR = 0.06240 GSR = 0.15302
Standard Deviation	GDR = 0.22971 GSR = 0.38706	GDR = 0.24438 GSR = 0.48962	GDR = 0.43685 GSR = 1.09321	GDR = 0.43495 GSR = 1.59364	GDR = 0.38358 GSR = 1.15968
Variance	GDR = 0.05276 GSR = 0.14981	GDR = 0.05972 GSR = 0.16723	GDR = 0.19084 GSR = 1.18879	GDR = 0.18917 GSR = 2.53969	GDR = 0.14714 GSR = 1.34485

Unpaired T test unequal variance	t = -1.9081 df = 5763.972 p-val = 0.05643	t = -12.7685 df = 5852.887, p-val < $2.2e^{-16}$	t = -3.3268 df = 4818.886 p-val = 0.000885	t = -6.7317 df = 4393.413 p-val = $1.893e^{-11}$	t = -4.8047 df = 4565.719 p-val = $1.599e^{-06}$
Mann-Whitney U test	W = 13516511 p-val = 0.8057	W = 14888238, p-val < $2.2e^{-16}$	W = 12575528 p-val = $5.964e^{-09}$	W = 14276956 p-val = $2.656e^{-07}$	W = 12640674 p-val = $6.737e^{-08}$
Significance	.	***	***	***	***

	Gain/loss *Pi* PIC99189	Gain/loss *Pi* 90128	Gain/loss *P. ipomoeae*	Gain/loss *P. mirabilis*	Gain/loss *P. phaseoli*
Average	GDR = 5 GSR = 1 Total = 13	GDR = 16 GSR = 0 Total = 25	GDR = 124 GSR = 9 Total = 210	GDR = 111 GSR = 11 Total = 194	GDR = 312 GSR = 45 Total = 616
Hypergeometric test	Cumulative prob. = 0.98063	Cumulative prob. = 0.999	Cumulative prob. = 0.999	Cumulative prob. = 1	Cumulative prob. = 1
Significance	**	***	***	***	***

	SNP frequency *Pi* PIC99189	SNP frequency *Pi* 90128	SNP frequency *P. ipomoeae*	SNP frequency *P. mirabilis*	SNP frequency *P. phaseoli*
Average	GDR = 1.42153 GSR = 1.35590	GDR = 1.46044 GSR = 1.38982	GDR = 11.67757 GSR = 10.60108	GDR = 16.84811 GSR = 16.70401	GDR = 18.57755 GSR = 16.85050
Standard Deviation	GDR = 6.16733 GSR = 4.97588	GDR = 4.95869 GSR = 4.56299	GDR = 26.6503 GSR = 21.0479	GDR = 30.07557 GSR = 29.86253	GDR = 41.17146 GSR = 37.52455
Variance	GDR = 38.03597 GSR = 24.76597	GDR = 24.5886 GSR = 20.8246	GDR = 710.2383 GSR = 443.0147	GDR = 904.542 GSR = 891.771	GDR = 1695.089 GSR = 1408.092

Unpaired T test unequal variance	t = 0.6034 df = 9853.482 p-val = 0.5463	t = 0.7511 df = 9043.841, p-val = 0.4526	t = -2.3157 df = 9968.459 p-val = 0.02060	t = -0.2413 df = 8537.269 p-val = 0.8093	t = -2.1729 df = 9106.801 p-val = 0.02982
Unpaired T test equal variance	t = 0.5725 df = 10717 p-val = 0.567	t = 0.7358 df = 10717 p-val = 0.4619	t = 2.1862 df = 10717 p-val = 0.02883	t = 0.2409 df = 10717 p-val = 0.8096	t = 2.1236 df = 10717 p-val = 0.03373
Significance			.		.

	dN/dS *Pi* PIC99189	dN/dS *Pi* 90128	dN/dS *P. ipomoeae*	dN/dS *P. mirabilis*	dN/dS *P. phaseoli*
Average	GDR = 0.29625 GSR = 0.31123	GDR = 0.29957 GSR = 0.30707	GDR = 0.30045 GSR = 0.44660	GDR = 0.29907 GSR = 0.34791	GDR = 0.31161 GSR = 0.51911
Standard Deviation	GDR = 0.4184 GSR = 0.4325	GDR = 0.415 GSR = 0.509	GDR = 0.3674 GSR = 0.5412	GDR = 0.6714 GSR = 0.5425	GDR = 0.35761 GSR = 0.62132
Variance	GDR = 0.175098 GSR = 0.187054	GDR = 0.172 GSR = 0.259	GDR = 0.134736 GSR = 0.292338	GDR = 0.121045 GSR = 0.450171	GDR = 0.127885 GSR = 0.386128

Unpaired T test unequal variance	t = -0.7378 df = 877.838 p-val = 0.4608	t = -0.3348 df = 841.197 p-val = 0.7379	t = -11.8552 df = 3067.755 p-val < $2.2e^{-16}$	t = -17.5716 df = 3214.95 p-val < 2.2e-16	t = -15.1017 df = 2890.305 p-val < $2.2e^{-16}$
Unpaired T test equal variance	t = -0.7519 df = 2664 p-val = 0.4522	t = -0.3759 df = 2835 p-val = 0.707	t = -14.0424 df = 8268 p-val < $2.2e^{-16}$	t = -22.3294 df = 8816 p-val < $2.2e^{-16}$	t = -19.0445 df = 8416 p-val < $2.2e^{-16}$
Significance			***	***	***

Table 2.4. Statistical tests supporting differences between GDR and GSR genes regarding gene expression *in planta*

	Sporangia	Zoospores	Pot_6hpi	Pot_16hpi
Average	GDR= 0.3078 GSR= -0.0320	GDR= 0.0175 GSR= 0.0273	GDR= 0.2317 GSR= 0.0774	GDR= 0.2057 GSR= 0.0836
Standard deviation	GDR= 0.6643 GSR= 0.5163	GDR= 0.7417 GSR= 0.5971	GDR= 0.9652 GSR= 1.1898	GDR= 0.8994 GSR= 1.1226
Variance	GDR= 0.4414 GSR= 0.2666	GDR= 0.5501 GSR= 0.3565	GDR= 0.9315 GSR= 1.4157	GDR= 0.8090 GSR= 1.2602
Unpaired T-test Equal Variance	t= 27.3786 df= 10509 p-value< 2.2e-16 95% conf. interval: 0.3155 0.3641 mean of x 0.3078 mean of y -0.0320	t= -0.7047 df= 10509 p-value= 0.481 95% conf. interval: -0.0373 0.0176 mean of x 0.0175 mean of y 0.0273	t= 7.2468 df= 10509 p-value= 4.566e-13 95% conf. interval: 0.1126 0.1961 mean of x 0.2317 mean of y 0.0774	t= 6.1152 df= 10509 p-value= 9.984e-10 95% conf. interval: 0.0830 0.1612 mean of x 0.2057 mean of y 0.0836
Unpaired T-test Unequal Variance	t= 29.2141 df= 9696.756 p-value<2.2e-16 95% conf. interval: 0.3170 0.3626 mean of x 0.3078 mean of y -0.0320	t= -0.7454 df= 9500.533 p-value= 0.4561 95% conf. interval: -0.0358 0.0161 mean of x 0.0175 mean of y 0.0273	t= 6.8645 df= 6836.583 p-value= 7.258e-12 95% conf. interval: 0.1103 0.1984 mean of x 0.2317 mean of y 0.0774	t= 5.7747 df= 6770.158 p-value= 8.053e-09 95% conf. interval: 0.0806 0.1635 mean of x 0.2057 mean of y 0.0836
Mann-Whitney U-test	W= 17669991 p-value<2.2e-16	W= 12628692 p-value= 0.1255	W= 13847882 p-value= 4.372e-11	W= 13713669 p-value= 1.221e-08
Significance	***		***	***

	Pot_2dpi	Pot_3dpi	Pot_4dpi	Pot_5dpi
Average	GDR= 0.0423 GSR= 0.0150	GDR= -0.0031 GSR= -0.0025	GDR= -0.0574 GSR= -0.0286	GDR= -0.2959 GSR= -0.0214
Standard deviation	GDR= 0.5756 GSR= 0.6791	GDR= 0.4316 GSR= 0.4984	GDR= 0.3675 GSR= 0.3921	GDR= 0.4530 GSR= 0.4038
Variance	GDR= 0.3313 GSR= 0.4612	GDR= 0.1862 GSR= 0.2484	GDR= 0.1350 GSR= 0.1538	GDR= 0.2052 GSR= 0.1631
Unpaired T-test Equal Variance	t= 2.1908 df= 10509 p-value= 0.02849 95% conf. interval: 0.0029 0.0517 mean of x 0.0423 mean of y 0.0150	t= -0.062 df= 10509 p-value= 0.9505 95% conf. interval: -0.0187 0.0176 mean of x -0.0031 mean of y -0.0025	t= -3.7918 df= 10509 p-value= 1.5e-04 95% conf. interval: - 0.0438 -0.0140 mean of x -0.0574 mean of y -0.0286	t= -31.1759 df= 10509 p-value< 2.2e-16 95% conf. interval: - 0.2917 -0.2572 mean of x -0.2959 mean of y -0.0214
Unpaired T-test Unequal Variance	t = 2.0985 df= 7081.182 p-value= 0.0359 95% conf. interval: 0.0018 0.0528 mean of x 0.0423 mean of y 0.0150	t= -0.0597 df= 7205.468 p-value= 0.9524 95% conf. interval: -0.0194 0.0182 mean of x -0.0031 mean of y -0.0025	t= -3.7277 df= 7688.17 p-value= 1.9e-04 95% conf. interval: 0.0441 -0.0137 mean of x -0.0574 mean of y -0.0286	t= -32.1269 df= 8867.323 p-value< 2.2e-16 95% conf. interval: - 0.2912 -0.2577 mean of x -0.2959 mean of y -0.0214
Mann-Whitney U-test	W= 13834206 p-value= 8.039e-11	W= 13353799 p-value= 0.0009	W= 12539278 p-value= 0.03335	W= 7672013 p-value< 2.2e-16
Significance	***	**		***

	Tom_2dpi	Tom_3dpi	Tom_5dpi
Average	GDR= -0.0589 GSR= 0.0306	GDR= -0.1605 GSR= -0.0295	GDR= -0.2052 GSR= -0.0316
Standard deviation	GDR= 0.4678 GSR= 0.5315	GDR= 0.4533 GSR= 0.5646	GDR= 0.4371 GSR= 0.4411
Variance	GDR= 0.2188 GSR= 0.2824	GDR= 0.2055 GSR= 0.3187	GDR= 0.1911 GSR= 0.1946
Unpaired T-test Equal Variance	t= -8.9894 df= 10509 p-value< 2.2e-16 95% conf. interval: -0.1090 -0.0699 mean of x -0.0589 mean of y 0.0306	t= -13.0221 df= 10509 p-value< 2.2e-16 95% conf. interval: -0.1506 -0.1112 mean of x -0.1605 mean of y -0.0295	t= -19.572 df= 10509 p-value< 2.2e-16 95% conf. interval: - 0.1909 -0.1562 mean of x -0.2052 mean of y -0.0316
Unpaired T-test Unequal Variance	t= -8.6947 df= 7302.693 p-value< 2.2e-16 95% conf. interval: -0.1096 -0.0693 mean of x -0.0589 mean of y 0.0306	t= -12.3034 df= 6781.478 p-value< 2.2e-16 95% conf. interval: -0.1518 -0.1100 mean of x -0.1605 mean of y -0.0295	t= -19.5251 df= 8048.5 p-value< 2.2e-16 95% conf. interval: - 0.1910 -0.1561 mean of x -0.2052 mean of y -0.0316
Mann-Whitney U-test	W= 11934000 p-value= 7.3e-10	W= 11294678 p-value< 2.2e-16	W= 9823851 p-value< 2.2e-16
Significance	***	***	***

Pot_, potato; Tom_, tomato; hpi, hours post inoculation; dpi, days post inoculation

2.4.13. Visualisation of polymorphism and gene expression relative to genome architecture

Genes were sorted in two-dimensional bins according to length of 5' and 3' FIRs (along Y and X axis respectively) as described earlier in Haas *et al.* (2009) (Haas et al., 2009). A color scale was used to represent either the (i) number of genes in bins or (ii) average polymorphism values (CNV, SNP frequency or dN/dS) or gene induction value (as log2 of the ratio of expression in sample over expression in mycelia grown in vitro) associated to genes in a given bin.

2.4.14. Fast-evolving genes and tribes enriched in GSRs and fast-evolving genes

GO mapping was performed on all 18,155 *P. infestans* T30-4 predicted proteins using the Blast2Go server (Conesa et al., 2005). Gene tribes were identified by Markov Clustering using the TribeMCL option in BioLayout Express3D 7 (Freeman et al., 2007). The output of a BlastP analysis of *P. infestans* T30-4 predicted proteome versus itself with 10E-05 e-value cutoffs was used as the input file. This method allowed grouping 9,418 proteins into 1,153 tribes. Considering that GSR and fast evolving genes correspond to 22% and 25% of all genes respectively, further analyses were limited to the 811 tribes containing at least 5 genes (7,993 genes included in 811 tribes), the minimum value from which statistical significance can arise. These tribes were manually checked for dominant functional annotation and GO terms and/or associated annotation whenever applicable.

Genes were classified as fast evolving when matching any of the following criteria: (i) CNV value > 1 in any strain other than *P. infestans* T30-4 (presumed duplicated gene), (ii) dN/dS>1 in any strain other than *P. infestans* T30-4, or (iii) absent in any strain other than *P. infestans* T30-4. Enrichment in genes matching a criterion 'C' (GSR or fast evolving) for a tribe 'T' was calculated as $((Genes_T \cap Genes_C) / Genes_T) / (Genes_C / Genes_{All})$; where $Genes_T$ is the number of genes in the tribe 'T', $Genes_C$ is the number of genes matching criterion 'C', and $Genes_{All}$

is the total number of genes. A list of 4,913 genes matching at least one of the criteria was retrieved. A chi-square test implemented in R was performed on all 811 tribes for enrichment in GSR genes and/or fast evolving genes. The following p-value thresholds were considered: ***, p-val. <0.01; **, p-val. <0.05; * p-val <0.1. Depletion and enrichment are defined relative to the proportion in the whole genome. Among the 811 tribes, we found 163 tribes (20.1%) enriched in GSR genes (88, 56 and 19 with 0.01, 0.05 and 0.1 p-value thresholds respectively), and 123 tribes (15.2%) enriched in fast evolving genes (66, 42 and 15 with 0.01, 0.05 and 0.1 p-value thresholds respectively) and 65 tribes in both. Then, I looked at the gene induction value (as log2 of the ratio of expression in sample over expression in mycelia grown in vitro) associated to the genes contained in the 65 tribes (see appendix 3.1).

2.5. Whole-genome expression analysis of *P. infestans* isolates

2.5.1. Gene expression analysis

Mycelia were harvested after growing for 10-12 days in V8 juice Agar or Rye Sucrose Agar (RSA), ground in liquid nitrogen and frozen prior RNA extraction. In addition to mycelia, I collected leaf discs infected with zoospores of *P. infestans* T30-4, 06_3928A and NL07434 at different days post inoculation: 2, 3 and 4 on potato. For *P. infestans* T30-4, I also analyzed day 5 and earlier time points of infection 6 and 16hpi on potato, and days 2, 3 and 5 on tomato. The infected material was ground in liquid nitrogen to a fine powder and frozen prior RNA extraction. Each sample and its biological replicate were homogenized with RLT buffer containing ß-mercaptoethanol from the RNAesy Plant Mini Kit (Qiagen, Cat No. 74904) proceeding with a modified manufacture's protocol. RNA quality and integrity were checked prior to cDNA synthesis using the Bioanalyzer (Agilent 2100). NimbleGen microarray services were utilized for cDNA preparations and subsequent chip hybridizations to a custom array design (080603_PI_BH_EXP) that include all predicted genes in *P. infestans* and tomato ESTs. Microarray normalization was done using the previously described methods in (Haas et al., 2009). Analysis of gene expression was performed using the MultiExperiment

viewer (MeV). Log2 transformed array intensity values were analyzed for differential gene expression using the t-test, as implemented in MeV (Saeed et al., 2003), assuming equal variances. For this study, only the array targets corresponding to annotated *P. infestans* genes were analyzed. T-tests were performed comparing two groups: Group A, consisting of sample replicates for mycelia grown in RSA and V8, and Group B, consisting of replicates for one of the days post-infection. False discovery rates were addressed by computing q-values for each test (Storey and Tibshirani, 2003). Such tests were performed for each pair of replicates post-infection, and all significantly ($p < 0.05$, $q < 0.05$) differentially expressed genes were reported. Significant differentially regulated genes exhibiting at least two-fold variation in gene expression between averaged media and infected potato sample replicates were considered induced during infection.

2.5.2. Measurement of biotrophic growth during infection

P. infestans strains T30-4, 06_3928A and NL07434 were grown in RSA plates for 12 days at 18°C. Sporangia were harvested from RSA plates by adding cold water to the plates and zoospores were collected after 3 hours of incubation at 4°C. Potato leaves were drop inoculated with a solution of 100,000 zoospores/ml. Droplets of 10 µl were applied onto abaxial sides of potato-detached leaves on wet paper towels. Two droplets per leaf with a total of 28 droplets in 14 leaves were applied separately for each of the three isolates. Infected leaves were exposed to UV light at 2, 3 and 4 days after inoculation (dpi) and whole leaves digital images were recorded with Gel Doc imaging system (Biorad). UV light exposed digitalized leaf images were loaded in Image J (1.43u) software (Rasband) and the areas (in mm^2) for the outer ring (include both non-necrotized and necrotized region) and the inner ring (necrotized region) were calculated with the area function of Image J. Then, I calculated the diameters from the outer and the inner areas by applying the formulas, $r = sqrt[a/\pi]$ (where a = *area* and $\pi = 3.1416$) and then $d = 2r$ (where d = *diameter* and r = *ratio*). Finally, I calculated the difference between outer and the inner ring diameters to estimate the extend of the biotrophic growth only in mm. For each time point I estimated the standard

error (28 replicates that were measured) for the diameter by using the formula *stderror = STDEV(range) / SQRT(COUNT(range))* in excel 2008 for Mac OSX, where *STDEV* is the sample standard deviation.

2.6. Gene expression profiling of *Puccinia monoica – Boechera stricta* interaction

2.6.1. Plant material

Table 2.5. List of infected and uninfected plant material collected in the field

Sample type	Sample ID	Sample description	Sample replicates
Infected plant*	Pseudoflower ('Pf')	*Pseudoflowers from Puccinia monoica infected plants*	Pseudoflower GOT1-4 Pseudoflower GOT1-6 Pseudoflower GOT1-8
Uninfected plant*	Stem and Leaves ('SL')	Uninfected *Boechera stricta* stem and leaves	Stem GOT-21B Stem GOT-22B Stem GOT-23B
	Flower ('F')	Uninfected *Boechera stricta* natural flowers	Flower GOT-21A Flower GOT-22A

*Plant material was collected from Gothic (2900m) about 5 miles away from the Rocky Mountain Biological Laboratory near Gunnison, CO, USA.

2.6.2. Gene expression analysis

For the microarray experiments of rust pseudoflowers, I extracted total RNA from pseudoflowers from *Puccinia monoica* infected plants ('Pf') and uninfected *Boechera stricta* plant stems and leaves ('SL'), and uninfected *B. stricta* flowers ('F') (Table 2.5). Tissue harvested in the field was stored in RNALater solution (Ambion) before being transported to the lab. Total RNA was purified from each independent sample using TRIzol Reagent (Invitrogen Corp.) according to the manufacturer's instructions. RNA quality and integrity were checked prior to cDNA synthesis using the Bioanalyzer (Agilent 2100). NimbleGen microarray services were utilized for cDNA preparations, chip hybridizations to an *Arabidopsis thaliana* custom array design (ATH6 60mer X4 exp, Cat No. A4511001-00-01) and subsequent normalization of the probe sets using Robust Multichip Average (RMA) (Bolstad et al., 2003).

For the microarray analysis, I combined the results from two independent analyses. First, a t-test using the log2 expression values was performed to detect and describe global gene expression changes. T-tests were calculated from three combinations: 'Pf' *vs* 'SL' (Group B consisting in rust infected plant with pseudoflowers sample replicates GOT-4, 6, 8 *vs* Group A equal to uninfected plant stem and leaves samples replicates GOT-21B, 22B and 23B), 'F' *vs* 'SL' (Group B consisting of uninfected plant flower sample replicates GOT-21A, 22A *vs* Group A equal to uninfected plant stem and leaves samples replicates GOT-21B, 22B and 23B) and 'Pf' *vs* 'F' (Group B consisting in rust infected plant with pseudoflowers sample replicates GOT-4, 6, 8 *vs* Group A consisting of uninfected plant flower sample replicates GOT-21A, 22A). T-test parameters included assumption of equal variances and p-value based on t-distribution. Then, all significantly ($p < 0.05$) differentially expressed genes obtained from the T-test were reported. Second, a False Discovery Rate (FDR) with Rank Products (RP) using the Log2 expression values was performed to identify biologically relevant gene changes from different environmental backgrounds (Breitling et al., 2004). RP function more reliable and consistently than non-parametric t-test in the analysis of samples subjected to genetic or environmental factors, for instance in samples collected in the field and not under controlled laboratory conditions (Kammenga et al., 2007). In RP genes are ranked based on up- or down-regulation in each experiment. Then, for each gene a combined probability is calculated as a False Discovery Rate (FDR) value based on permutations. For this study, FDR values were calculated using 5,000 permutations from three combinations: 'Pf' *vs* 'SL' (Group B consisting in rust infected plant with pseudoflowers sample replicates GOT-4, 6, 8 *vs* Group A equal to uninfected plant Stem and Leaves samples replicates GOT-21B, 22B and 23B), 'F' *vs* 'SL' (Group B consisting of uninfected plant flower sample replicates GOT-21A, 22A *vs* Group A equal to uninfected plant stem and leaves samples replicates GOT-21B, 22B and 23B) and 'Pf' *vs* 'F' (Group B consisting in rust infected plant with pseudoflowers sample replicates GOT-4, 6, 8 *vs* Group A consisting of uninfected plant flower sample replicates GOT-21A, 22A). Genes with FDR values less than 0.05 were considered differentially expressed between the comparisons used for each combination.

2.6.3. Gene ontology enrichment and pathways analysis

Gene Ontology (GO) annotations data was extracted from the Arabidopsis database TAIR (Berardini et al., 2004). Over-represented groups of GO terms and functional domains were identified using a hypergeometric test, with a threshold of p-value of 0.05 using BINGO (Maere et al., 2005) plugging installed in Cytoscape. The hypergeometric test compared the 27,822 GO annotated genes, with the GO terms associated to the significantly regulated genes in each experiment: 948 genes in 'Pf' *vs* 'SL' and 859 genes in 'F' *vs* 'SL' ('Pf', pseudoflowers from *Puccinia monoica* infected plants; 'SL', uninfected *Boechera stricta* stem and leaves; 'F', uninfected *B. stricta* flowers). Pathways were analyzed using AraCyc database (http://plantcyc.org/release_notes/aracyc/aracyc_release_notes.faces).

CHAPTER 3: Functional validation of signal peptides of *Phytophthora infestans* RXLR effectors

3.1. Introduction

Two crucial findings have facilitated the computational prediction of effectors in oomycetes and their use in high throughput functional assays. The first crucial finding was the validation of the concept that effectors proteins must be secreted in order to reach their targets proteins in the apoplast or cytoplasm of the host cell (Torto et al., 2003). To be secreted effectors must encode N-terminal signal peptides that direct the transport of the mature proteins to the secretory pathway. Prediction of signal peptides in pathogens proteins aimed at generating catalogues of secreted proteins (secretome) is important step in the identification of effectors genes involved in pathogen infection and host-pathogen interactions (Grell et al., 2011; Kamoun, 2009; Mueller et al., 2008; Raffaele et al., 2010b). The second crucial finding was the identification in oomycetes of secreted effectors with a conserved translocation motif, RXLRs (Whisson et al., 2007). RXLRs are modular proteins that carry N-terminal signal peptides and the RXLR motif that functions in secretion and targeting and a variable C-terminal domain that carries the effector activity and functions inside the host cell (Birch et al., 2006; Morgan and Kamoun, 2007; Schornack et al., 2009). Both the secretion signals and the RXLR motifs led to *ab initio* identification of RXLR effectors in pathogenic oomycetes (Win et al., 2007). All known oomycete effectors identified so far with avirulence activity (AVR proteins) belong to the host-translocated RXLR class and show *in planta* gene induction (Vleeshouwers et al., 2011).

Phytophthora infestans, a pathogenic oomycete that causes late blight in potato is predicted to secrete hundreds of RXLR effector proteins (Haas et al., 2009; Kamoun, 2006; Raffaele et al., 2010b). An *in planta* screening enabled the discovery of four *P. infestans* RXLR effectors, three of them being highly induced during infection in tomato (Oh et al., 2009). PexRD6/AVRblb1, PexRD39/AVRblb2 and PexRD40/AVRblb2 RXLR effectors are AVR proteins

that are recognized by the cognate *Rpi-blb1* and *Rpi-blb2* genes, respectively (Oh et al., 2009; Vleeshouwers et al., 2008). PexRD8 RXLR effector suppresses cell death produced by another secreted protein (Oh et al., 2009). To functionally validate the signal peptide predictions of these four *P. infestans* representative RXLR effector genes induced *in planta*, I used a genetic assay called Signal Sequence Trap (SST) system, based on the requirement of yeast cells for invertase secretion to grow on sucrose or raffinose media (Jacobs et al., 1997; Klein et al., 1996; Lee et al., 2006). Here, using the SST method I report that the signal peptides of these four *P. infestans* RXLR effectors are functional (Lee et al., 2006; Menne et al., 2000; Oh et al., 2009; Schneider and Fechner, 2004). Moreover, recent studies confirm that the SST method is a very useful resource and suggest that this method can be expanded for the analysis of secretion signals in effectors from unrelated oomycetes (Tian et al., 2011).

3.2. Results and discussion

3.2.1. Features of host translocated RXLR effectors of *P. infestans* with avirulence activity

Phytophthora infestans host translocated RXLR effectors are modular proteins with a N-terminal domain consisting of a signal peptide, followed by the RXLR motif that functions in secretion and translocation and a C-terminal domain that carries the effector activity (Fig. 3.1) (Kamoun, 2006; Morgan and Kamoun, 2007; Schornack et al., 2009). 86% (483 out of 563) of the RXLR effectors in *P. infestans* genome are predicted to be secreted with HMM probabilities scores above 0.9 (Haas et al., 2009; Raffaele et al., 2010b) (Fig. 3.2) and only 16% (79 out of the 483) are induced during infection on potato (Fig. 3.2, see appendix 1.1). All known *P. infestans Avr* genes (*Avr1, Avr2, Avr3a, Avr4, Avrblb1, Avrblb2,* and *Avrvnt1*) with avirulence activity belong to the RXLR class and are also induced *in planta* (Fig. 3.1) (Rehmany et al., 2005; Vleeshouwers et al., 2011). AVR proteins can also act as virulence factors, like the effector AVR3a that manipulates the host ubiquitin proteosome system by stabilizing the ubiquitin E3-ligase CMPG1 to suppress plant immunity. This suggests that *in planta*-

induced RXLR effectors are candidate effectors with avirulence or virulence activities in plant cells.

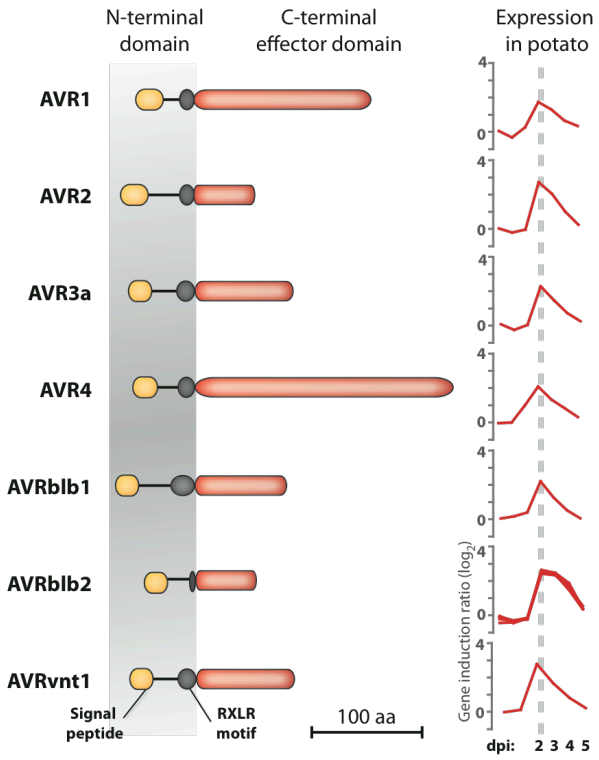


Fig. 3.1. Features of characterized *Phytophthora Avr* gene products
The figure depicts AVR1, AVR2, AVR3a, AVR4, AVRblb1, AVRblb2, and AVRvnt1. The domain structure of *P. infestans* AVR proteins shows a typical RXLR effector modular structure with N-terminal (signal peptide) domain, RXLR motif, and C-terminal effector domain. The N-terminal domain functions in secretion and host translocation whereas the variable C-terminal domain carries the effector biochemical activity. Expression in potato panels illustrates a time course expression pattern of the *Avr* genes during infection of potato [2–5 days post infection (dpi)] with the y-axis showing gene induction. For each gene the line graph shows the gene induction in log2 during infection in potato using mycelia as baseline with a t-test (see chapter 2 section 2.5.1, appendix 1.1). Each of the *Avr* genes is maximally induced at 2 dpi in potato during the early phase of the disease.

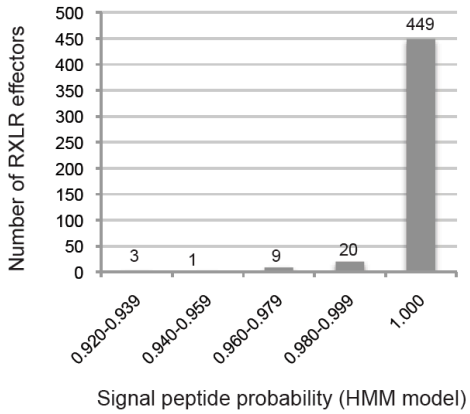


Fig. 3.2. Distribution of signal peptide probabilities in RXLR effectors predicted to be secreted in *P. infestans*
A total of 483 RXLRs was classified in bins according to the signal peptide probabilities calculated from HMM model.

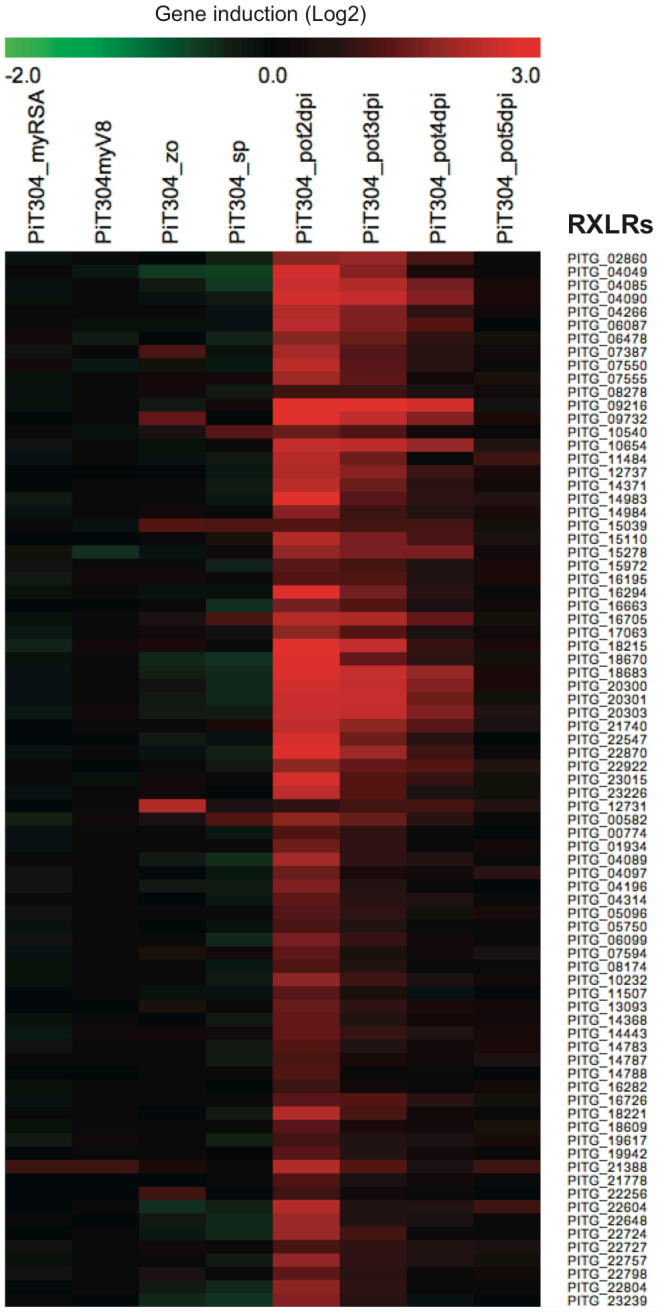


Fig. 3.3. *P. infestans* secreted RXLR effector genes that are induced during infection in potato

79 secreted RXLRs were statistically significantly induced during infection on potato at 2 and/or 3 dpi using mycelia as baseline with a t-test (see chapter 2 section 2.5.1, appendix 1.1). MyRSA, mycelia grown in Rye Sucrose Agar; myV8, mycelia grown in V8 agar; sp, sporangia, zo, zoospores; pot, potato; dpi, days post infection.

3.2.2. RXLR effectors of *P. infestans* used for functional validation of signal peptides

In this study, I selected three RXLR effector genes representative of the 79 induced during infection on potato in *P. infestans* T30-4 (see appendix 1.1) ((Haas et al., 2009), Liliana Cano, unpublished). *PexRD6/ipiO (Avrblb1), PexRD39 (Avrblb2)* and *PexRD40 (Avrblb2)* are avirulence genes that are recognized by their cognate *R* genes resulting in the induction of hypersensitive cell death and immunity (Oh et al., 2009; Vleeshouwers et al., 2008). In addition, I selected the effector gene *PexRD8* that is induced during infection on tomato and that encodes a protein that has been described to suppress the hypersensitive cell death produced by the *P. infestans* INF1 elicitin protein (Oh et al., 2009).

3.2.3. Invertase secretion assay using sucrose/raffinose-containing media

To verify that the predicted signal peptides of the selected RXLR effector genes function in secretion of the corresponding proteins, I used the Signal Sequence Trap system (SST) (Jacobs et al., 1997; Klein et al., 1996; Lee et al., 2006). Deletion of the signal peptide from the invertase gene blocks secretion and prevents growth on sucrose or raffinose. Cloning functional foreign signal peptide sequences in frame with the truncated invertase restores the ability of yeast cells to growth in sucrose and raffinose (Fig. 3.4).

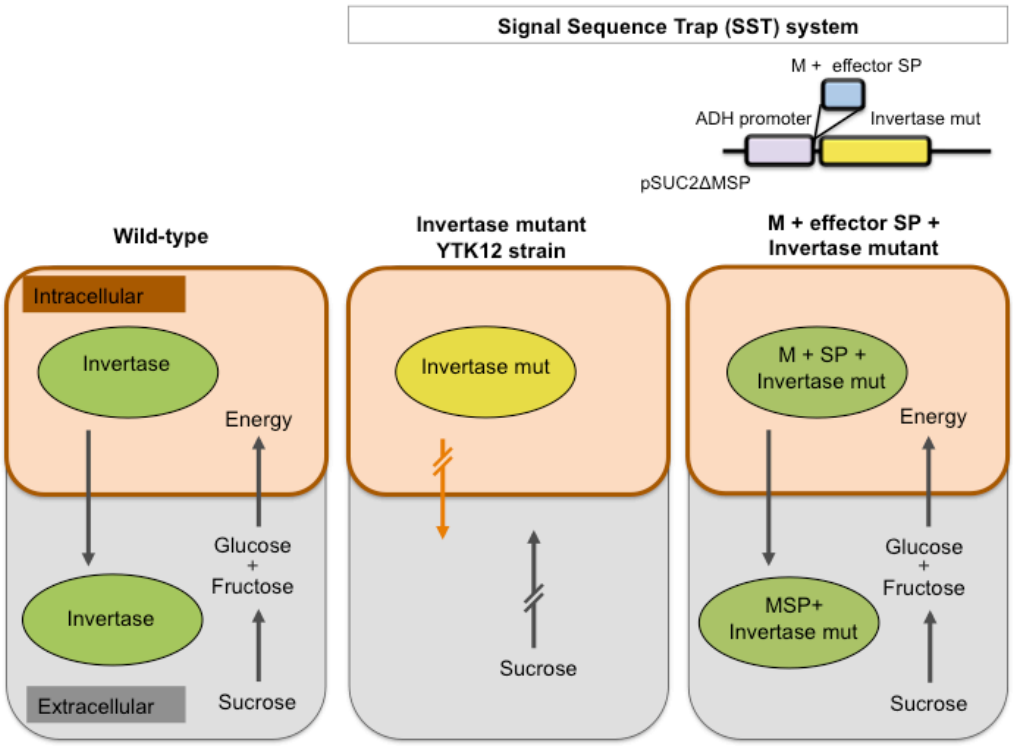


Fig. 3.4. Schematic diagram for identification of secreted proteins using Signal Sequence Trap (SST)
Wild-type yeast is able to grow on sucrose medium by secreting invertase, which metabolizes sucrose and thereby provides glucose as an energy source. An invertase-deficient yeast *Saccharomyces cerevisiae* strain (YTK12) is not able to grow on sucrose medium. A signal peptide sequence carrying their own Methionine (M + SP) is fused to the vector pSUC2ΔMSP in front of a mutated invertase gene (invertase mut) that lacks the N-terminal signal sequence for secretion and then these constructs are transformed the invertase-deficient yeast YTK12 strain. Only clones carrying a effector signal peptide sequence encoding for a secreted protein and whose sequences are in frame are able to secrete invertase, which enables them to grow on sucrose-containing selection medium.

I cloned the predicted signal peptide sequences and the following two amino acids of the four genes encoding selected PexRD proteins (PexRD6/ipiO, PexRD8, PexRD39, and PexRD40), fused them in frame to the mature sequence of yeast invertase in the pSUC2 vector and transformed them in the invertase deficient yeast strain YTK12 (Jacobs et al., 1997) (see chapter 2 Table 2.1, appendix 1.2). Fig. 3.5 shows that untransformed invertase-deficient yeast strain YTK12 was not able to growth in complete minimal medium (CMD-W) media which lacks tryptophan or in the yeast peptone raffinose antimycin (YPRAA) which contains raffinose, a complex sugar that without invertase can not be used by yeast. Invertase-deficient yeast *Saccharomyces cerevisiae* strain YTK12

transformed with an empty pSUC2 vector (construct that carries a tryptophan gene, see appendix 1.2) enabled the YTK12 strain to grow in the CMD-W medium lacking tryptophan. However, I found no growth of the yeast YTK12 mutants carrying the empty pSUC2 in the YPRAA medium, which suggest that there was no invertase secretion activity and no change in the inability of this strain to hydrolyze complex raffinose sugars (Fig. 3.4). On the contrary, all four PexRD constructs enabled the invertase-deficient yeast strain YTK12 to grow on YPRAA medium (with raffinose instead of sucrose, growth only when invertase is secreted) (Fig. 3.5).

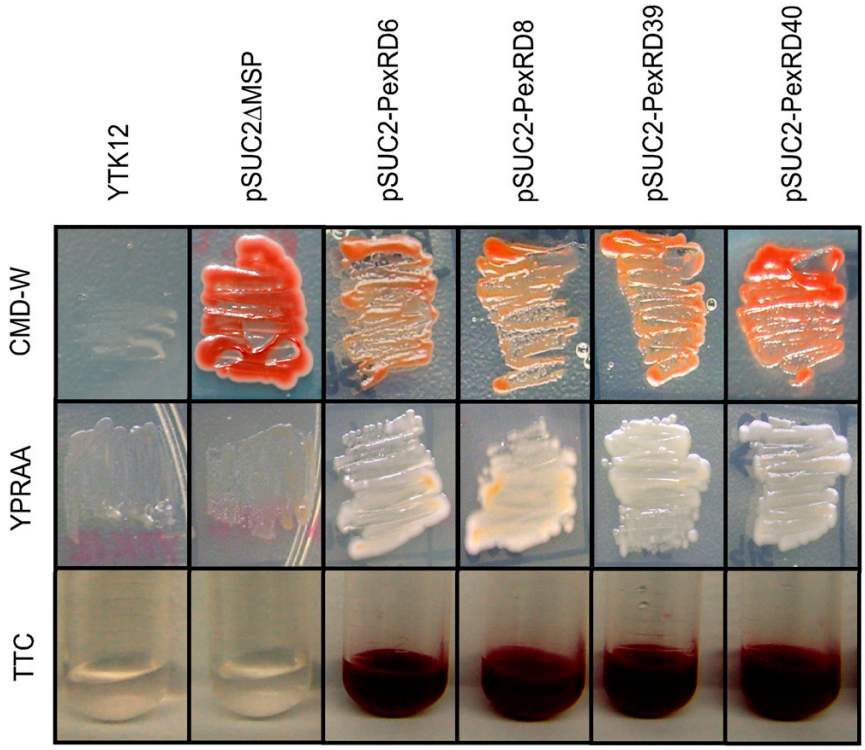


Fig. 3.5. Functional validation of the signal peptides of RXLR effectors of *P. infestans*
Functional validation of the signal peptides of PexRD6/IpiO/AVRblb1, PexRD8, PexRD39/AVRblb2, and PexRD40/ AVRblb2 was performed using the yeast invertase secretion assay. Invertase-deficient yeast *Saccharomyces cerevisiae* YTK12 strain carrying the PexRD signal peptide fragments fused in frame to the invertase gene in the pSUC2 vector are able to grow in both the complete minimal medium lacking tryptophan (CMD-W) and yeast peptone raffinose antimycin (YPRAA) media and reduce the dye 2,3,5-triphenyltetrazolium chloride (TTC) to red formazan, indicating secretion of invertase. The controls include the untransformed invertase-deficient YTK12 strain and invertase-deficient YTK12 carrying the pSUC2 vector.

3.2.4. Invertase secretion assay using a colorimetric test

In addition, invertase secretion was confirmed with an enzymatic activity test based on reduction of the dye 2,3,5-triphenyltetrazolium chloride (TTC) to the insoluble red colored 1,3,5-triphenylformazan (TPF) (Fig. 3.5) (Klotz, 2004). TTC (2,3,5-triphenyltetrazolium chloride) is a colorimetric indicator that detects the enzymatic invertase activity products glucose and fructose in TTC-treated yeast culture filtrates (Klotz, 2004; Vitolo and Borzani, 1983). I found that the TTC-treated culture filtrates in both negative controls: 1) invertase-deficient yeast strain YTK12 and 2) the invertase-deficient yeast YTK12 transformed with pSUC2 empty vector; remained colorless (Fig. 3.5). In contrast, the all four PexRD constructs enabled the invertase-deficient yeast strain YTK12 to secrete invertase and generate glucose in the presence of sucrose which resulted in the TTC-treated culture filtrates change of colorless to dark red in about 6 minutes.

3.3. Conclusions

Secretory effector proteins of oomycete pathogens alter the host cell environment by triggering or suppressing the immune system of the host (Schornack et al., 2009; Stassen and Van den Ackerveken, 2011). Bioinformatic identification and functional validation of *in planta*-induced secretory proteins carrying RXLR translocation motifs in *P. infestans* with putative roles during pathogen infection will lead to the discovery of large set of potential candidate effectors and their functions. With this study I showed that signal peptides of four representative RXLR effectors of *P. infestans* are functional in yeast and confirmed earlier observations that predictions obtained with the SignalPv2.0 program are highly accurate (Lee et al., 2006; Menne et al., 2000; Schneider and Fechner, 2004). These findings also support putative additional *in planta* effects of the four validated RXLR effectors. For example, the inhibition of secretion of plant proteases by PexRD40/AVRblb2 that is currently under investigation (Tolga Bozkurt, unpublished).

CHAPTER 4: The serine and cysteine protease inhibitor effector families are conserved across diverse pathogenic oomycetes

4.1. Introduction

Plant pathogenic oomycetes secrete an arsenal of effector proteins acting in the intracellular or extracellular space to reprogram the host and enable parasitic infection (Kamoun, 2006, 2007). Protease inhibitors are secreted in the extracellular space (apoplastic effector proteins) where they interact and inhibit plant proteases to repress or induce defence reactions (Schornack et al., 2009; Song et al., 2009). The presence of protease inhibitors in oomycetes was first described in the potato late blight pathogen *Phytophthora infestans* with two major structural classes: (1) Kazal-like serine protease inhibitors (EPIs) (14 proteins) and (2) cystatin-like cysteine protease inhibitors (EPICs) (6 proteins) (Kamoun, 2006; Song et al., 2009; Tian et al., 2005; Tian et al., 2004; Tian and Kamoun, 2005; Tian et al., 2007). Further studies in various oomycete pathogens based on transcriptome analysis described the identification of related genes encoding extracellular protease inhibitors from both structural classes in the sunflower downy mildew *Plasmopara halstedii,* the root rot pathogen *Aphanomyces euteiches,* the fish pathogen *Saprolegnia parasitica* and the broad host range pathogen *Pythium ultimum* (Bouzidi et al., 2007; Cheung et al., 2008; Gaulin et al., 2008; Torto-Alalibo et al., 2005). The genome sequence of the oomycetes pathogens *P. infestans (Pi)*, *P. ultimum (Pu), S. parasitica (Sp)*, *Hyaloperonospora arabidopsidis (Hpa)* and *Albugo laibachii (Al)* offers the opportunity to extend the annotation of novel or existing effector families in these genomes (Baxter et al., 2010; Haas et al., 2009; Levesque et al., 2010) (see chapter 1 Table 1.1). Here, I report the identification of 24 additional protease inhibitors of *P. infestans (Pi)* and their gene expression patterns *in planta*. I investigated the expression patterns of a total of 41 protease inhibitors of *Pi* and found that 30 out of 41 were induced at early and/or late stages of infection in potato and/or tomato suggesting a putative role in counter-defense for the

majority of the members of these families. Also, I predicted a total of 21, 14, 5 and 7 protease inhibitors in *Pu, Sp, Hpa* and *Al,* respectively. These findings confirm previous observations that protease inhibitors of both structural classes are common features of oomycetes pathogens probably because they provide a powerful counter-defense mechanism to target a diverse set of host proteases. In *Pi* and six other pathogenic oomycetes, serine protease inhibitor proteins can contain several tandemly arranged Kazal-like domains. I found variations in the structure of the Kazal-like domains with domains that lack cysteines in position 3 (Cys_3) and 6 (Cys_6). This specific variation of cysteines in the Kazal-like domains was only detected in *Phytophthora* and not in other oomycetes analyzed in this study.

Obligate biotroph parasites are hypothesized to activate less defense responses than non-obligate parasites by modifications (or reprogramming) of the host cell that result in less proteases being produced by the host. In that case, the need of counter-defense protease inhibitors in the pathogen might also be reduced. Consistent with this hypothesis, I found that protease inhibitors, particularly Kazal-like inhibitors are less abundant in the obligate parasites *Hpa* and *Al*.

4.2. Results and discussion

4.2.1. Protease inhibitors of *Phytophthora infestans* and their expression patterns *in planta*

Phytophthora infestans, the potato and tomato late blight hemibiotroph oomycete pathogen, secretes two major structural classes of extracellular protease inhibitor proteins: (1) Kazal-like serine protease inhibitors (EPIs) and (2) cystatin-like cysteine protease inhibitors (EPICs) (Tian et al., 2005; Tian et al., 2004; Tian and Kamoun, 2005; Tian et al., 2007). Both classes of extracellular protease inhibitors effectors in *P. infestans* are transcriptionally induced during pre-infection stages (germinated cyst) and early stages of infection of potato, suggesting a role during host colonization (Haas et al., 2009; Judelson et al., 2008; Randall et al., 2005; Tian et al., 2004). Prior to the genome sequence of *P. infestans*, based on

expressed sequence tags (ESTs), analyses revealed the presence of 19 extracellular protease inhibitors, 14 containing Kazal-like (EPI) and 6 containing cystatin-like (EPIC) domains (Song et al., 2009; Tian et al., 2004; Tian et al., 2007). Annotation of the complete genome sequence of *P. infestans* revealed a total of 41 extracellular protease inhibitors, 33 containing Kazal-like (EPI) and 8 containing cystatin-like (EPIC) domains (Haas et al., 2009) (Table 4.1). Therefore, analysis of *P. infestans* genome sequence allowed the identification of protease inhibitors genes that were not predicted in previous studies. For example, *epi11* was initially predicted to encode for three Kazal-like domains and potentially more, due to an incomplete open reading frame (ORF) (Tian et al., 2004). Based on the *P. infestans* genome sequence, *epi11* ORF was completed and predicted to encode for seven Kazal-like domains, the largest number of Kazal-domains among all EPIs in *P. infestans* (Table 4.1).

EPI1 and EPI10 are two Kazal-like protease inhibitors with a role in *P. infestans*-host interactions having the property of binding and inhibiting the pathogenicity related proteins (PR) P69 subtilisin serine-like protease of tomato (Tian et al., 2005; Tian et al., 2004). In addition, *P. infestans* EPIC2B is a cystatin-like protease inhibitor that binds and inhibits the plant papain-like extracellular Cys protease (PIP1, *Phytophthora* Inhibited Protease 1) (Tian et al., 2007). *P. infestans* protease inhibitors, EPI1, EPI10 and EPIC2B and their host plant targets P69B and PIP1, respectively are induced during infection in tomato, which suggest an important role in defense and counter-defense during *P. infestans*-host interaction (Tian et al., 2005; Tian et al., 2004; Tian et al., 2007). I carried out a microarray analysis of a time course infection, including early and late stages of infection on potato and tomato (see chapter 2 section 2.5.1), and exploited this data to investigate the expression patterns for the 41 genes encoding protease inhibitors in *P. infestans*. The majority of *P. infestans* protease inhibitors from both structural classes are induced during infection. I found that 23 out of the 33 Kazal-like *epi* genes and 7 out of the 8 cystatin-like *epiC* genes were induced at early/late stages on potato/tomato, respectively. In summary, 73% (30 out of 41) protease inhibitors genes of both families in *P. infestans* were induced *in planta* (Table 4.1). Besides protease inhibitor genes many other effectors reside in the repeat-rich gene-sparse regions which are regions

enriched in genes with fast-evolving features and genes that are induced *in planta* (see details in chapter 5 section 5.2.4, Table 5.1, Fig. 5.5) (Haas et al., 2009). Therefore the genes encoding protease inhibitors annotated in this chapter in Table 4.1 are likely to be rapidly evolving genes and this features could be have a beneficial effect to the pathogen by having protease inhibitor effectors with better inhibition affinity to new host proteases (Haas et al., 2009; Judelson et al., 2008; Randall et al., 2005; Tian et al., 2004).

Table 4.1. *P. infestans* secreted protease inhibitor effector families and their expression *in planta*

Gene ID	Secreted	Gene name	Type of domain	No. of domains	P1 residue	Comments	Gene induction* on	
							Potato	Tomato
PITG_22681	Yes	*epi1*	Kazal-like	2	Asp, Asp	Complete ORF	6-16hpi, 2dpi	-
PITG_23119	Yes	*epi1-like*	Kazal-like	2	Asp, Asp	Complete ORF	6-16hpi	-
PITG_01369	Yes	*epi2*	Kazal-like	2	Asp, Asp	Complete ORF	-	-
PITG_22936	Yes	*epi2-like1*	Kazal-like	2	Ala, Asp	Complete ORF	2dpi	3dpi
PITG_22692	Yes	*epi2-like2*	Kazal-like	1	His	Complete ORF	-	-
PITG_16827	Yes	*epi3*	Kazal-like	1	Glu	Complete ORF	-	-
PITG_12131	Yes	*epi4*	Kazal-like	3	Thr, Asp, Asp	Complete ORF	16hpi	-
PITG_22995	Yes	*epi5*	Kazal-like	1	Arg	Complete ORF	-	-
PITG_22739	Yes	*epi5-like*	Kazal-like	1	Asp	Complete ORF	6-16hpi	-
PITG_05440	Yes	*epi6*	Kazal-like	3	Gln, Asp, Asp	Complete ORF	2-3dpi	2-3dpi
PITG_05437	Yes	*epi6-like1*	Kazal-like	3	Gln, Asp, Asp	Complete ORF	2-3dpi	2-3dpi
PITG_22171	Nd	*epi6-like2*	Kazal-like	3	Ala, Ala, Asp	Misannotated ORF, upstream start codon	16hpi, 2-3dpi	2-3dpi
PITG_05430	Yes	*epi6-like3*	Kazal-like	3	Lys, Asp, Asp	Complete ORF	2dpi	-
PITG_22950	Yes	*epi7*	Kazal-like	1	Asp	Complete ORF	6-16hpi	-
PITG_11898	Yes	*epi7-like*	Kazal-like	1	Asp	Complete ORF	6-16hpi, 2dpi	2-3dpi
PITG_23032	Yes	*epi8*	Kazal-like	2	Asp, Asp	Complete ORF	-	-
PITG_13292	Yes	*epi9*	Kazal-like	1	Arg	Complete ORF	-	-
PITG_23195	Yes	*epi9-like*	Kazal-like	1	Arg	Complete ORF	6-16hpi	-
PITG_12129	Yes	*epi10*	Kazal-like	3	Asp, Asp, Asp	Complete ORF	6hpi	-
PITG_07096	Yes	*epi11*	Kazal-like	7	Asp, Lys, Glu, Glu, Glu, Glu, Ala	Complete ORF	6-16hpi	-
PITG_07452	Yes	*epi12*	Kazal-like	1	Ser	Complete ORF	16hpi, 2-3dpi	2-3dpi
PITG_22920	Yes	*epi12-like*	Kazal-like	1	Asp	Complete ORF	6hpi	-
PITG_11899	No	*epi15*	Kazal-like	1	Asp	Complete ORF	6-16hpi, 2dpi	3dpi
PITG_07094	Yes	*epi16*	Kazal-like	1	Asp	Complete ORF	6-16hpi	-
PITG_07095	Yes	*epi16-like*	Kazal-like	1	Gln	Complete ORF	6-16hpi	-
PITG_12138	Yes	*epi17*	Kazal-like	2	Asp, Met	Complete ORF	6-16hpi, 2dpi	2-3dpi
PITG_14708	Yes	*epi18*	Kazal-like	2	Leu, Gln	Complete ORF	-	-
PITG_23178	Yes	*epi19*	Kazal-like	3	Ala, Arg, Tyr	Misannotated ORF, downstream start codon	-	-
PITG_22942	Yes	*Kazal-like1*	Kazal-like	1	Pro	Complete ORF	6-16hpi	-
PITG_22940	Yes	*Kazal-like2*	Kazal-like	1	Pro	Complete ORF	-	-
PITG_22941	Yes	*Kazal-like3*	Kazal-like	1	Pro	Complete ORF	6-16hpi	-
PITG_23147	Yes	*Kazal-like4*	Kazal-like	1	Pro	Complete ORF	6hpi	-
PITG_23012	Yes	*Kazal-like5*	Kazal-like	1	Pro	Complete ORF	-	-
PITG_09169	Yes	*epiC1*	cystatin-like	1	Na	Complete ORF	6hpi, 2dpi	-
PITG_09175	Yes	*epiC2A*	cystatin-like	1	Na	Complete ORF	6hpi, 2dpi	-
PITG_09173	Yes	*epiC2B*	cystatin-like	1	Na	Complete ORF	2-3dpi	2-3dpi
PITG_14891	Yes	*epiC3*	cystatin-like	1	Na	Complete ORF	6hpi	-
PITG_00058	Yes	*epiC4*	cystatin-like	1	Na	Complete ORF	-	-
PITG_13320_NS	No	*epiC5*	cystatin-like	1	Na	Complete ORF	6-16hpi	-
PITG_14924	Yes	*epiC6*	cystatin-like	1	Na	Complete ORF	6-16hpi	-
PITG_22881	Yes	*epiC-like*	cystatin-like	1	Na	Complete ORF	6hpi	-

* Gene induction in Log2 (*in planta* expression relative to mycelia) in hours post infection (hpi) and days post infection (dpi) (see chapter 2 section 2.5.1). Nd, Not determined, this is because secretion signal could be only estimated once the 5' end sequence is obtained (not presented here); NS, Not secreted; Na, not applied this is because P1 residues are only present from proteins containing Kazal-like domains). Secretion signals predicted with SignalPv2.0 program (see chapter 2 section 2.2) (Nielsen et al., 1997).

The oomycete domain structure of Kazal-like inhibitors usually follows the conserved motif C-$X_{3,4}$-C-X_7-C-X_6-Y-X_3-C-X_6-C$X_{9,12,13,14}$-C (Tian et al., 2004). The majority of *P. infestans* Kazal-like EPI proteins contain the 6 conserved cysteines that define the family. However, some multidomain Kazal-like EPI proteins of *P. infestans* like EPI1 and EPI10 were shown to contain atypical Kazal-like domains characterized by the lack of Cys_3 and Cys_6 that result in the formation of two disulfide bridges instead of three (see purple domains with two bridges and blue domains with three bridges in chapter 1 Fig. 1.2) (Tian et al., 2004). Among, all *P. infestans* annotated Kazal-like domains annotated in this chapter; I found that 19 EPI domains lacked the Cys_3 and Cys_6 in their Kazal-like domain structure (see atypical domains in Fig. 4.1A, Fig. 4.2A, see appendix 2.1). These 19 atypical Kazal-like domains with two disulfide bridges occur in 15 *epi* genes, and 12 out 15 *epi* genes were induced *in planta* (Fig. 4.1A, Table 4.1 and Fig. 4.2B). The atypical Kazal-like domains with two disulfide bridges present in EPI1 and EPI10 proteins are predicted to inhibit plant subtilisins, which indicate these atypical domains are functional (Tian et al., 2005).

The specificity of the Kazal-like inhibitor proteins is dictated by the predicted active site P1 (Lu et al., 2001). The P1 residue in *P. infestans* Kazal-like inhibitors was variable with 13 amino acids represented (Asp, Glu, Pro, Arg, Ala, Gln, Lys, Thr, Met, His, Ser, Tyr, Leu) (Table 4.1). In agreement to previous studies, I found that in *P. infestans* half (30 out 60) the P1 residues correspond to aspartate (Asp), an uncommon P1 amino acid in other natural Kazal inhibitors (Table 4.1 and Fig. 4.2B) (Tian et al., 2004).

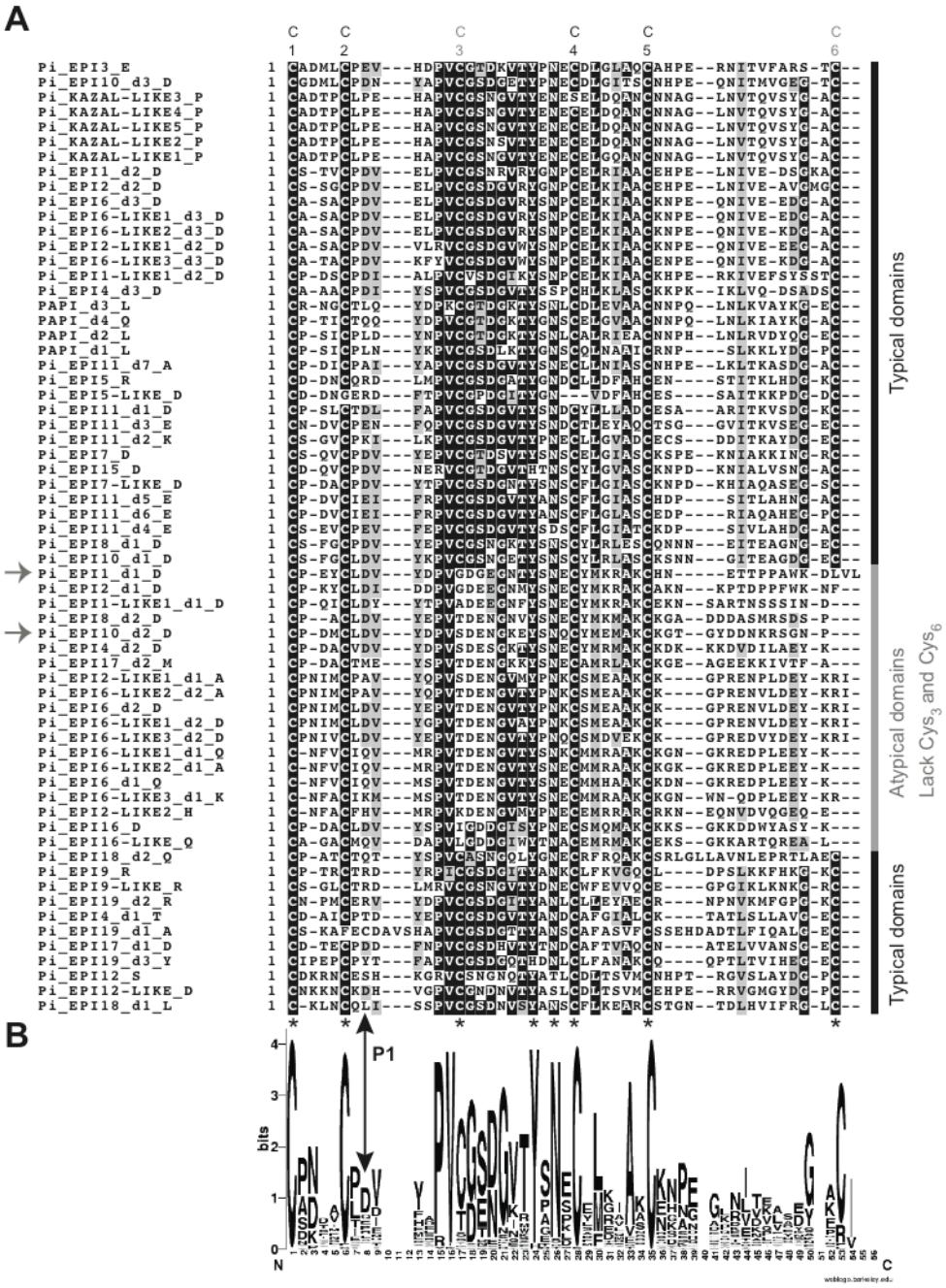


Fig. 4.1. Sequence alignment of 60 Kazal domains of *Phytophthora infestans* and their corresponding consensus sequence pattern

(A) Multiple sequence alignment of 60 EPI domains in *P. infestans* with representative Kazal family inhibitor domains with their predicted P1 residues indicated by the double-headed arrow (see chapter 2 section 2.2). The alignment also includes 4 additional Kazal-like inhibitors from the crayfish *Pacifastus leniusculus* (PAPI-1_d1-d4, CAA56043). The amino acid residues that defined the Kazal-like family protease inhibitor domain are marked with asterisks (bottom). Conserved cysteines and their positions are shown (top). Kazal-like domains with variable Cys_3 and Cys_6 are highlighted with a grey bar on the left. Arrows point to atypical domains d1 of EPI1 and d2 of EPI10 that can inhibit the plant protease P69B (see chapter 1 Fig. 1.2) and that belong to the group with variations in the

Cys_3 and Cys_6 of the Kazal-like domain (Tian et al., 2004). A group of 19 Kazal-like domains, which are atypical, are marked with a grey bar of the left side of the alignment. (B) Consensus sequence pattern of oomycete Kazal domains. Consensus sequence was calculated at http://weblogo.berkeley.edu/logo.cgi. The bigger the letter, the more conserved the amino acid site is for that position. The positions of amino acids in the consensus sequence correspond to the positions in the sequence alignment. The P1 positions are indicated by a double-headed arrow.

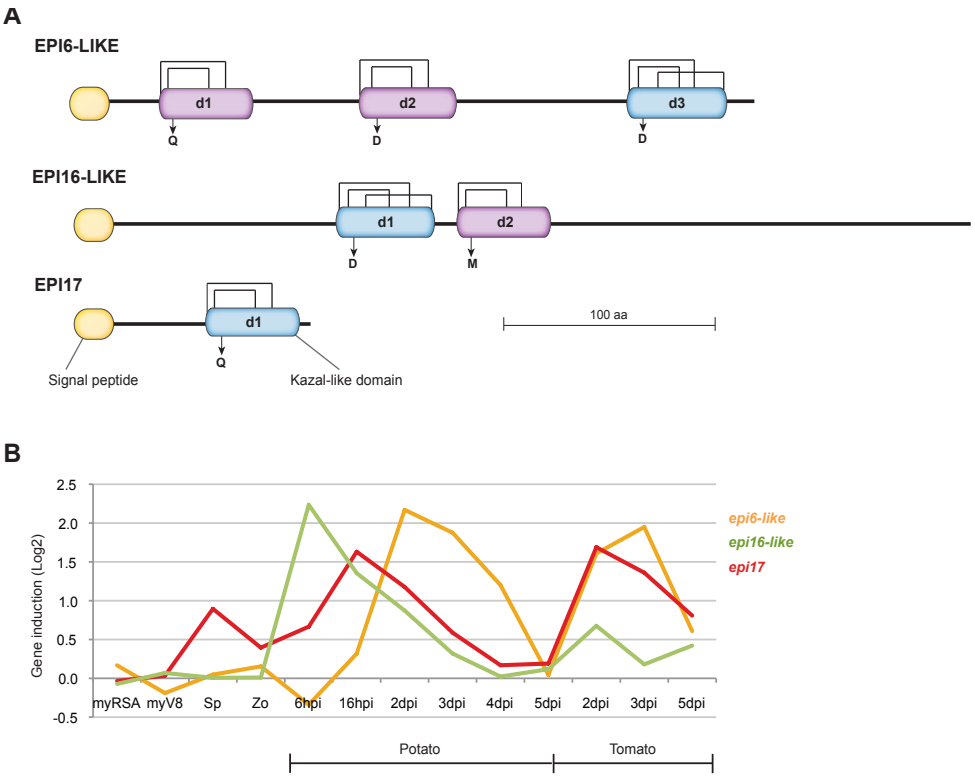


Fig. 4.2. Structure and gene expression Kazal-like EPI inhibitors of *Phytophthora infestans*
(A) Schematic representation of EPI6-like, EPI16-LIKE and EPI17 effector protein domains. The signal peptide is indicated in yellow, the atypical Kazal-like domains are shown purple and the typical domains in blue (see chapter 2 section 2.2). The disulfide linkages predicted based on the structure of other Kazal domains are shown with bars. Note that protease inhibitors can present different types of Kazal-like domains within the same effector protein. For example, EPI6-like has three Kazal-like domains, the first two domains are atypical with only two disulfide bridges and the third domain is typical with three disulfide bridges. EPI16-like has two Kazal-like domains, the first is typical and the second is atypical. EPI17 has only one Kazal-like domain and it is atypical. The positions and amino acid letter for the P1 residues are marked with arrows. (B) Gene expression of Kazal-like inhibitors *epi6-like*, *epi16-like* and *epi17* of *P. infestans.* Line graph shows *in planta* gene induction as log2 estimated for each sample (Sp, Zo, Potato and Tomato time points) relative mycelia (see microarray analysis in chapter 2 section 2.5.1). MyRSA, mycelia in Rye Sucrose Agar (RSA); myV8, mycelia in V8 agar; Sp, Sporangia, Zo, Zoospores, hpi, hours post inoculation; dpi, days post inoculation.

4.2.2. Prediction of protease inhibitors in pathogenic oomycetes

To search for protease inhibitor-encoding genes in the recently sequenced genomes of *Pu, Sp*, *Hpa,* and *Al*, I performed a BLASTP search using *P. infestans* protease inhibitor proteins as queries (see chapter 2 section 2.2). I also did a TBLASTN search to find whether additional protease inhibitor genes, meaning genes not covered by the original gene models, could be predicted from the scaffolds.

4.2.2.1. Protease inhibitors of *Pythium ultimum*

P. ultimum is a necrotroph oomycete pathogen and one of the most pathogenic *Pythium* species. *P. ultimum* is the causal agent of a variety of diseases, including damping off, and affects multiple monocot and dicot hosts (Martin and Loper, 1999). Previous studies based on transcriptome analysis in *P. ultimum* revealed the presence of two protease inhibitors similar to Kazal-like and cystatin-like of *P. infestans* (Cheung et al., 2008). In this study, I identified 15 proteins in *P. ultimum* with similarity to *P. infestans* Kazal-like serine protease inhibitors: 12 secreted and 3 non-secreted proteins (Table 4.2, see chapter 2 section 2.2 and appendix 2.1). Sequence alignment to other oomycete Kazal-like protease inhibitors showed conservation of the cysteine backbone (Fig. 4.3). I also identified 6 proteins with similarity to *P. infestans* cystatin-like cysteine protease inhibitors: 3 secreted and 3 non-secreted proteins (Table 4.2, see appendix 2.2). Sequence alignment of their putative cystatin-like inhibitor domains highlights the conserved amino acids in the N-terminal trunk (NT) and loop1 (L1) and loop 2 (L2) domains (Fig. 4.5).

In *P. infestans* there was a wide distribution of P1 residues and in *P. ultimum* the most common residues were Asp, Ala, Glu and Met (Fig. 4.4 and Table 4.2). This suggests that there is also diversity in specificities of Kazal-like inhibitors of *P. ultimum*, which could have implications in the ability to inhibit multiple proteases and successfully infect a wide range of hosts.

Table 4.2. Predicted protease inhibitor effector families in *P. ultimum* genome

Gene ID	Secreted	Type of domain	No. of domains	P1 residue	Comments
Pu_PYU1_T010209	Yes	Kazal-like	5	Ala, Lys, Met, Ala, Lys	Complete ORF
Pu_PYU1_T009699	Yes	Kazal-like	4	Asp, Leu, Arg, Ser	Complete ORF
Pu_PYU1_T000142	Yes	Kazal-like	4	Glu, Ser, Lys, Thr	Complete ORF
Pu_PYU1_T009700	Yes	Kazal-like	3	Met, Asp, Pro	Complete ORF
Pu_PYU1_T000511_NS	No	Kazal-like	3	Met, Asp, Gln	Complete ORF
Pu_PYU1_T013339	Yes	Kazal-like	2	Ala, Arg	Complete ORF
Pu_PYU1_T012159	Yes	Kazal-like	2	Val, Glu	Complete ORF
Pu_PYU1_T012158	Yes	Kazal-like	2	Val, Glu	Complete ORF
Pu_PYU1_T012161	Yes	Kazal-like	2	Ala, Asp	Complete ORF
Pu_PYU1_T014337	Yes	Kazal-like	2	Val, Leu	Complete ORF
Pu_PYU1_T012160	Yes	Kazal-like	2	Gly, Asp	Complete ORF
Pu_PYU1_T014335	Yes	Kazal-like	2	Ala, Met	Complete ORF
Pu_PYU1_T005024_NS	No	Kazal-like	2	Leu, Glu	Complete ORF
Pu_PYU1_T012156_NS	No	Kazal-like	1	Ser	Complete ORF
Pu_PYU1_T012157	Yes	Kazal-like	1	Thr	Complete ORF
Pu_PYU1_T011854	Yes	cystatin-like	1	na	Complete ORF
Pu_PYU1_T012817_NS	No	cystatin-like	1	na	Complete ORF
Pu_PYU1_T012816	Yes	cystatin-like	1	na	Complete ORF
Pu_PYU1_T012805_NS	No	cystatin-like	1	na	Complete ORF
Pu_PYU1_T011856_NS	No	cystatin-like	1	na	Complete ORF
Pu_PYU1_T012815	Yes	cystatin-like	1	na	Complete ORF

NS, Not secreted; Na, not applied this is because P1 residues are only present from proteins containing Kazal-like domains). Secretion signals predicted with SignalPv2.0 program (see chapter 2 section 2.2) (Nielsen et al., 1997).

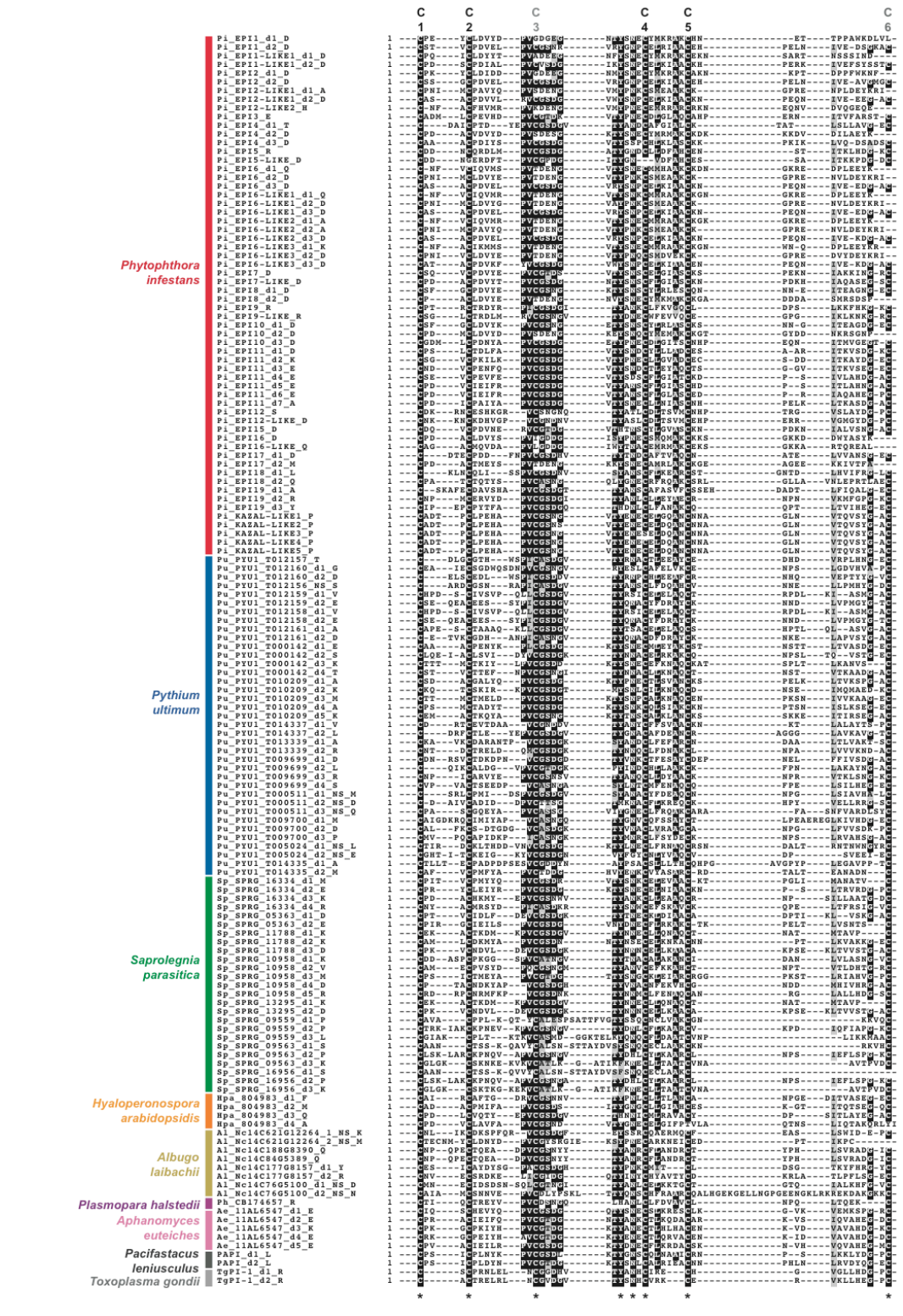


Fig. 4.3. Sequence alignment of 137 Kazal-like domains of seven pathogenic oomycetes

Multiple sequence alignment of 140 Kazal-like domains present in 64 serine-like protease inhibitors (EPIs) of seven pathogenic oomycetes. Out of the 140 oomycete Kazal-like

domains of oomycetes, 60 are from *Phytophthora infestans (Pi)*, 37 are from *Pythium ultimum (Pu),* 25 are from *Saprolegnia parasitica (Sp),* 4 are from *Hyaloperonospora arabidopsidis (Hpa),* 8 are from *Albugo laibachii (Al),* 1 is from *Plasmopara halstedii (Ph)* and 5 are from *Aphanomyces euteiches (Ae)*. The alignment also includes 8 additional known Kazal-like domains present in 2 protease inhibitors from crayfish and a protozoan parasite species, respectively. Out of the 8 additional Kazal-like inhibitors, 4 are from the crayfish *Pacifastus leniusculus* (PAPI-1_d1-d2, CAA56043) and 4 from the apicomplexan protozoan parasite *Toxoplasma gondii* (TgPI-1_d1-d2, AF121778). Appendix 2.1 contains the list of the137 Kazal-like domain sequences used in this alignment. The amino acid residues that defined the Kazal-like family protease inhibitor domain are marked with asterisks (bottom). The conserved cysteines and their position are numbered in the alignment (top). Cysteine positions three and six shown in grey are missing in some protease inhibitors domains of *P. infestans*. The first suffix indicates the number of the Kazal-like domain from left to right of the C-terminal effector domain in multidomain proteins. The second suffix indicates the P1 amino acid residue, which is the central to the specificity of Kazal inhibitors (Lu et al., 2001). The third suffix "NS" if present, indicates protease inhibitor domains from proteins not predicted to be secreted.

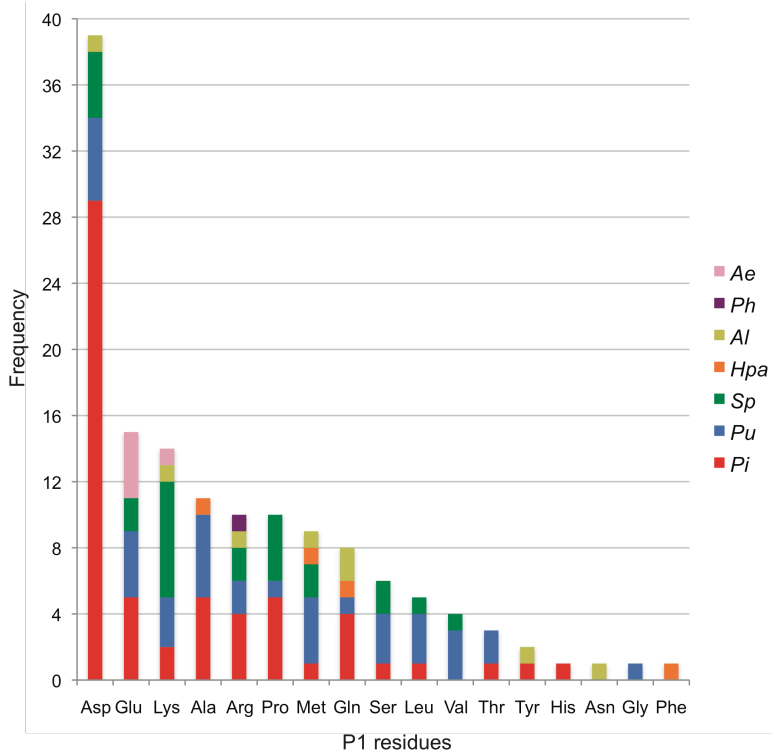


Fig. 4.4. Distribution of P1 residues among seven oomycete Kazal-like domains
Frequency represents the number of Kazal-like domains containing a given amino acid residue at the P1 position.

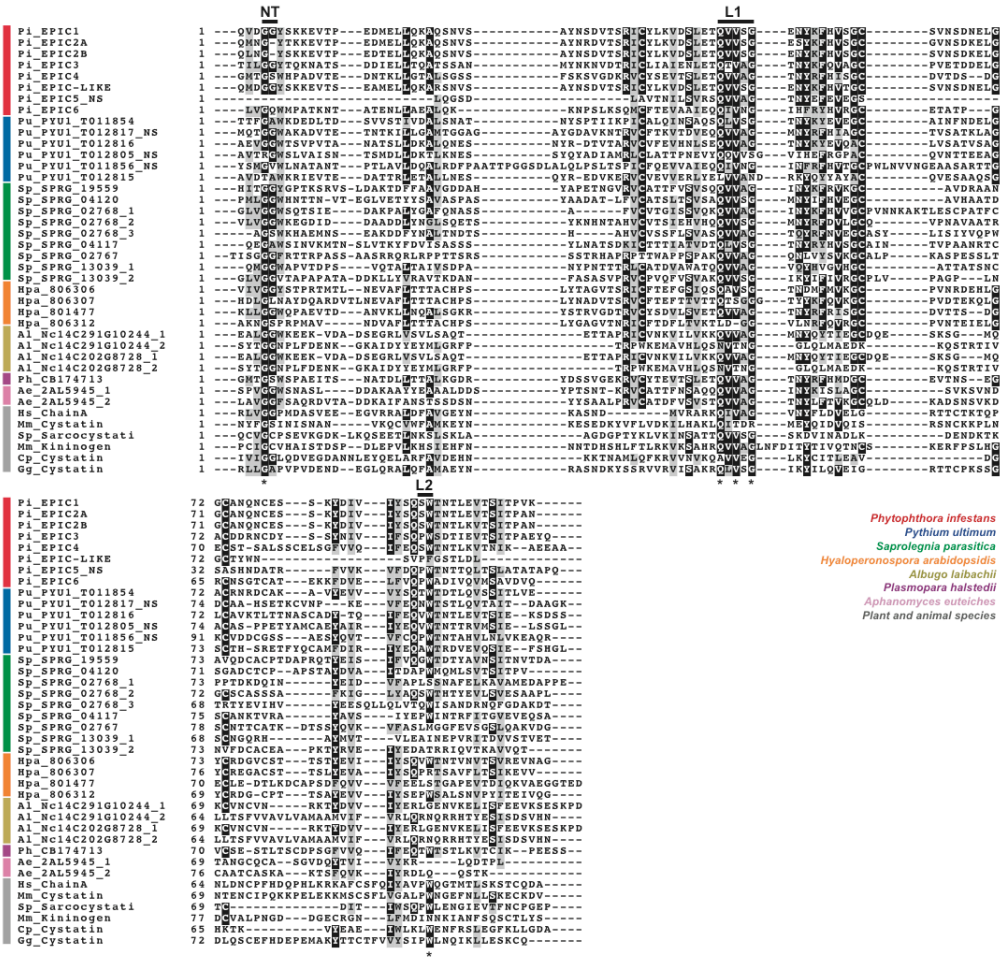


Fig. 4.5. Sequence alignment of 34 cystatin-like domains of seven pathogenic oomycetes

Multiple sequence alignment of 34 cystatin-like domains present in 28 cysteine protease inhibitors (EPICs) of seven pathogenic oomycetes (see chapter 2 section 2.2). Out of the 34 oomycete cystatin-like domains, 8 are from *Phytophthora infestans (Pi)*, 6 are from *Pythium ultimum (Pu),* 9 are from *Saprolegnia parasitica (Sp),* 4 are from *Hyaloperonospora arabidopsidis (Hpa),* 4 are from *Albugo laibachii (Al),* 1 is from *Plasmopara halstedii (Ph)* and 2 are from *Aphanomyces euteiches (Ae)*. The alignment also includes 6 cystatin-like domains present in 6 cysteine protease inhibitors from plants (*Carica papaya* Cp_cystatin gi|311505), from animals (insect *Sarcophaga peregrina* Sp_Sarcocystatin gi|399335, chicken Gg_cystatin P01038, mouse Mm_cystatin gi|6226846 Mm_Kininogen gi|12643495, human Hs_Chain A gi14278690). Appendix 2.2 contains the list of the 34 cystatin-like domain sequences used in this alignment. The proposed active-site residues in cystatins, including the N-terminal trunk (NT), first binding loop (L1) and second binding loop (L2) are indicated in the sequence with a bar (top). The amino acids that define cystatins are marked with asterisks (bottom). The first suffix indicates the number of the cystatin-like domain from left to right of the C-terminal effector domain in multidomain proteins. The second suffix "NS" was added to protease inhibitor domains from proteins that were not predicted to be secreted.

4.2.2.2. Protease inhibitors of *Saprolegnia parasitica*

S. parasitica is an opportunistic oomycete pathogen of fish (both saprophytic and necrotrophic growth) and one of the most important pathogens on salmon and trout species (Hatai and Hoshiai, 1994). Previous studies based on transcriptome analysis of *S. parasitica* showed the presence of two secreted proteins classified as one Kazal-like and one of the cystatin-like protease inhibitors similar to those reported in *P. infestans* (Torto-Alalibo et al., 2005). In this study, I identified 8 secreted proteins in *S. parasitica* with similarity to *P. infestans* Kazal-like serine protease inhibitors (Table 4.3, see appendix 2.1). Sequence alignment to other oomycete Kazal-like protease inhibitors showed conservation of the six-cysteine backbone (Fig. 4.3).

The most common amino acids for the P1 residues in *S. parasitica* Kazal-like inhibitors were Lys, Asp and Pro. Lys P1 residue is present in *Toxoplasma gondii* Kazal inhibitor with trypsin inhibition specificity (Fig. 4.4 and Table 4.3). The skin mucus of many fish species contains trypsin-like activity with the ability to lyse dead bacterial cells, suggesting a role in defence (Aranishi and Mano, 2000; Hjelmeland, 1983). It is possible that Kazal-like proteins of *S. parasitica* with Lys are putative inhibitors of trypsin proteases in fish and this hypothesis could be explored in the future.

In addition, I identified 6 secreted proteins with similarity to *P. infestans* cystatin-like cysteine protease inhibitors (Table 4.3, see appendix 2.2). Sequence alignment of their putative cystatin-like inhibitor domains highlights the conserved amino acids in the N-terminal trunk (NT) and loop1 (L1) and in some of them the loop 2 (L2) domains (Fig. 4.5).

Table 4.3. Secreted protease inhibitor effector families predicted in *S. parasitica* genome

Gene ID	Other gene name	Type of domain	No. of domains	P1 residue	Comments
Sp_SPRG_10958	-	Kazal-like	5	Lys, Val, Met, Asp, Arg	Complete ORF
Sp_SPRG_16334	-	Kazal-like	4	Met, Glu, Lys, Arg	Complete ORF
Sp_SPRG_09559	Sp_001_0127[a]	Kazal-like	3	Pro, Pro, Leu	Complete ORF
Sp_SPRG_09563	-	Kazal-like	3	Ser, Pro, Lys	Complete ORF
Sp_SPRG_16956	-	Kazal-like	3	Ser, Pro, Lys	Incomplete ORF, missing stop codon
Sp_SPRG_11788	-	Kazal-like	3	Lys, Lys, Glu	Complete ORF
Sp_SPRG_05363	-	Kazal-like	2	Asp, Glu	Complete ORF
Sp_SPRG_13295	-	Kazal-like	2	Lys, Asp	Complete ORF
Sp_SPRG_19559	Sp_001_01374[a]	cystatin-like	1	na	Complete ORF
Sp_SPRG_04120	-	cystatin-like	1	na	Complete ORF
Sp_SPRG_02768	-	cystatin-like	3	na	Complete ORF
Sp_SPRG_04117	-	cystatin-like	1	na	Complete ORF
Sp_SPRG_02767	-	cystatin-like	1	na	Complete ORF
Sp_SPRG_13039	-	cystatin-like	2	na	Complete ORF

[a] Previously reported protease inhibitor sequence (Torto-Alalibo et al., 2005). Na, not applied this is because P1 residues are only present from proteins containing Kazal-like domains). All proteins listed in this table are predicted to be secreted using SignalPv2.0 program (see chapter 2 section 2.2) (Nielsen et al., 1997).

4.2.2.3. Protease inhibitors of *Hyaloperonospora arabidopsidis*

H. arabidopsidis (Hpa), an obligate biotrophic parasite, and its natural host *Arabidopsis thaliana* are widely used as a model pathosystem for downy mildew pathogens (Slusarenko and Schlaich, 2003). In this study, I identified only one secreted protein in *H. arabidopsidis* with similarity to *P. infestans* Kazal-like serine protease inhibitors (Table 4.4, see chapter 2 section 2.2 and appendix 2.1). Sequence alignment to other oomycete Kazal-like protease inhibitors showed conservation of the six-cysteine backbone (Fig. 4.3). I also identified 4 secreted proteins with similarity to *P. infestans* cystatin-like cysteine protease inhibitors (Table 4.4, see chapter 2 section 2.2 and appendix 2.2). Sequence alignment of their putative cystatin-like inhibitor domains highlights the conserved amino acids for some of them in the N-terminal trunk (NT) and loop1 (L1) and loop 2 (L2) domains (Fig. 4.5).

Table 4.4. Secreted protease inhibitor effector families predicted in *H. arabidopsidis* genome

Gene ID	Type of domain	No. of domains	P1 residue	Comments
Hpa_804983	Kazal-like	4	Phe, Met, Gln, Ala	Complete ORF
Hpa_806306	cystatin-like	1	na	Complete ORF
Hpa_806307	cystatin-like	1	na	Complete ORF, start codon misannotated
Hpa_801477	cystatin-like	1	na	Complete ORF
Hpa_806312	cystatin-like	1	na	Complete ORF

Na, not applied this is because P1 residues are only present from proteins containing Kazal-like domains). All proteins listed in this table are predicted to be secreted using SignalPv2.0 program (see chapter 2 section 2.2) (Nielsen et al., 1997).

4.2.2.4. Protease inhibitors of *Albugo laibachii*

A. laibachii (Al) is another obligate biotrophic oomycete, recently described as highly specialized on *Arabidopsis thaliana* (Slusarenko and Schlaich, 2003; Thines et al., 2009). In this study, I identified 5 secreted proteins in *A. laibachii* with similarity to *P. infestans* Kazal-like serine protease inhibitors (Table 4.5, see chapter 2 section 2.2 and appendix 2.1). Sequence alignment to other oomycete Kazal-like protease inhibitors showed conservation of the six-cysteine backbone (Fig. 4.3). I also identified 2 secreted proteins with similarity to *P. infestans* cystatin-like cysteine protease inhibitors (Table 4.4, see chapter 2 section 2.2 and appendix 2.2). Sequence alignment of their putative cystatin-like inhibitor domains highlights the conserved amino acids for some of them in the N-terminal trunk (NT) and loop1 (L1) and loop 2 (L2) domains (Fig. 4.5).

Table 4.5. Secreted protease inhibitor effector families predicted in *A. laibachii* genome

Gene ID	Secreted	Type of domain	No. of domains	P1 residue	Comments
Al_Nc14C621G12264_NS	No	Kazal-like	2	Lys, Met	Complete ORF
Al_Nc14C76G5100_NS	No	Kazal-like	2	Asp, Asn	Complete ORF
Al_Nc14C177G8157	Yes	Kazal-like	2	Tyr, Arg	Complete ORF
Al_Nc14C188G8390	Yes	Kazal-like	1	Gln	Complete ORF, start codon misannotated
Al_Nc14C84G5389	Yes	Kazal-like	1	Gln	Complete ORF, start codon misannotated
Al_Nc14C291G10244	Yes	cystatin-like	2	na	Complete ORF
Al_Nc14C202G8728	Yes	cystatin-like	2	na	Complete ORF

Na, not applied this is because P1 residues are only present from proteins containing Kazal-like domains). All proteins listed in this table are predicted to be secreted using SignalPv2.0 program (see chapter 2 section 2.2) (Nielsen et al., 1997).

4.2.3. Comparative analysis of oomycetes protease inhibitors

Kazal-type serine protease (EPI) inhibitors are single or multi-domain proteins with domains that usually have different specificities towards a particular protease, with the P1 residue contributing to this specificity (Lu et al., 2001). Although, aspartic acid is the most abundant P1 residue of Kazal-like inhibitors (EPIs) of *P. infestans*, the P1 residue can be variable within *P. infestans* and across various oomycetes studied in this chapter (Fig. 4.4) (Tian et al., 2005).

Phylogenetic analysis of the Kazal-like domains revealed that atypical domains with two disulfide bridges that lack Cys_3 and Cys_6 were present in *P. infestans* but not in other oomycete species (Fig. 4.6 and Fig. 4.3). Interestingly, atypical domains are also present in two other *Phytophthora* species, *Phytophthora ramorum* and *Phytophthora sojae* (Miaoying Tian, unpublished), besides *P. infestans* (Tian and Kamoun, 2005). These observations suggest that Kazal-like atypical domains are specific to the *Phytophthora* lineage.

EPI1 and EPI10 are *in planta*-induced genes of *P. infestans* encoding for multidomain Kazal-like protease inhibitors (Tian et al., 2005; Tian et al., 2004). These two protease inhibitors present both atypical and typical Kazal-like domains (see chapter 1 Fig. 1.2 and this chapter Fig. 4.1A, Table 4.1) (Tian et al., 2005). However, only the atypical Kazal-like domains of EPI1 and EPI10

inhibitors have been predicted to inhibit the plant subtilisin A (see chapter 1 Fig. 1.2) (Tian et al., 2005; Tian et al., 2004). It is possible that atypical two-disulfide Kazal-like domains are of importance to the pathogenicity of *Phytophthora.* Besides EPI1 and EPI10, there are 13 other multidomain Kazal-like protease inhibitors in *P. infestans* that have at least one atypical domain (Fig. 4.1). Further experiments to characterize the atypical Kazal-like domains will help to understand the biological functions of the diverse Kazal-like inhibitors in *Phytophthora*.

Sequence analyses of the cystatin-like inhibitors, show that although there are significant amino acid differences in the overall cystatin proteins among the seven oomycetes, their tertiary structures are conserved: N-terminus trunk (NT), Loop1 (L1) containing the highly conserved region (QXVXG) and Loop2 (L2) with the region (PW) (Fig. 4.5). Although the conservation in the tertiary structure, phylogenetic analysis of the cystatin-like inhibitors shows that all animal and plants cystatins formed a distant group compared to oomycete cystatins (Fig. 4.7). I suggest this could be explained by the possibility of having of another conserved region with the motif RXC (an Arg, a variable amino acid Ile/Val/Leu/Met/Pro, and Cysteine) before Loop1 (L1) that is only present in oomycetes and not in plant or animals cystatins (Fig. 4.5). Only two oomycete proteins showed mutated cysteines in this putative RXC motif, Al_Nc14C202G8728_2 and Al_Nc14C202G8728_2. As a consequence these two proteins grouped closer to plant and animal cystatins (Fig. 4.5 and Fig. 4.7). In some mammalian cystatins, a second inhibitory site that lies just before the Loop 1, SND (a Ser, an Asn and Asp) motif is shown to block legumain or asparaginyl endopeptidase (AEP) enzymes (Alvarez-Fernandez et al., 1999). More experiments will help to understand whether the putative RXC motif has a biological function and their relevance in oomycetes.

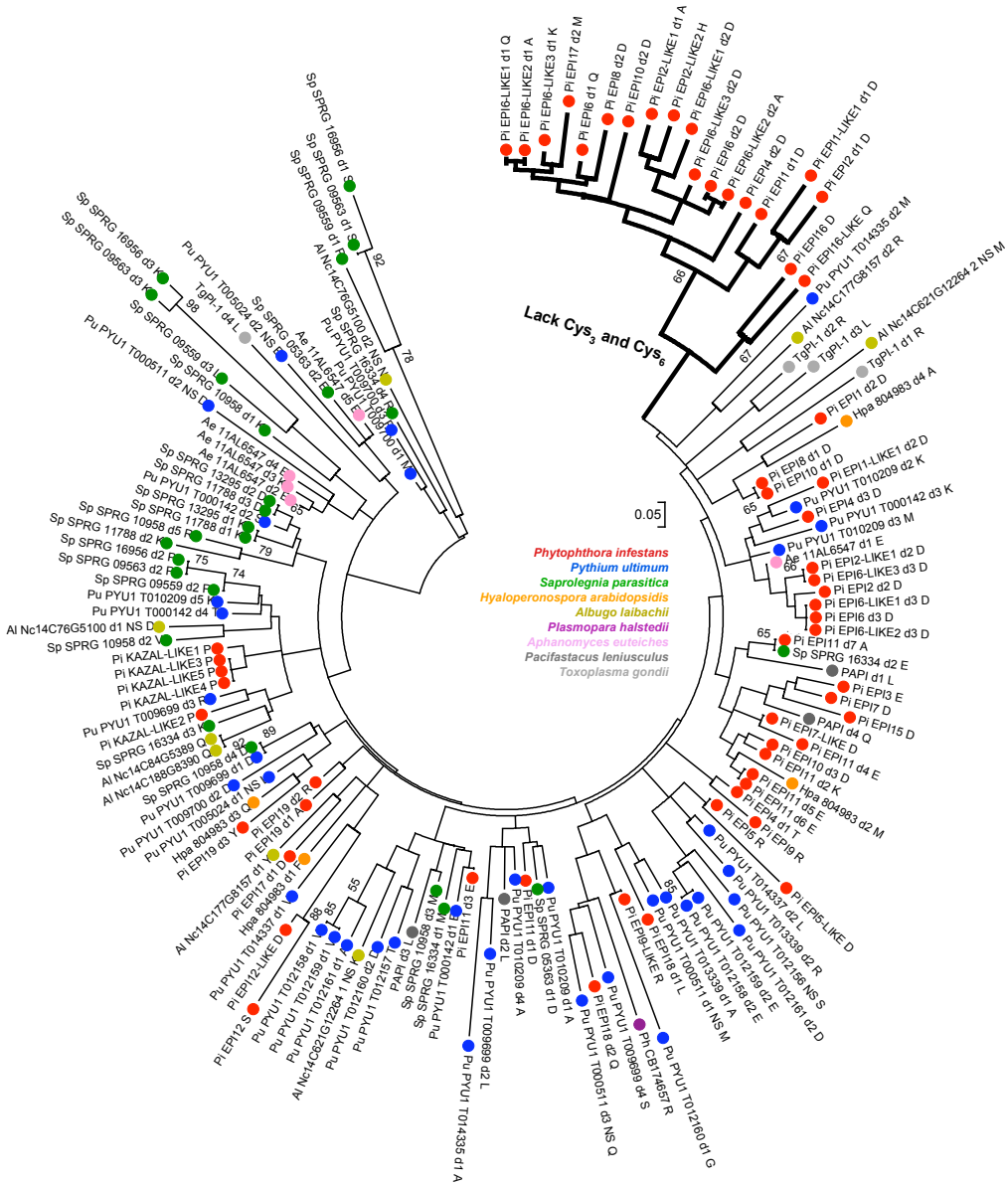


Fig. 4.6. Phylogenetic analysis of 140 Kazal-like domains of seven pathogenic oomycetes

The neighbor-joining tree was constructed with 140 Kazal-like domains present in 64 serine-like protease inhibitors (EPIs) of seven oomycete pathogens (see chapter 2 section 2.2). Out of the 140 oomycete Kazal-domains, 60 are from *Phytophthora infestans (Pi)*, 37 are from *Pythium ultimum (Pu),* 25 are from *Saprolegnia parasitica (Sp),* 4 are from *Hyaloperonospora arabidopsidis (Hpa),* 8 are from *Albugo laibachii (Al),* 1 is from *Plasmopara halstedii (Ph)* and 5 are from *Aphanomyces euteiches (Ae)*. The neighbor-joining tree also includes 8 additional known Kazal-like domains present in 2 protease inhibitors from crayfish and a protozoan parasite species, respectively. Out of the 8 additional Kazal-like inhibitors, 4 are from the crayfish *Pacifastus leniusculus* (PAPI-1_d1-d4, CAA56043) and 4 from the apicomplexan protozoan parasite *Toxoplasma gondii* (TgPI-1_d1-d4, AF121778). Appendix 2.1 contains the list of the 137 Kazal-like domain sequences used in this alignment. The first suffix indicates the number of the Kazal-like domain from left to right of the C-terminal effector domain in multidomain

proteins. The second suffix indicates the P1 amino acid residue, which is the central to the specificity of Kazal inhibitors (Lu et al., 2001). The third suffix “NS” if it is present, indicate proteins are not predicted to be secreted. Group with branches highlighted in black are indicative of Kazal-like protease inhibitor domains of *P. infestans* that lack third and sixth cysteine positions (Tian et al., 2004). Bootstrap values were obtained with 1000 replications and values equal or higher than 50% are shown.

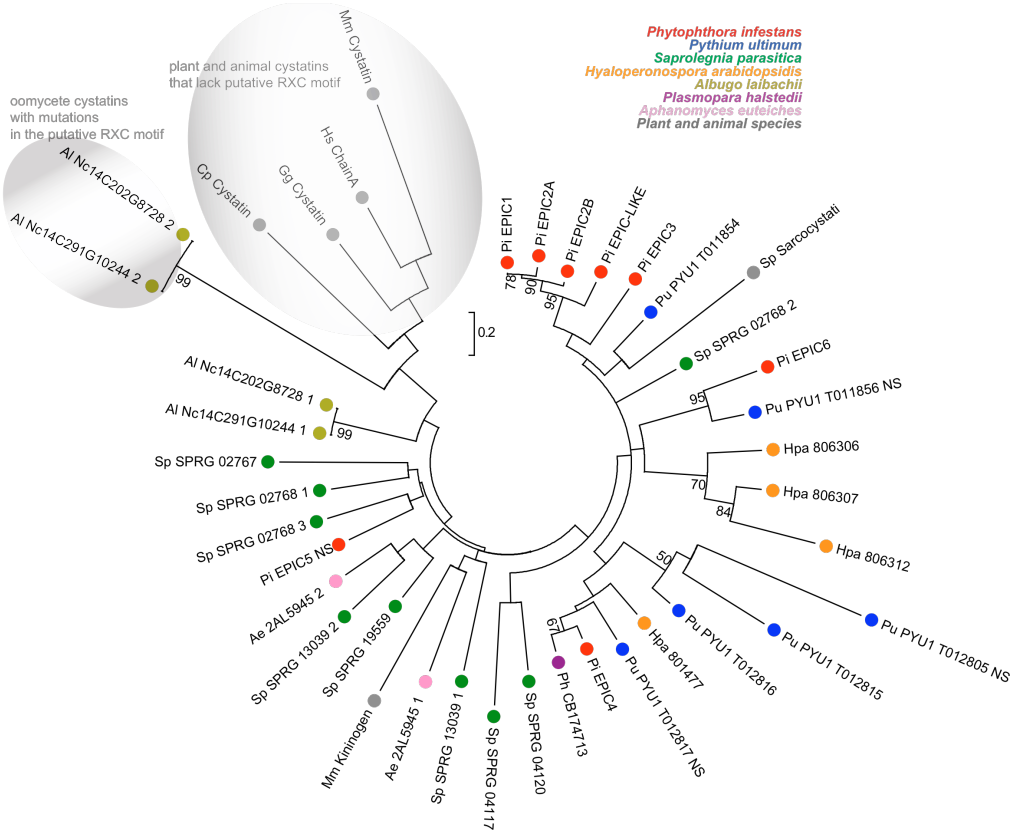


Fig. 4.7. Phylogenetic analysis of 34 predicted cystatin-like domains of seven pathogenic oomycetes

The neighbor-joining tree was constructed with 34 cystatin-like domains present in 28 cysteine protease inhibitors (EPICs) of seven pathogenic oomycetes (see chapter 2 section 2.2). Out of the 34 oomycetes cystatin-like domains shown in the tree, 8 are from *Phytophthora infestans (Pi)*, 6 are from *Pythium ultimum (Pu),* 9 are from *Saprolegnia parasitica (Sp),* 4 are from *Hyaloperonospora arabidopsidis (Hpa),* 4 are from *Albugo laibachii (Al),* 1 is from *Plasmopara halstedii (Ph)* and 2 are from *Aphanomyces euteiches (Ae)*. The neighbor-joining tree also includes 6 cystatin-like domains present in 6 cysteine protease inhibitors from plants (*Carica papaya* Cp_cystatin gi|311505), from animals (insect *Sarcophaga peregrina* Sp_Sarcocystatin gi|399335, chicken Gg_cystatin P01038, mouse Mm_cystatin gi|6226846 Mm_Kininogen gi|12643495, human Hs_Chain A gi14278690). The plant and animal cystatins are highlighted due the absence of a putative RXC motif present in oomycete cystatins. Appendix 2.2 contains the list of the 34 cystatin-like domain sequences used to construct this phylogenetic tree. The first suffix indicates the number of the cystatin-like domain from left to right of the C-terminal effector domain in multidomain proteins. The second suffix “NS” was added to protease inhibitor domains from proteins that were not predicted to be secreted. Group highlighted in a grey

circle correspond to Kazal-like domains that lack both cysteine positions three and six only occurring in *P. infestans* (Tian et al., 2004). Bootstrap values were obtained with 1000 replications and values equal or higher than 50% are shown.

Protease inhibitors of both structural classes were found in all seven oomycete species (Table 4.6). These findings confirm that protease inhibitors are present across a diverse range of pathogenic oomycetes species. By comparing the number of predicted domains, I found multiple Kazal-like domains were present in six out of seven oomycete species, *P. infestans, P. ultimum, Saprolegnia parasitica, Hyaloperonospora arabidopsidis*, *Albugo laibachii* and *Aphanomyces euteiches* (Table 4.6). In contrast multiple cystatin-like domains were only present in two out of seven oomycete species, *S. parasitica* and *Aphanomyces euteiches* (Table 4.6). From these observations, I conclude that tandem duplications of Kazal-like domains are more widespread that duplications of cystatin-like domains in oomycetes.

The *P. ultimum*, *P. infestans* and *S. parasitica* oomycete genomes showed the largest repertoire of protease inhibitors among the species examined, particularly in Kazal-like protease inhibitors compared to the number of protease inhibitors detected in the genomes of *H. arabidopsidis* and *A. laibachii* (Table 4.6). This observation suggests that protease inhibitors are less abundant in number in obligate parasites. Fewer protease inhibitors are also reported in *A. euteiches* and *Plasmopara halstedii*. However, the number of protease inhibitors predicted in these oomycete species is based on expressed sequence tags (ESTs) and not whole-genome analysis, which raises the possibility of additional protease inhibitors in these pathogens (Bouzidi et al., 2007; Gaulin et al., 2008). For example, in *S. parasitica*, only two protease inhibitors, one Kazal-like and one cystatin-like were previously described before, but in this study I found the presence of 14 in total (Torto-Alalibo et al., 2005).

Table 4.6. Summary of protease inhibitors from seven oomycete pathogen species

Description	*Phytophthora infestans*[1]	*Pythium ultimum*[2]	*Saprolegnia parasitica*[3]	*Hyaloperonospora arabidopsidis*[4]	*Albugo laibachii*[5]	*Aphanomyces euteiches*[6]	*Plasmopara halstedii*[7]
No. of Kazal-like protease inhibitors	$31^{++}/33^{ins}$	$12^{+}/15^{ins}$	8	1	$5/8^{ins}$	1	1
Highest No. of Kazal-like domains in a protein	7	5	5	4	2	3	1
No. of cystatin-like protease inhibitors	$7^{++}/8^{ins}$	$3/6^{ins}$	6^{+}	4^{+}	2	1	1
Highest No. of cystatin-like domains in a protein	1	$1/1^{ns}$	3	1	2	2	1
No. of protease inhibitors, all*	$38^{++}/41^{ins}$	$15^{+}/21^{ns}$	14	5	7	2	2

[1, 2, 3, 4, 5] Oomycete species with available genome-sequencing data, that can be downloaded from www.broad.mit.edu (Baxter et al., 2010; Haas et al., 2009; Kemen et al., 2011; Levesque et al., 2010; Torto-Alalibo et al., 2005),
[6, 7] Oomycete species with available expressed sequence tag (EST) data (Bouzidi et al., 2007; Gaulin et al., 2008). These genomes may contain more protease inhibitors that were not detected in the transcriptome analysis.
[++] Highest number of secreted proteins,
[+] Second highest number of secreted proteins. Secretion signals predicted using SignalPv2.0 program (see chapter 2 section 2.2) (Nielsen et al., 1997).
[ins] Count that includes protease inhibitors that are predicted not to be secreted.
*Count only includes Kazal-like and Cystatin-like families of protease inhibitors.

4.3. Conclusions

The presence of protease inhibitors of both structural classes among various oomycete pathogens despite the diversity of hosts and lifestyles suggest that these effector families are common features in oomycetes. High numbers of protease inhibitors are induced *in planta* in *P. infestans* implicating them in virulence.

It was previously described that *P. infestans* Kazal-like protease inhibitors have atypical domains with two disulfide bridges (Tian et al., 2004; Tian and Kamoun, 2005). In this study, I show that these atypical domains are not present in other oomycetes outside the genus *Phytophthora*. Although EPI1 and EPI10 genes encoding protease inhibitors are divergent in sequence, they both are induced *in planta* and have atypical domains that were predicted to inhibit plant subtilisin A (see chapter 1, Fig. 1.2 and this chapter Table 4.1) (Tian et al., 2005; Tian et al., 2004). It is probable that other protease inhibitors of *P. infestans* that are also divergent in sequence but that contain atypical Kazal-like inhibitors are of importance to the pathogenicity of *Phytophthora*.

CHAPTER 5: Genome analyses of the *Phytophthora* clade1c species reveals families of fast evolving and *in planta*-induced genes

5.1. Introduction

Many plant pathogens, including those in the lineage of the Irish potato famine organism *Phytophthora infestans*, evolve by host jumps followed by specialization. However, how host jumps affect genome evolution remains largely unknown. Sylvain Raffaele (postdoc), Rhys Farrer (predoc) and I performed the genome analysis of six genomes of four sister species in order to determine patterns of sequence variation in the *P. infestans* lineage (Raffaele et al., 2010a). The genome analyses revealed uneven evolutionary rates across genomes with genes in repeat-rich regions showing higher rates of structural polymorphisms and positive selection. Importantly, in this study I highlight the finding that the gene sparse regions are enriched in *in planta*-induced genes, implicating host adaption in genome evolution. More specifically I report the gene expression patterns of a group of 65 genes encoding for rapidly evolving protein families that reside in the gene-sparse regions and show that within these families a high number of effector genes are induced *in planta*. Altogether, these results demonstrate that dynamic repeat-rich genome compartments underpin accelerated gene evolution following host jumps in this pathogen lineage.

5.2. Results and discussion

5.2.1. Sequence variation in effector genes of *Phytophthora* clade1c species

Phytophthora infestans is an economically important specialized pathogen that causes the destructive late blight disease on *Solanum* plants, including potato and tomato. In central Mexico, *P. infestans* naturally co-occurs with two closely

related species, *Phytophthora ipomoeae* and *Phytophthora mirabilis*, that specifically infect plants as diverse as morning glory (*Ipomoea longipedunculata*) and four-o'clock (*Mirabilis jalapa*), respectively. Elsewhere in North America, a fourth related species, *Phytophthora phaseoli*, is a pathogen of lima beans (*Phaseolus lunatus*). Altogether these four *Phytophthora* species form a very tight clade of pathogen species that share ~99.9% identity in their ribosomal DNA internal transcribed spacer regions (Kroon et al., 2004). Phylogenetic inferences clearly indicate that species in this *Phytophthora* clade 1c [commonly used nomenclature (Blair et al., 2008)] evolved through host jumps followed by adaptive specialization on plants belonging to four different botanical families (Blair et al., 2008; Grunwald and Flier, 2005). Adaptation to these host plants most likely involves mutations in the hundreds of disease effector genes that populate gene poor and repeat-rich regions of the 240-megabase pair (Mbp) genome of *P. infestans* (Raffaele et al., 2010a). However, comparative genome analyses of specialized sister species of plant pathogens have not been reported, and the full extent to which host adaptation affects genome evolution remains unknown.

To determine patterns of sequence variation in a phylogenetically defined species cluster of host-specific plant pathogens, Illumina reads for six genomes representing the four clade 1c species were generated (see chapter 2 section 2.4.1 and 2.4.2). The previously sequenced *P. infestans* strain T30-4 was included and used to optimize bioinformatic parameters (see chapter 2 chapter 2.4.7, Fig. 2.2 and Fig. 2.4) (Haas et al., 2009). By aligning Illumina reads of the five resequenced genomes to the reference genome strain T30-4 (see chapter 2 section 2.4.3) we could identified a total of 746,744 nonredundant coding sequence single-nucleotide polymorphisms (SNPs) (homozygous SNPs) (Fig. 5.1). We also investigated copy number variation (CNV) events (duplication or deletions) in coding genes of the five resequenced genomes relative to T30-4. To estimate gene copy number variation (CNV) we used average read depth per gene and GC content correction (see chapter 2 section 2.4.11 and Fig. 2.5, Fig. 2.6) (Yoon et al., 2009). In total, 3,975 CNV events were detected in coding genes of the five genomes relative to T30-4, among which there are 1,046 deletion events (see chapter 2 section 2.4.10, section Fig. 5.1).

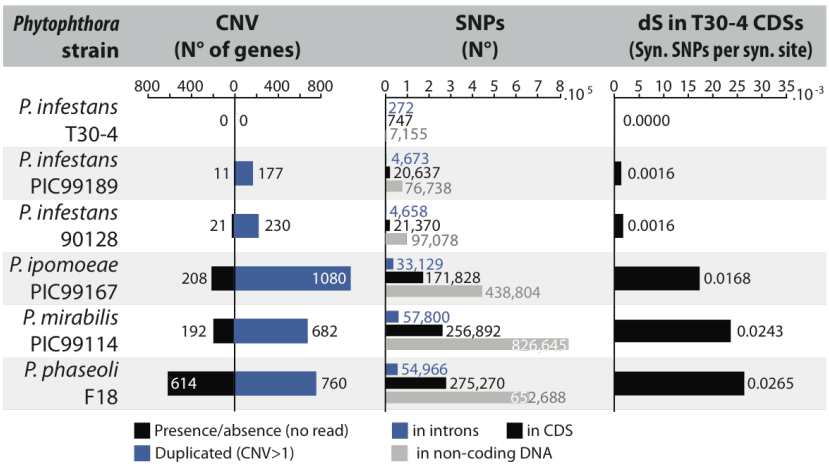


Fig. 5.1. Summary of genome sequences obtained for *Phytophthora* clade 1c species
Six strains representing four species were analyzed. *P. infestans* T30-4 previously sequenced was included for quality control (Haas et al., 2009). CDS, coding sequence; CNV, copy number variation; SNP, single-nucleotide polymorphism; syn., synonymous

To determine signatures of positive selection in the *Phytophthora* clade 1c species, rates of synonymous (dS) and nonsynonymous (dN) substitutions were calculated for every gene (see chapter 2 section 2.4.8) (Yang and Nielsen, 2000). Average dS divergence rates relative to *P. infestans* T30-4 were consistent with previously reported species phylogeny (Fig. 5.1) (Blair et al., 2008). We detected a total of 2,572 genes (14.2% of the whole genome) with dN/dS ratios >1 indicative of positive selection in the clade 1c species, with the highest number in *P. mirabilis* (1,004 genes) (Fig. 5.2A). A high proportion of genes annotated as effector genes show signatures of positive selection (300 out of 796) (Fig. 5.2B). This supports previous observations that effector genes are under strong positive selection in oomycetes (Allen et al., 2004; Liu et al., 2005; Win et al., 2007).

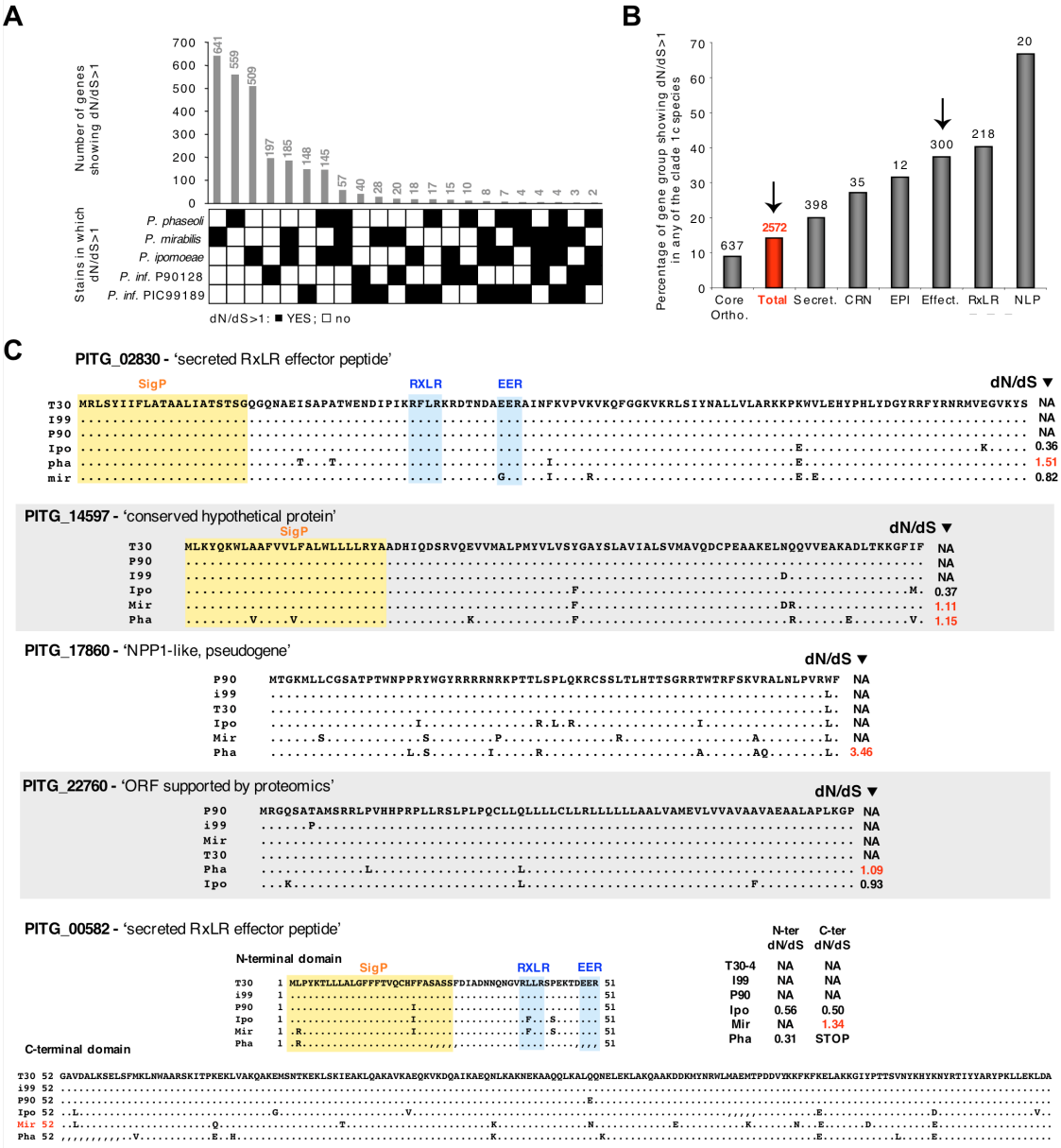


Fig. 5.2. Genes showing dN/dS>1 in the *Phytophthora* clade 1c species
(A) Number of genes with dN/dS>1 (Y-axis) and pairwise comparisons with the reference genome in which dN/dS>1 (X-axis black boxes - white if dN/dS<1 for comparison with this strain). Values are ordered by decreasing number of genes with dN/dS>.1 (B) Proportion of whole genome, core ortholog genes, secretome genes and various effector family genes showing dN/dS>1 as a percentage of the total number of genes in the examined group. The number of genes showing dN/dS>1 in the group is indicated as a label. (C) Examples of genes showing dN/dS>1 (including RXLR effectors). Alignments of homolog sequences in the resequenced strains are provided with polymorphic residues shown. Unresolved positions are indicated by a coma. PITG_00582 example illustrates a case where dN/dS>1 in the C-terminal domain of a RXLR effector using yn00 program of PAML (Yang, 2007) by implementing the Yang and Nielson method (Yang and Nielsen, 2000) (see chapter 2 section 2.4.8) . NA, not applicable; T30, *P. infestans* T30-4; i99, *P. infestans* PIC99189; P90, *P. infestans* P90128; ipo, *P. ipomoeae* PIC99167; mir, *P. mirabilis* PIC99114; pha, *P. phaseoli* F18; SigP, signal peptide.

5.2.2. Gene-sparse regions are enriched in genes with increased rates of CNV and positive selection

P. infestans genome has experienced a repeat-driven expansion relative to distantly related *Phytophthora* spp. with an unusual discontinuous distribution of gene density (Haas et al., 2009). Disease effector genes localize to expanded, repeat-rich and gene-sparse regions of the genome, in contrast to core ortholog genes, which occupy repeat-poor and gene-dense regions (Haas et al., 2009). We exploited our sequence data to determine the extent to which genomic regions with distinct architecture evolved at different rates. Statistical tests and random sampling species were analyzed. *P. infestans* T30-4 previously sequenced was used to determine the significance of differences in CNV, presence/absence polymorphisms, SNP frequency, and dN/dS values in genes located in gene-dense versus gene-sparse regions (see chapter 2 Table 2.3). Although averages of gene copy numbers were similar in both regions, significantly higher frequency of CNV and gain/loss were observed in genes located in the repeat-rich regions (Fig. 5.3A). Notably, presence/absence polymorphisms were 13 times as abundant in the gene-sparse compared to the gene-dense regions. In addition, even though SNP frequency was similar across the genomes, average dN/dS was significantly higher in gene-sparse regions, indicating more genes with signatures of positive selection (Fig. 5.3A). Indeed, 23% of the genes in the gene-sparse regions showed dN/dS > 1 in at least one of the resequenced genomes compared to only 11.5% of genes in the gene-dense regions. In total, 44.6% of the genes in the gene-sparse regions showed signatures of rapid evolution (deletion, duplication, or dN/dS > 1) compared to only 14.7% of the remaining genes. The uneven distribution in gene density in the *P. infestans* genome can be visualized with plots of two-dimensional bins of 5' and 3' flanking intergenic region (FIR) lengths (Haas et al., 2009). These plots were adapted to illustrate the relationships between gene density and polymorphism. This plots confirmed that in the gene-sparse regions there is increased rates of polymorphisms including CNV (duplications and deletion events and positive selection (Fig. 5.3B). These findings indicate that different regions of the examined genomes evolved at markedly different rates, with the gene-sparse, repeat-rich regions experiencing accelerated rates of evolution.

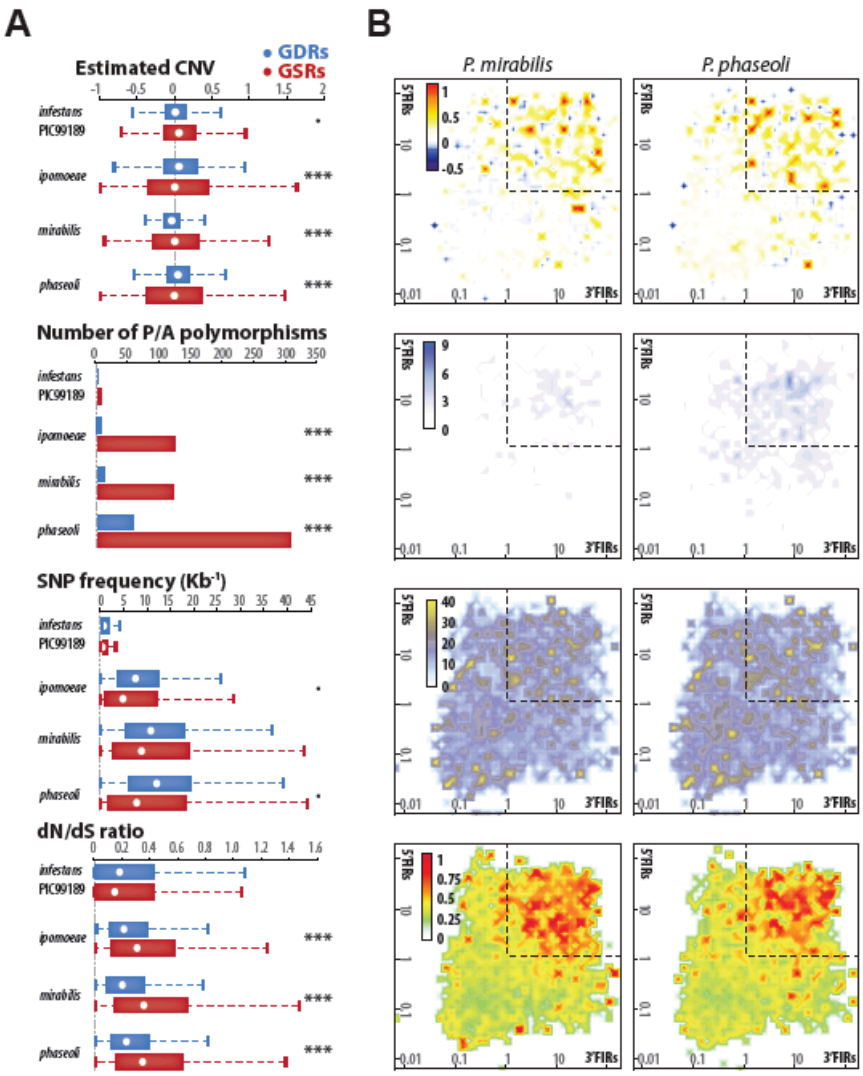


Fig. 5.3. The two-speed genome of *P. infestans*
(A) Distribution of copy number variation (CNV), presence/absence (P/A) and single-nucleotide polymorphisms (SNP), and dN/dS in genes from gene-dense regions (GDRs) and gene-sparse regions (GSRs). Statistical significance was assessed by unpaired t test assuming unequal variance (CNV, dN/dS); assuming equal variance (SNP frequency); or by Fisher's exact test (P/A) (•P<0.1; ***P<10−4) (see chapter 2 section 2.4.12). Whiskers show first value outside 1.5 times the interquartile range. (B) Distribution of polymorphism in *P. mirabilis* and *P. phaseoli* according to local gene density (measured as length of 5' and 3' flanking intergenic regions, FIRs). The number of genes (P/A polymorphisms) or average values (CNV, SNP, dN/dS) associated with genes in each bin are shown as a color-coded heat map (see chapter 2 section 2.4.13).

5.2.3. Gene-sparse regions are enriched in genes that are induced *in planta*

To gain insights into the functional basis of the uneven evolutionary rates detected in the gene-sparse versus gene-dense regions of the clade 1c species, I

used wide-genome microarray expression from a time course infection on potato and tomato of *P. infestans* T30-4 during and plotted on the FIR length maps (Fig. 5.4, see chapter 2 section 2.5.1) (Haas et al., 2009). Gene-dense regions were enriched in genes that are induced in sporangia, the asexual spores that are produced by all *Phytophthora* species. In marked contrast, distribution patterns of genes that are induced during pre-infection and infection stages on potato and tomato indicate enrichment of these genes to gene-sparse loci (Fig. 5.4A) (and 2.4.13). I performed χ^2 tests to show that the relationships between gene density (FIR length) and patterns of gene expression are significant (Fig. 5.4B, see chapter 2 section 2.4.12, Table 2.4). These suggest that the gene-sparse, repeat rich regions are highly enriched in *in planta*-induced genes, therefore implicating host adaptation in genome evolution.

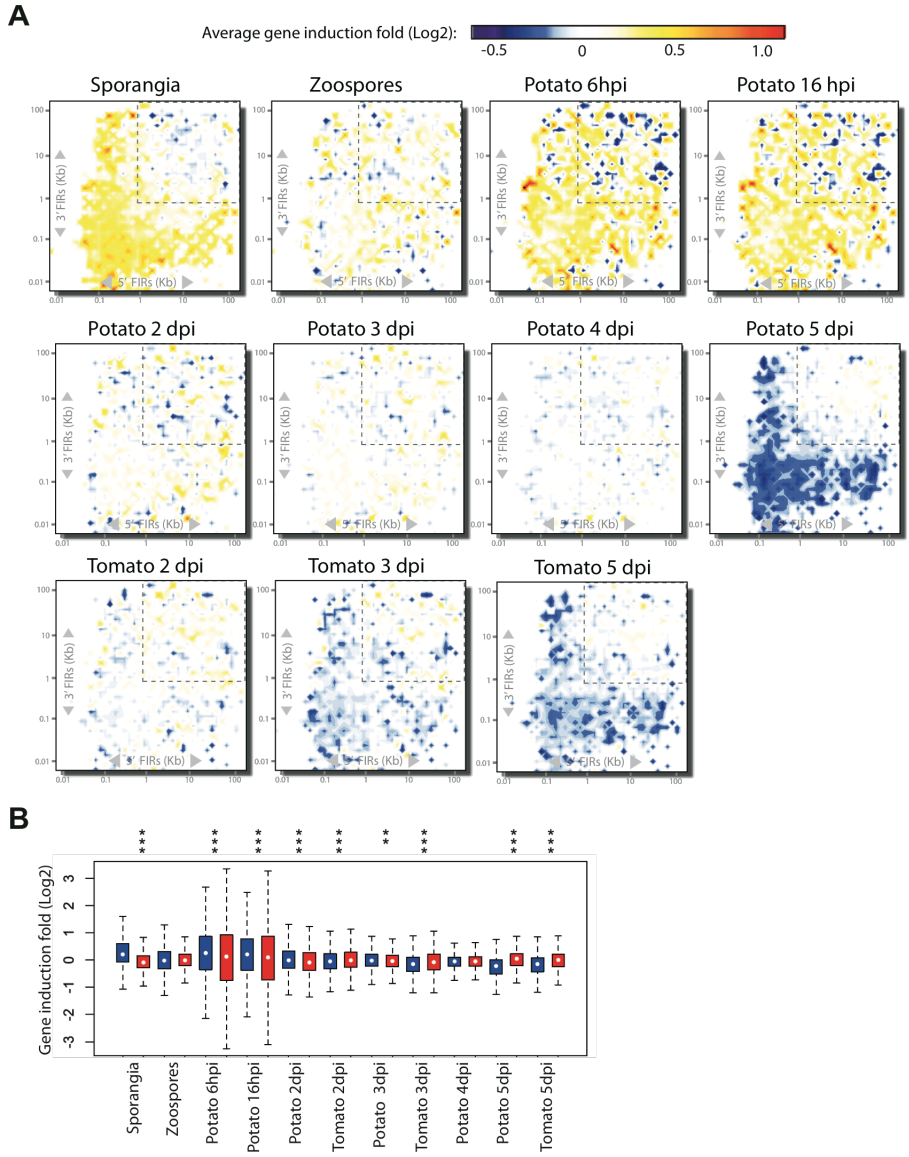


Fig. 5.4. The gene-sparse regions (GSRs) of *P. infestans* genome are highly enriched of *in planta*-induced genes
(A) Distribution of gene induction according to local gene density (measured as length of flanking intergenic regions, FIRs). Genes were sorted into two-dimensional bins according to the length of their 5' (y-axis) and 3' (x-axis) FIR lengths. Average induction values associated to genes in each bin are shown as a color-coded heat map for sporangia, zoospores, infection of tomato 2, 3 and 5 days post-inoculation (dpi) and infection of potato 6, 16 hpi and 2-5 dpi. Values are relative to expression in *in vitro* grown mycelium (see chapter 2 section 2.5.1 and section 2.4.13). (B) Distribution of fold of gene induction (as log2 compared to expression value in mycelia) for genes located in gene-dense regions (GDRs in blue) and GSRs (red). Whiskers of the box plots show first value outside of 1.5 times the interquartile range. Statistical significance was assessed by Mann-Whitney U-test. Probabilities as shown as: *, p<0.01; **, p<0.001; and ***, p<10E-04 (see chapter 2 section 2.4.12).

5.2.4. Protein families containing fast evolving genes and present in gene-sparse regions

To assign biological functions to genes with accelerated rates of evolution that populate the gene-sparse, repeat-rich regions, a Markov clustering was performed on the predicted proteome of *P. infestans* and implemented gene ontology mapping. Protein families (tribes) significantly enriched or deficient in genes that locate to gene-sparse regions or are rapidly evolving were identified with Fisher's exact test (see chapter 2 section 2.4.14). In total, 811 tribes with five or more proteins were generated, containing 7,993 proteins out of 18,155 of the predicted proteins (equivalent to 44% of proteome). Of these, 163 tribes were statistically enriched (p-value<0.1) in genes from gene sparse regions (GSR), 123 tribes were enriched (p-value<0.1) in Fast-Evolving (FE) genes and 65 tribes were enriched (p-value<0.1) in both (see chapter 2 section 2.4.14). I found that 67% of the Tribes with genes from gene-sparse regions (GSR) and fast evolving genes (44 out of the 65) show at least 1 member that is induced *in planta* (see appendix 3.1). As expected, several of these tribes (19 out of 65) consist of effector families (Kamoun, 2006; Oh et al., 2009; Tian et al., 2007). Tribe171 consisting of protease inhibitors shown to suppress host defences by targeting host proteases, exhibited the highest frequency of *in planta* induced-genes (30 out of 41) and particularly rich in genes located in gene-sparse regions and exhibiting presence/absence polymorphisms, duplications and positive selection (Fig. 5.5) (Song et al., 2009; Tian et al., 2005; Tian et al., 2004; Tian and Kamoun, 2005; Tian et al., 2007). Tribes consisting of RXLR effectors included homologs of *Avrblb2* (Tribe123) from *P. infestans*, which had previously been shown to be highly induced *in planta* and to be under positive selection (Oh et al., 2009). In addition, I detected high number of duplication events in genes that belong involved in cell wall degradation, including pectate lyase (Tribe016), pectin lyase (Tribe061), glycosyl hydrolases, and an unique RXLR family (Tribe225) with hydrolase activity annotation (GO:0016787) effectors that suggests substantial changes in the cell wall degrading enzyme repertoire (Table 5.1). In addition to known effector families, tribes annotated as histone (Tribe032 and Tribe486) and ribosomal RNA (rRNA) methyltransferases (Tribe066), which are involved in epigenetic maintenance, were particularly rich in genes located in

gene-sparse regions and exhibiting presence/absence polymorphisms (Table 5.1, Fig. 5.6) (Peng and Karpen, 2009).

Table 5.1. Gene expression patterns of *P. infestans* tribes (with annotations) enriched in genes residing in gene-sparse regions (GSR) and are rapidly evolving

Description (Tribe ID)	Gene Ontology (GO) ID	No. of genes in Tribe	Secreted[a]	No. of genes *in planta*-induced genes[b]	Genes in GSR repeat-rich regions		Rapidly evolving genes			
					No.	P-value[c]	Presence/ Absence[d]	Duplicated[e]	dN/dS>1[f]	P-value[g]
Effectors										
Protease Inhibitor (171)	GO0008233	41	37	30	18	3.88E-05	7	5	18	6.94E-05
RXLR effector Avrblb2 (123)	na	14	14	11	10	3.87E-05	9	4	2	5.39E-05
NPP1-like family (052)	na	21	15	8	10	1.09E-02	4	3	5	1.80E-02
RXLR effector, hydrolase (225)	GO0016787	10	8	6	9	1.73E-06	2	2	6	6.44E-04
RXLR effector (074)	na	18	17	5	9	1.05E-02	1	8	3	4.35E-04
RxLR effector Avr2 (429)	na	7	7	4	7	6.74E-06	7	0	1	8.91E-05
RXLR effector (0174)	na	12	7	4	8	7.71E-04	0	5	6	3.46E-02
RXLR effector (154)	na	12	9	4	6	4.85E-02	2	4	4	3.46E-02
RXLR effector (305)	na	8	8	3	6	1.52E-03	2	5	3	5.61E-04
RXLR effector (555)	na	6	6	3	4	3.30E-02	0	1	3	8.47E-02
RXLR Effector (610)	na	5	4	3	4	1.01E-02	0	2	2	3.07E-02
RXLR effector (805)	na	5	5	3	4	1.01E-02	1	4	1	3.07E-02
RXLR effector (536)	na	6	0	3	6	4.18E-05	1	1	4	8.22E-03
RXLR effector (349)	na	8	4	3	6	1.52E-03	1	3	4	6.31E-02
RXLR effector (576)	na	6	1	3	5	1.84E-03	2	2	4	8.22E-03
Crinkler effector (034)	na	21	6	2	14	3.37E-06	0	15	1	1.48E-05
RXLR effector (551)	na	6	6	1	4	3.30E-02	5	0	1	3.68E-04
Crinkler effector (022)	na	31	1	0	21	3.76E-09	0	27	4	2.36E-13
Elicitin (024)	GO0009405	28	3	0	15	1.61E-04	11	5	0	7.47E-04
DNA and RNA maintenance processes										
DOT1-like Histone-Lysine N-methyltransferase (032)	GO0018024	25	0	3	14	1.27E-04	7	6	1	9.81E-03
Centromere protein CENP-B, helix turn helix domain (218)	GO0045449	10	1	2	6	1.25E-02	0	3	4	4.67E-02
DNA-binding domain (200)	GO0043565	11	0	1	10	2.97E-07	0	4	5	1.68E-02
SET domain histone methyltransferase (486)	GO0008168	6	0	0	6	4.18E-05	6	0	0	3.68E-04
SpoU rRNA methyltransferase (066)	GO0008173	19	13	0	12	5.70E-05	1	0	0	5.99E-02
Cell wall degrading enzymes and carbohydrate binding proteins										
Pectate lyase (016)	GO0030570	37	14	9	20	7.71E-06	0	15	1	4.22E-02
Pectin lyase (061)	GO0047490	20	14	7	11	1.09E-03	3	7	5	2.18E-03
Chitin binding protein (095)	GO0008061	16	4	3	7	7.57E-02	0	2	7	7.44E-02
Hydrolase of O-glycosyl compounds (396)	GO0004553	6	1	2	5	1.84E-03	1	2	5	8.22E-03
Other enzymes										
Phosphoenolpyruvate carboxykinase (450)	GO0004611	7	0	2	4	7.64E-02	0	6	0	2.16E-03
Serine protease (416)	GO0006508	7	1	1	4	7.64E-02	0	6	0	2.16E-03
Cysteine protease (085)	GO0008234	16	0	0	8	1.74E-02	12	0	1	4.25E-06

[a] Secretion signals were predicted with SignalPv2.0 program (Nielsen et al., 1997) with a HHM signal peptide probability of 0.9 or higher (Torto et al., 2003). In addition to signalP predictions, sequences that contained putative transmembrane domains (TM) predicted with TMHMM program (Krogh et al., 2001) were filtered out. [b] Number of genes in tribe induced during the biotrophic phase of infection on potato (at any of the time points: 6 hpi, 16 hpi, 2 dpi, 3 dpi) and/or on tomato (at any of the time points: 2, 3 dpi) using mycelia as baseline (see chapter 2 section 2.5.1 and appendix 3.1). Hour post inoculation (hpi); Days post inoculation (dpi). [c and g] P-value of chi-square test for the enrichment of genes with the indicated attribute (see chapter 2 section 2.4.14). [d, e and f] Number of genes within a Tribe with the indicated attribute.

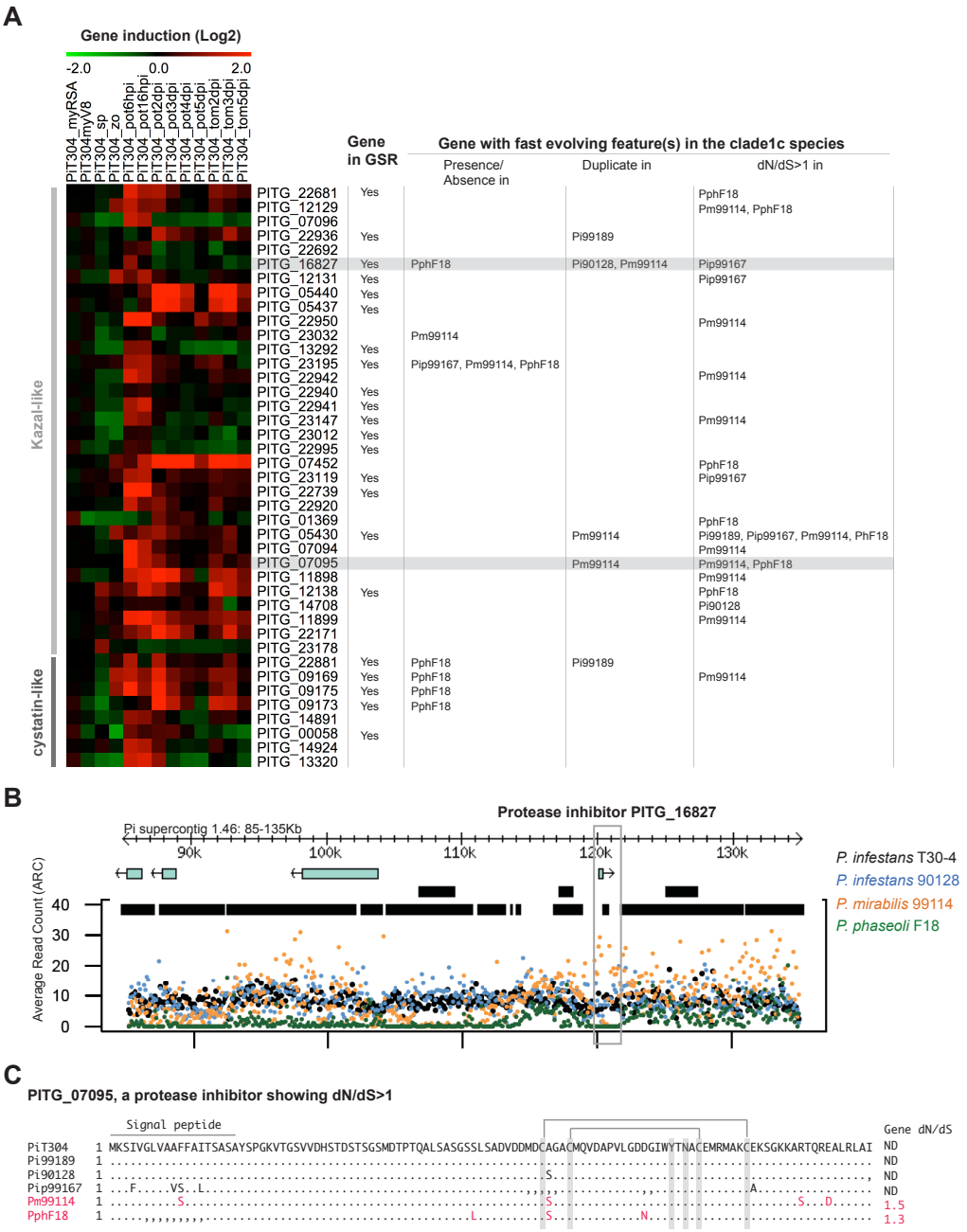


Fig. 5.5. Expression patterns and polymorphisms in *Phytophthora* protease inhibitors effector families

(A) Left panel show genes induction patterns during infection of Kazal-like and cystatin-like protease inhibitors from Tribe 171 (see appendix 3.1). Gene expression values are estimated relative to mycelia (see chapter 2 section 2.5.1). MyRSA, mycelia in Rye Sucrose Agar (RSA); MyV8; mycelia in V8 agar; sp, sporangia; zo, zoospores; pot, potato; tom, tomato; hpi, hours post inoculation; dpi, days post inoculation. Right panel show genes that locate to Gene sparse regions (GSR) and have fast evolving feature(s) in the *Phytophthora* clade1c species. Pi99189, *P. infestans* 99189; Pi90128, *P. infestans* 90128; Pip99167, *P. ipomoeae* PIC99167; Pm99114, *P. mirabilis* PIC99114; PphF18, *P. phaseoli* F18. Two examples of Kazal-like protease inhibitors exhibiting structural (B) and sequence polymorphisms (C) from this tribe are marked highlighted in grey. (B)

Duplication events spanning the CDSs of the Kazal-like protease inhibitor PITG_16827 in *P. infestans* 90128 and *P. mirabilis* PIC99114 and deletion events in *P. phaseoli* F18 identified using Average Read Count (ARC) along 100 bp windows. The upper ribbon shows the corresponding window illustrated using *P. infestans* genome browser. (C) Example of a Kazal-like protease inhibitor PITG_07095 showing dN/dS>1 in *P. mirabilis* and *P. phaseoli* shown in red. dN/dS ratios were calculated using yn00 program of PAML (Yang, 2007) by implementing the Yang and Nielson method (Yang and Nielsen, 2000) (see chapter 2 section 2.4.8). Alignment of homologous sequences in the re-sequenced strains is provided with only the polymorphic residues shown. Unresolved portions are indicated by comas. In cases where dN/dS values could not be calculated, the dN/dS for this gene is indicated with ND, not determined. Amino acid residues that define the Kazal family protease inhibitor domain are highlighted in grey. The putative disulfide linkages formed by cysteine residues within the predicted Kazal domains are drawn. Sequence encoding for Signal peptide in the N-terminal region is shown with a bar. Secretion signals were predicted using SignalPv2.0 program (Nielsen et al., 1997).

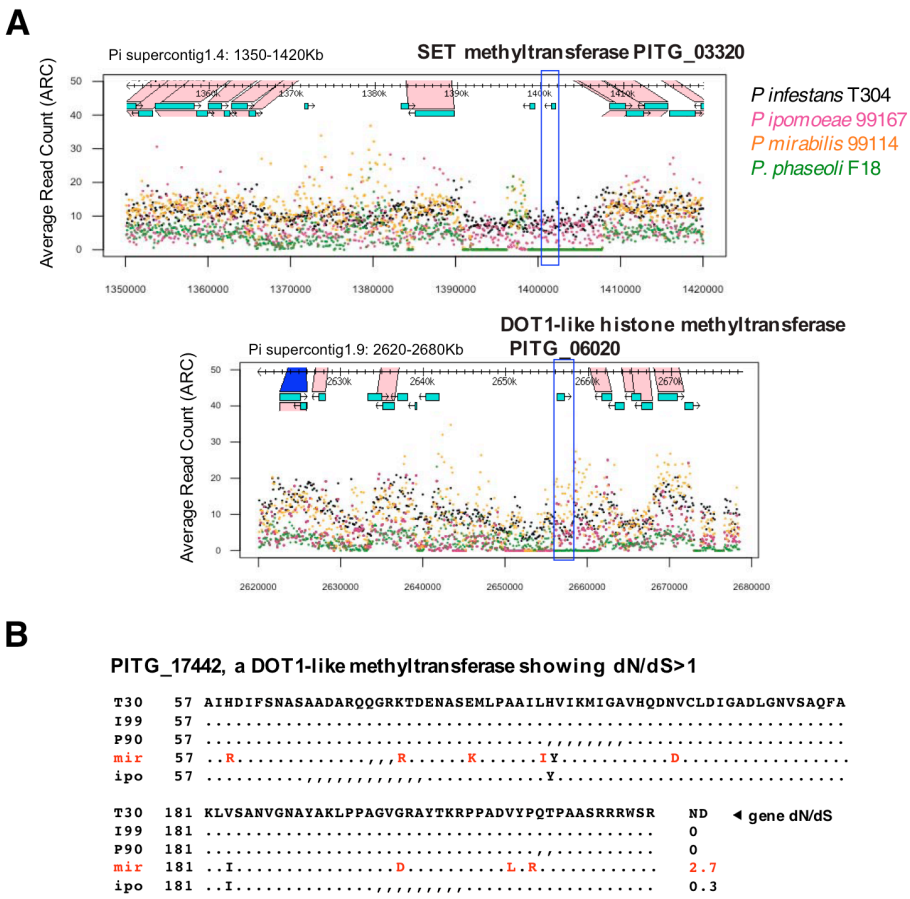


Fig. 5.6. Illustration of polymorphism in *Phytophthora* SET-domain and DOT1-like histone methyltransferases

(A) Deletions events spanning the CDS of histone methyltransferases in some *P. infestans* related species identified using Average Read Count (ARC) along 100 bp windows. The upper ribbon shows the corresponding window illustrated using *P. infestans* genome browser SybilLite (see chapter 2 section 2.3). (B) Example of a histone methyltransferase showing dN/dS>1 in *P. mirabilis*. Portions of the alignment of homologous sequences in the re-sequenced strains are provided with only the

polymorphic residues shown. Unresolved positions are indicated by comas. dN/dS ratios were calculated using yn00 program of PAML (Yang, 2007) by implementing the Yang and Nielson method (Yang and Nielsen, 2000) (see chapter 2 section 2.4.8). In cases where dN/dS values could not be calculated, the dN/dS for this gene is indicated with ND, not determined.

5.3. Conclusions

This study demonstrates that highly dynamic genome compartments enriched in noncoding sequences underpin accelerated gene evolution following host jumps. Gene-sparse regions that drive the extremely uneven architecture of the *P. infestans* genome are highly enriched in *in planta*-induced genes, particularly effectors, therefore implicating host adaptation as a driving force of genome evolution in this lineage. *In planta*-induced and rapidly evolving effector families that resides largely in gene-spares regions included RXLRs, protease inhibitors and a variety of cell wall degrading enzymes. In addition to known effector families, several rapidly evolving genes annotated as histone and RNA methyltransferases involved in epigenetic processes were also significantly enriched in the gene-sparse regions. Histone methylation indirectly modulates gene expression in various eukaryotes and could underlie concerted and heritable gene induction patterns through long-range remodeling of chromatin structure (Elizondo et al., 2009; Kouzarides, 2002; Zhang and Reinberg, 2001). Histone acetylation and methylation are thought to be key regulators of gene expression in *P. infestans* and could modulate expression patterns of genes located in the gene-sparse regions (van West et al., 2008). In addition, histone hypomethylation reduces DNA stability and may have contributed to genome plasticity in the *P. infestans* lineage by regulating transposons activity as well as genomic and expression variability (Elango et al., 2008; Peng and Karpen, 2009; Peters et al., 2001; Zeh et al., 2009). Finally, understanding *P. infestans* genome evolution should prove useful in designing rational strategies for sustainable late blight disease management based on targeting the most evolutionarily stable genes in this lineage.

CHAPTER 6: Genome analyses of a clonal lineage 13_A2 of *Phytophthora infestans* uncover expression and genetic polymorphisms in effector genes

6.1. Introduction

The Irish potato famine pathogen *Phytophthora infestans* causes the late blight disease, an enduring problem for world agriculture and a threat to global food security. *P. infestans* is an oomycete (eukaryotic) microbe capable of both sexual and asexual reproduction. It is remarkable for its ability to rapidly adapt to genetically resistant potatoes and agrochemicals. In agricultural systems, *P. infestans* has experienced major population shifts driven by migration and successive emergence of asexual clonal lineages.

The 2007 late blight season in the United Kingdom (UK) was the worst reported in the last 50 years, mainly due to the emergence and rapid spread of an aggressive clone of *P. infestans* termed genotype 13_A2 (Chapman et al., 2010; Fry et al., 2009; Vleeshouwers et al., 2011). 13_A2 isolates are able to overcome previously effective forms of plant host resistance - adaptive phenotypic traits that probably drove the population displacement (Chapman et al., 2010). In order to investigate the molecular basis of the enhanced aggressiveness and virulence phenotypes observed in infected plants by *P. infestans* 13_A2, I performed genome analyses (whole-genome sequencing and whole-genome expression analyses) of a 13_A2 representative isolate named 06_3928A. This work revealed significant genetic and expression polymorphisms, particularly within disease effector genes. Also, this work uncovered diverse evolutionary events associated with effector genes that could have contributed to the enhanced virulence. Importantly, I highlight that 13_A2 isolate 06_3928A carry intact *Avrblb1, Avrblb2* and *Avrvnt1* effector genes that are induced *in planta*. Consistent with these findings, 06_3928A isolate cannot infect potato lines that carry the corresponding *R* immune receptor genes *Rpi-blb1*, *Rpi-blb2*, *Rpi-vnt1.1*. These findings point to a genetic strategy for mitigating the impact of 13_A2

epidemics and illustrate how pathogen genome analysis can benefit the management of a devastating plant disease epidemic.

6.2. Results and discussion

6.2.1. Genome sequencing analysis of *P. infestans* 06_3928A isolate

P. infestans delivers inside plant cells disease effector proteins to promote host colonization, for instance by suppressing plant immunity (Oh et al., 2009). The major class of host translocated effectors are the RXLR proteins, which are encoded by ~550 genes in the *P. infestans* T30-4 genome (Haas et al., 2009). A number of RXLR effectors trigger hypersensitive cell death and late blight resistance in plants expressing the matching R immune receptor (Vleeshouwers et al., 2011). In such cases, the RXLR effectors are said to have an “avirulence” activity acting as triggers of plant immunity.

To determine the effector gene repertoire and unravel genetic features of the 13_A2 Multilocus Genotype (MLG), I generated ~58-fold genome coverage Illumina paired-end reads of isolate 06_3928A Table 6.1 and see chapter 2 section 2.4.2). I processed the sequences first by aligning the reads to the previously sequenced genome of *P. infestans* strain T30-4 (see chapter 2 section 2.4.3) (Haas et al., 2009), and then by performing de novo assembly of the unaligned reads (see chapter 2 section 2.4.4). In total, 95.6% of the 06_3928A reads aligned to the T30-4 sequence (Table 6.1).

Table 6.1. Genome alignment statistics of *P. infestans* 06_3928A isolate

Run ID	Lane ID	No. of reads* (76 bp X 2)	No. of mapped reads	% of mapped reads	No. of reads mapped in pairs	% of reads mapped in pairs	No. of unmapped reads	% of unmapped reads
ID99	Lane 5	25,308,382	24,630,707	97.3	24,156,004	95.4	677,675	2.7
ID101	Lane 8	27,448,558	26,432,076	96.3	25,770,274	93.9	1,016,482	3.7
ID103	Lane 5	35,037,640	33,174,380	94.7	31,971,342	91.2	1,863,259	5.3
ID103	Lane 6	34,627,312	32,693,366	94.4	31,527,118	91.0	1,933,946	5.6
ID103	Lane 7	35,689,613	33,930,316	95.1	32,770,200	91.8	1,759,296	4.9
ID103	Lane 8	33,410,173	31,938,448	95.6	30,956,230	92.7	1,471,725	4.4
	Total	191,521,678	182,799,293	95.6	177,151,168	92.7	8,722,383	4.4
		Estimated Genome Depth	58x					

* Count after filtering for reads containing Ns and/or abnormal read length.

6.2.1.1. RXLR effector genes show higher rates of dN/dS

In this study, I focused in the identification of coding genes from the sequenced genome of *P. infestans* 06_3928A. To exclude missing genes from further analyses, I looked at genes with an average breadth of coverage greater than 0 (see chapter 2 section 2.4.10). I identified 18,106 coding sequences with an average breadth of coverage of 99.2% (Table 6.2). I optimized bioinformatic parameters for calling single nucleotide polymorphisms (SNPs) to reach 99.9% accuracy and 85.8% sensitivity (see chapter 2 section 2.4.7 and Fig. 2.3). Using these parameters, I identified 22,523 SNPs in 5,879 coding sequences of 06_3928A (Table 6.2). This value is in the same range as the 20,637 and 21,370 SNPs reported for *P. infestans* isolates PIC99189 and 90128, respectively (Raffaele et al., 2010a) (Table 6.2). Of the total SNPs discovered, 11,795 were unique to 06_3928A among the four examined strains indicating a considerable degree of variation in the 13_A2 MLG (Table 6.2, Fig. 6.1A).

Table 6.2. Genome features of three *P. infestans* isolates

Genome features	*P. infestans* 06_3928A	*P. infestans* isolates (clade1c)	
		PIC99189	90128
Predicted genome size (Mb)	240	-	-
Average genome coverage	58x	-	-
Average breadth of coverage in coding sequences (%)	99.2	-	-
Average depth of coverage in coding sequences	70.2x	-	-
No. of SNPs in coding sequences	22,523	20,637	21,370
No. of SNPs causing loss of stop codons	90	72	73
No. of unique SNPs in coding genes *	11,795	9,935	11,645
No. of genes with at least one SNP *	5,879	6,784	7,361
No. of SNPs in introns	6,043	4,673	4,658
No. of SNPs in non-coding DNA	155,996	76,738	97,078
dS in T30-4 CDSs (syn. SNPs per syn. site)	0.0018	0.0016	0.0016
No. of genes with presence/absence (no reads)	47	11	21
Uncovered regions in the genome (no reads) (Mb)	6.5	7.8	13.4
No. of genes showing CNV>1	320	177	230
No. of genes showing dN/dS>1 †	288	232	270

* Count of SNPs causing loss of stop codons were omitted
† dN/dS rates were calculated using Yang method (see chapter 2 section 2.4.8) (Yang and Nielsen, 2000).

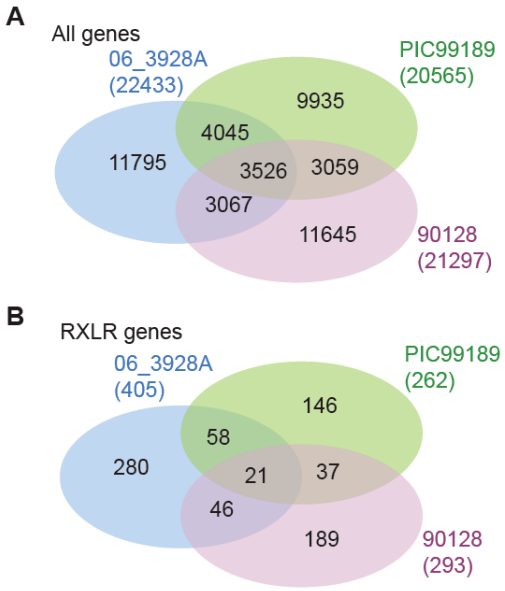


Fig. 6.1. Venn diagrams with number of SNPs in coding genes of three *P. infestans* isolates
SNPs in the *P. infestans* isolate 06_3928A were called in positions with 90% of consensus bases and a minimum read depth of 10. SNPs in the *P. infestans* PIC99189 and 90128 isolates were called as reported in Raffaele et al., (Raffaele et al., 2010a). SNPs causing loss of stop codons were excluded for each category. (A) Total number of SNPs in all coding genes. (B) Total number of SNPs in secreted RXLR effector genes.

To detect signatures of positive selection in the 13_A2 lineage relative to T30-4, I calculated rates of synonymous (dS) and nonsynonymous (dN) substitutions for every gene (see chapter 2 chapter 2.4.8). Of the 22,523 coding sequence SNPs, 11,421 are nonsynonymous (51%) corresponding to an average dN/dS rate of 0.34 (Table 6.3). Of the 405 SNPs detected in RXLR genes, 278 are nonsynonymous (69%) corresponding to an average dN/dS rate of 0.53 (Table 6.3 and see appendix 4.1).

Table 6.3. Summary of nonsynonymous and synonymous SNPs in coding genes of *P. infestans* 06_3928A

SNP count *	*P. infestans* 06_3928A		
	All genes	Core orthologs	RXLRs
Total No. of SNPs in coding genes	22,433	11,612	405
Total No. of nonsynonymous SNPs in coding genes	11,421	5,439	278
Total No. of synonymous SNPs in coding genes	11,012	6,173	127
No. of genes with at least one SNP	5,879	2,754	118
Average dN/dS †	0.34	0.30	0.53

* Count of SNPs causing loss of stop codons were omitted
† dN/dS rates were calculated using Yang method (see chapter 2 section 2.4.8) (Yang and Nielsen, 2000).

6.2.1.2. RXLR effector genes of *P. infestans* 06_3928A isolate show higher dN/dS rates compared to T30-4

Secreted protein genes, particularly RXLR effector genes, show higher rates of dN/dS compared to other gene categories indicative of positive selection (Fig. 6.2). RXLR effectors are modular proteins with their N-termini involved in secretion and host-translocation while the C-termini encode the effector biochemical activity (Morgan and Kamoun, 2007). I noted that the C-terminal domains of RXLR effector genes are highly enriched in nonsynonymous substitutions as previously described in other oomycete species (Win et al., 2007) (Fig. 6.3 and Fig. 6.4). These observations indicate that the effector domains of a number of RXLR genes of 13_A2 MLG may have been targeted by positive selection possibly contributing to enhanced aggressiveness and virulence. This also extends the work of Win et al. (2007) by showing that elevated rates of nonsynonymous substitutions at the C-termini of RXLR genes is detectable at the intra species level.

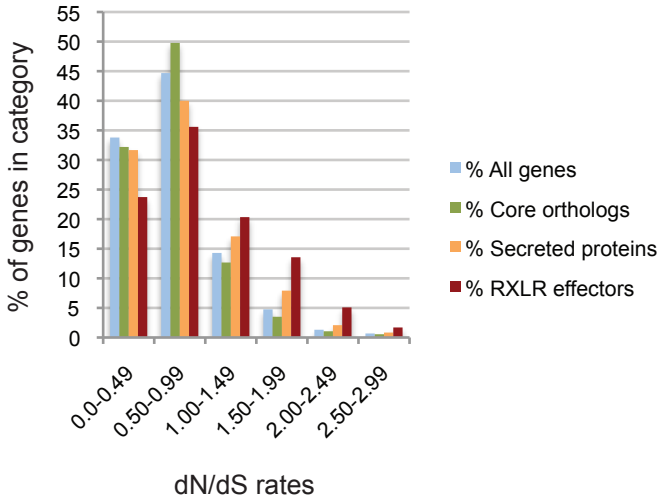


Fig. 6.2. Distribution of dN/dS in coding genes of *P. infestans* 06_3928A
dN/dS rates were calculated using Yang method (see chapter 2 section 2.4.8) (Yang and Nielsen, 2000). Genes where the Yang method was not applicable were omitted. A total of 3,975 (all), 1,997 (core orthologs), 240 (secreted) and 59 (RXLR) genes were analyzed.

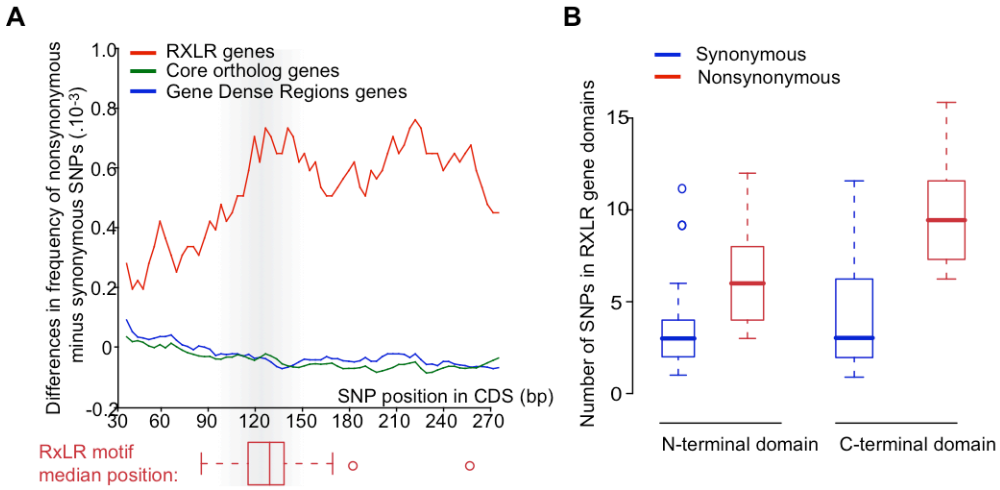


Fig. 6.3. Frequency of synonymous and nonsynonymous SNPs in RXLR genes of *P. infestans* 06_3928A
SNP count was considered for genes having at least one SNP. (A) Differences in the frequency of nonsynonymous minus synonymous SNPs in RXLRs (118 genes), Core orthologs (3,077 genes) and all gene dense region genes (2,442 genes) (see chapter 2 section 2.4.9) according to the position of SNP in the CDSs without signal peptides sequences. (B) Number of SNPs detected in the N-terminal and C-terminal domains of RXLR genes (see chapter 2 section 2.4.9).

A

PITG_14203- Secreted RxLR effector peptide (RXLRfam33) **dN/dS = 1.13**

N-terminal domain

```
T30-4      1  MRLSCVLLIALVALLSSFDQVSTKTSDDALTMKALRQGNAVATVTNNKRLLQTELPGSSAILEDNEER
06_3928A   1  ............................................................N.........
```

C-terminal domain

```
T30-4     69  ASWGLKKLVNLFMFKGASSNFSTAKLTKMLSDPAFKSEMFKTWNAKYSTEKVIAILDLSKKKNIPYGPMLQDFRNTYGRRPIKPQ
06_3928A  69  .........S........D.A....E.........................T...I..............L...........
```

B

PITG_00619- Secreted RxLR effector peptide (RXLRfam14) **dN/dS = 1.03**

N-terminal domain

```
T30-4      1  MKSFFVFVMFIVSFLAIGEAMETAGCRNNKLAIDPNQRFLRTTTTTEDDTATEEER
06_3928A   1  ............................................................
```

C-terminal domain

```
T30-4     57  GIQTRVFDKLGRATGNLRMPQKLKFWIWKKGKWDYNRLKEYLFKGVPKSVYEKDPRFAKILHKYGYYWRKSAMNGHQVRS
06_3928A  57  ..............................................A.............V.....D.............
```

C

PITG_11952- Secreted RxLR effector peptide (RXLRfam23) **dN/dS = 1.04**

N-terminal domain

```
T30-4      1  MSRYSMMLLLRLAVLITFGLSCAVALATDSNNVVKSNPTIKDTQALRFLRRYAFDEEANRDNEDQER
06_3928A   1  ...............................................................E..
```

C-terminal domain

```
T30-4     68  GIGVSKQLDDVLLKADEVIGILKNVVRKVGGALIKNSRETDEYLRLTRAVSGKYPTAKELSLSTLRQL
06_3928A  68  ....................................................................

T30-4    136  KKIEAVRKKDIEKGIDVSKAATDGIHRDIKLFDGNKRVPDKYMGAHVGRDQQRFTEADTRSLVAGVVT
06_3928A 136  ......................M........I.........................A.....

T30-4    204  RTNKKGEQEILLVSSSKPTKYEFMISKGVGKKTRVLKWQRCVRSLKKEGYVDRSWILGAVTIYTNSR
06_3928A 204  ........K...................R...............................H..
```

Fig. 6.4. Examples of RXLR effectors showing dN/dS ratios >1 in *P. infestans* 06_3928A
(A) PITG_14203 secreted RXLR effector with dN/dS ratio of 1.13. (B) PITG_00619 secreted RXLR effector with dN/dS ratio of 1.03. (C) PITG_11952 secreted RXLR effector with dN/dS ratio of 1.04. dN/dS rates were calculated using Yang method (see chapter 2 section 2.4.8) (Yang and Nielsen, 2000). N-terminal domain is shown in grey, signal peptide sequence in yellow, RXLR-EER motif in green and the C-terminal effector domain is in pink. Conserved amino acids are indicated with dots in the gene from 06_3928A isolate.

6.2.1.3. *P. infestans* 06_3928A isolate shows copy number variation in RXLR effector genes

To estimate copy number variation (CNV) in the resequenced genome of 06_3928A relative to T30-4, I used average read depth per gene and GC content correction (see chapter 2 chapter 2.4.11). I detected 367 CNV events among

06_3928A genes, of which there are 320 duplications and 47 deletions (see appendix 4.2 and appendix 4.3). RXLR effector genes show higher rates of CNV compared to other gene categories (Fig. 6.5 and see appendix 4.1). Two RXLR effector genes showed high levels of CNV with ~4-5X additional copies present in 06_3928A compared to other *P. infestans* reference strain T30-4 (Fig. 6.6). Remarkably, 21% (10 out of 47) of the genes that are deleted in 06_3928A encode RXLR effectors (see appendix 4.3). 13_A2 MLG isolates are able to infect potatoes carrying the *R1* gene (David Cooke, unpublished). I discovered that a ~18 Kb deletion encompassing the *Avr1* RXLR effector gene underpins the ability of the 06_3928A isolate to infect *R1* potatoes (van der Lee et al., 2001; Vleeshouwers et al., 2011) (Fig. 6.7).

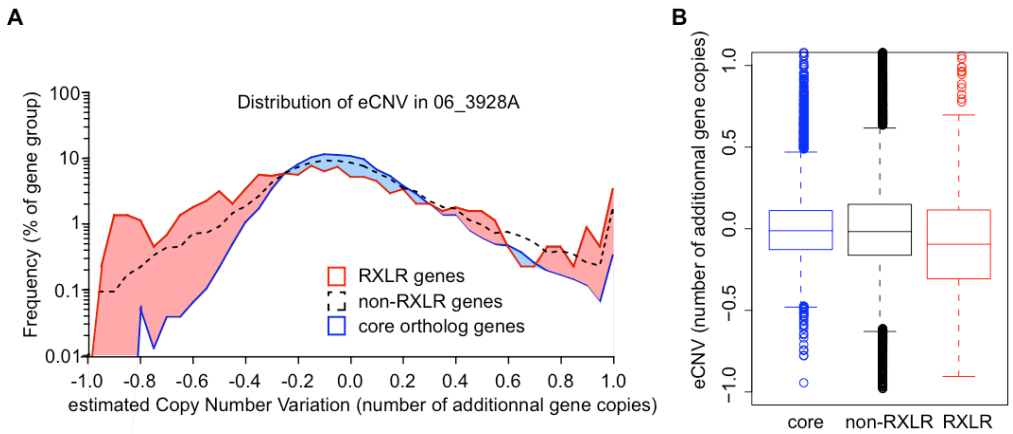


Fig. 6.5. Distribution of CNV in *P. infestans* 06_3928A genome
(A) Percentage of genes showing CNV in RXLRs, non-RXLRs and core ortholog genes (see chapter 2 section 2.4.11). (B) Box plot showing the distribution of estimated eCNV in RXLRs, non-RXLRs and core ortholog gene groups (average, first and third quartile, first values outside 1.5 times the interquartile range are shown).

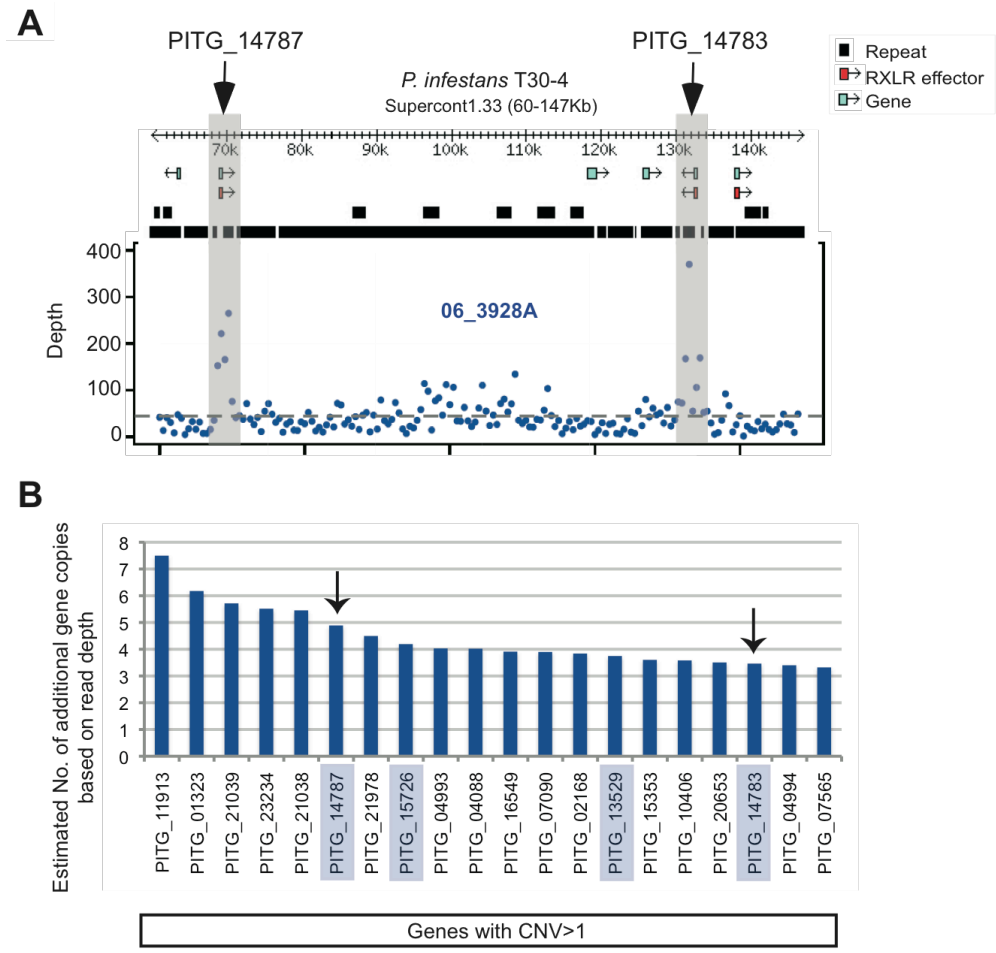


Fig. 6.6. Example of duplication events found in RXLR genes of *P. infestans* 06_3928A

(A) Depth of coverage plot showing duplication events in PITG_14787 and PITG_14783 RXLR genes in 06_3928A isolate. Alignment view of a 70 Kb genomic region of *P. infestans* from supercont1.33 containing PITG_14787 and PITG_14783 RXLR genes. Screen shot image at the top of the alignment is taken from *P. infestans* SybilLite genome browser (see chapter 2 section 2.3). Repeats are in black, genes are in green and RXLR effector genes are in red. The 70 Kb genomic region was scanned with a window size of 500 bp in the genome 06_3928A isolate with blue dots representing the average of 500 bp. Region where sequence reads from 06_3928A aligned to PITG_14787 or PITG_14783 genes are highlighted within grey vertical bars. Dashed grey lines indicate the genome average depth of coverage. (B) Histogram showing the top 20 genes from 06_3928A with additional gene copies compare to T30-4 strain (see chapter 2 section 2.4.11). The genes that are highlighted in blue boxes in the x-axis correspond to RXLR effectors. PITG_14787 gene shows the highest number of additional genes copies (~4-5) among the RXLRs. PITG_14787 and its paralog gene PITG_14783 are pointed with black arrows.

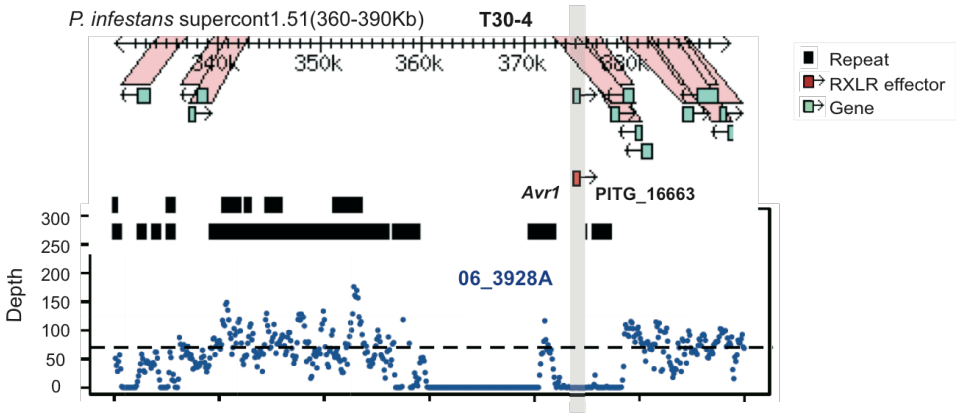


Fig. 6.7. Example of *Avr1* deletion in *P. infestans* 06_3928A
(A) Plot of sequencing depth of coverage of Illumina reads from isolate 06_3928A aligned to the region of supercontig 1.51 from T30-4 strain containing the avirulence effector *Avr1* (PITG_16663) (in red). The 30 Kb genomic region was scanned with a window size of 200 bp in the genome 06_3928A isolate with blue dots representing the average of 200 bp. Screen shot image at the top of the alignment is taken of *P. infestans* SybilLite genome browser (see chapter 2 section 2.3). Repeats are in black, genes are in green and RXLR effector genes are in red. Region where sequence reads from 06_3928A aligned to *Avr1* gene is highlighted within grey vertical bars. Dashed grey lines indicate the genome average depth of coverage. Note the ~18 kb sub-region (from 361 to 379 Kb) that shows reduced coverage in reads from isolate 06_3928A indicating high sequence divergence in this isolate.

6.2.1.4. Assembly of unmapped reads from *P. infestans* 06_3928A isolate reveal novel candidate RXLR effectors

To identify 06_3928A sequences absent from T30-4 genome, I performed *de novo* assembly of the unmapped Illumina reads and identified a total of 2.77 Mb contigs that did not align to T30-4 sequences (see chapter 2 section 2.4.4). *Ab initio* and homology based gene calling revealed six novel candidate RXLR effector genes that are absent in T30-4 strain (Fig. 6.8, see chapter 2 section 2.4.5, appendix 4.4). PCR validation showed the absence of these six assembled RXLR genes in the *P. infestans* strain T30-4 (PCR data from David Cooke, unpublished) (Table 6.5, see chapter 2 section 2.4.5). One of these *de novo* assembled RXLR genes is a highly polymorphic variant of *Avr2* that evades recognition by the *R2* resistance gene and explains virulence of 06_3928A on *R2* potatoes (Gilroy et al., 2011). These findings point to a series of genetic events that may explain the aggressiveness and virulence phenotype of the 13_A2 MLG.

Table 6.5. PCR validation of candidate assembled RXLR effectors from unmapped Illumina reads of *P. infestans* 06_3928A

P. infestans strain	MLG	PCR product amplification for					
		Pex644	*Pex50259*	*Pex30588*	*Pex46622*	*Pex15083**	*Pex14182*
T30-4	**Misc**	**-**	-	-	**-**	-	-
2006_3928A	**13_A2**	**+**	**+**	**+**	**+**	**+**	**+**
2006_3884B	13_A2	+	+	+	+	+	+
2006_3964A	13_A2	+	+	+	+	+	+
2006_4132B	13_A2	+	+	-	+	+	+
2006_4012F	3_A2	-	+	-	+	+	-
2006_4244E	3b_A2	-	+	-	+	+	+
2006_3936C2	10_A2	-	+	-	+	-	+
2006_4440C	10_A2	-	+	-	+	-	+
2004_7804B	15_A2	-	-	-	-	-	+
2006_3992G	16_A2	+	+	+	+	+	-
2006_4388E	17_A2	-	+	-	-	+	+
2003_25_1_3	22_A2	-	+	+	-	+	+
2003_25_3_1	22_A2	+	+	+	-	+	+
2006_3984C	1_A1	+	+	+	-	+	+
2006_4304A	1_A1	+	+	+	-	+	+
2006_3888A	2_A1	+	+	+	+	+	+
2006_4068B	2_A1	+	+	+	+	+	+
2006_3960A	2_A1	-	+	+	+	+	+
2006_4352E	4_A1	+	-	+	+	-	+
1996_9_5_1_C4	5_A1	-	-	+	-	+	+
07_5866C	5g_A1	-	-	+	+	+	+
2006_3920A	6_A1	+	+	-	+	-	+
2006_4100A	6_A1	+	+	-	+	-	+
2006_4168B	7_A1	+	-	-	-	+	+
2006_4168C	7_A1	+	-	-	-	+	+
2006_4232E	8_2a_A1	+	-	-	-	+	+
2006_4256B	8a_A1	+	-	-	-	+	+
2006_4320F	12_A1	+	-	-	-	+	+

* *Pex15083* was identified in this study as a candidate assembled RXLR effector gene. Pex15083 amino acid sequence corresponds to the Avirulence protein AVR2-LIKE variant in *P. infestans* 06_3928A isolate (Gilroy et al., 2011).

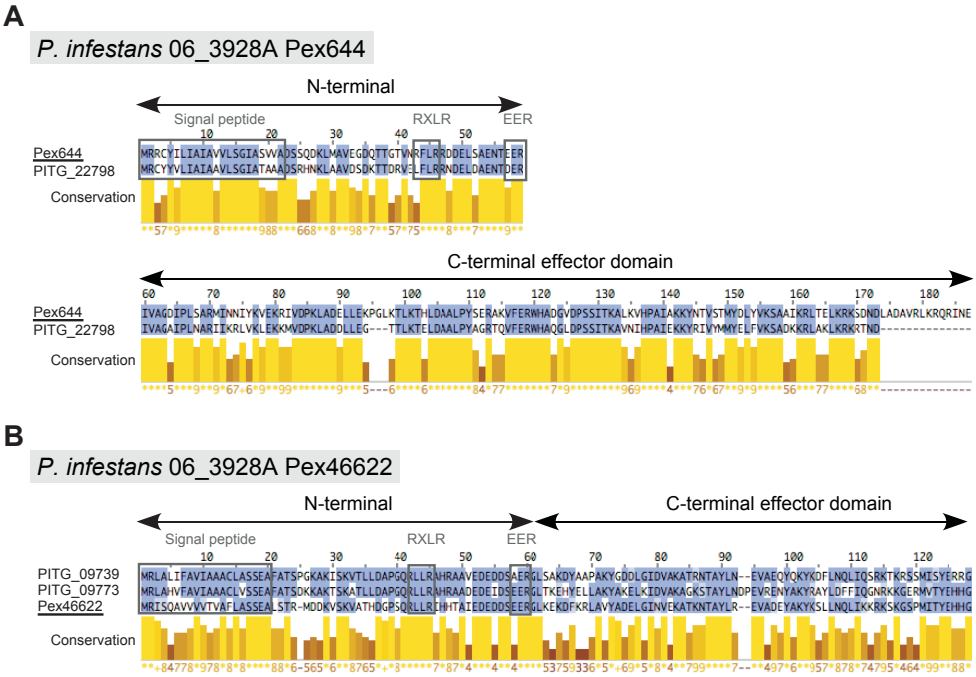


Fig. 6.8. Sequence alignment of *de novo* assembled RXLRs of *P. infestans* 06_3928A with similarity to RXLRs of the reference genome strain T30-4
(A) Pex644 candidate RXLR in 06_3928A show similarity to *P. infestans* T30-4 PITG_22798, a RXLR gene with no paralogs. (B) Pex46622 candidate RXLR in 06_3928A show similarity to PITG_09739 and PITG_09773, two genes that belong to the RXLR family6 in T30-4 strain. Signal peptides, RXLR and EER motifs are marked in grey boxes.

6.2.2. Genome-wide expression analysis of *P. infestans* 06_3928A

6.2.2.1. Gene expression polymorphisms: gain and loss of gene induction in RXLR effectors of *P. infestans* 06_3928A

I hypothesized that the phenotype of the 13_A2 multilocus genotype (MLG) not only results from changes in the gene coding sequences documented above, but also in changes in the regulation of gene expression. To identify gene expression polymorphisms we performed an infection time course by hybridizing NimbleGen microarrays with RNA from potato leaves harvested 2, 3 and 4 days post inoculation (dpi) with the 06_3928A isolate. I then compared the gene expression profiles obtained with 06_3928A to T30-4 and to NL07434, an isolate that originates from the sexual populations of the Netherlands, where the 13_A2 MLG was first detected (see chapter 2 section 2.5.1). I observed significant expression

polymorphisms between the three strains with 1,123 genes specifically induced in 06_3928A, compared to 110 in T30-4 and 891 in NL07434 (Fig. 6.9A). In total, only 398 out of 4,934 genes were induced in all three strains indicating distinct sets of genes induced during infection of potato (Fig. 6.9A) (see appendix 4.5). *P. infestans* effector genes are sharply induced during the biotrophic phase of infection, when the pathogen associates closely with living plant cells (Haas et al., 2009; Vleeshouwers et al., 2011). I identified 104 RXLR effector genes in 06_3928A induced during biotrophy compared to only 79 and 68 in strains T30-4 and NL07434, respectively (Fig. 6.9A and see appendix 4.1). Of these 104 RXLR genes, 20 were specifically induced in 06_3928A isolate but not in the other two strains (Fig. 6.9A and Fig. 6.10A). In contrast, 18 RXLR effector genes are not induced in 06_3928A but are induced in at least one of the other strains (Fig. 6.9A and Fig. 6.10B). One of these genes is *Avr4*, encoding AVR4 avirulence effector recognized by the R4 resistance protein (van Poppel et al., 2008) (Fig. 6.10B). The lack of induction of *Avr4* in 06_3928A is consistent with the observed virulence of 13_A2 isolates on *R4* potatoes (David Cooke, unpublished).

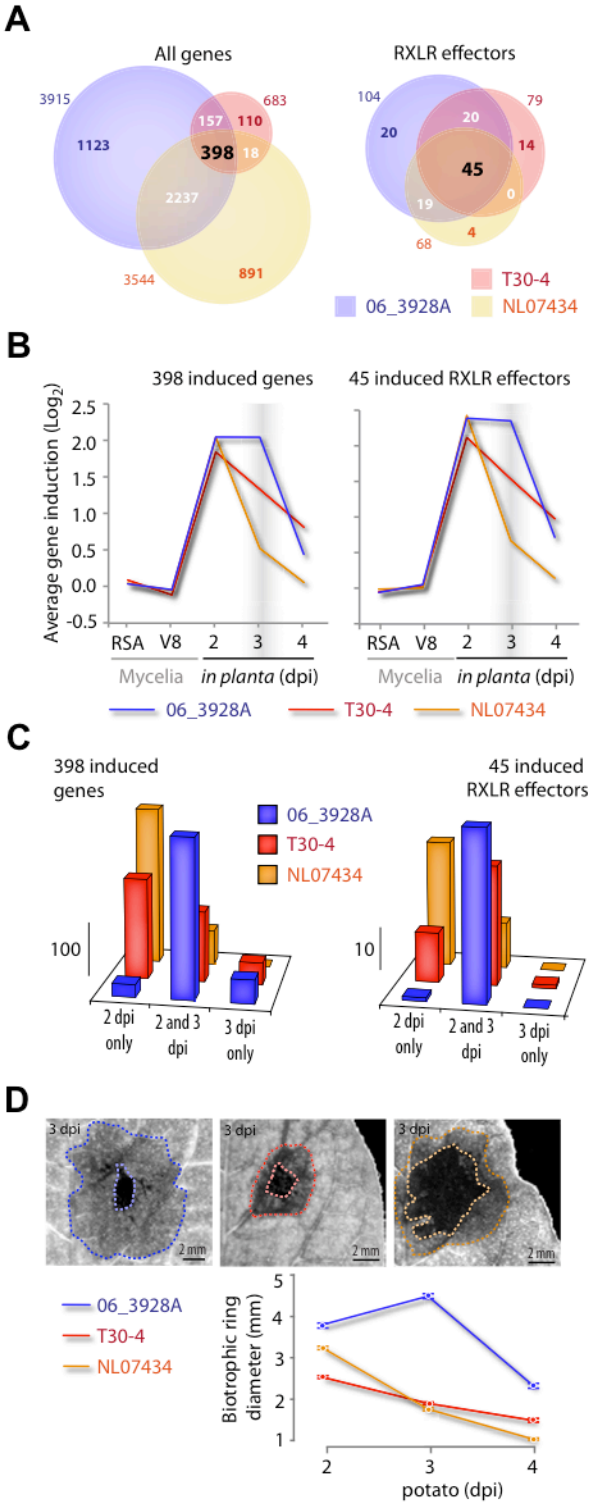


Fig. 6.9. Sustained induction of genes during the biotrophic phase of infection in *P. infestans* 06_3928A

(A) Venn diagram showing the distribution of *in planta*-induced genes between 06_3928A, NL07434 and T30-4 strains. Gene induction in potato time points relative to mycelia (see chapter 2 section 2.5.1). (B) Sustained induction at 2 and 3 dpi during infection in potato in the strain 06_3928A. (C) Number of genes induced according to the

time of induction in potato: (i) 2 dpi only, (ii) 2 and 3 dpi and (iii) 3 dpi only. In panels (A), (B) and (C) Left side correspond to all genes and right side to RXLRs. (D) Diameter measurements (mm) equivalent to the biotrophic growth during infection in potato shows a longer biotrophic growth 2-3 dpi in the strain 06_3928A compared to the strains T30-4 and NL07434 (see chapter 2 section 2.5.2).

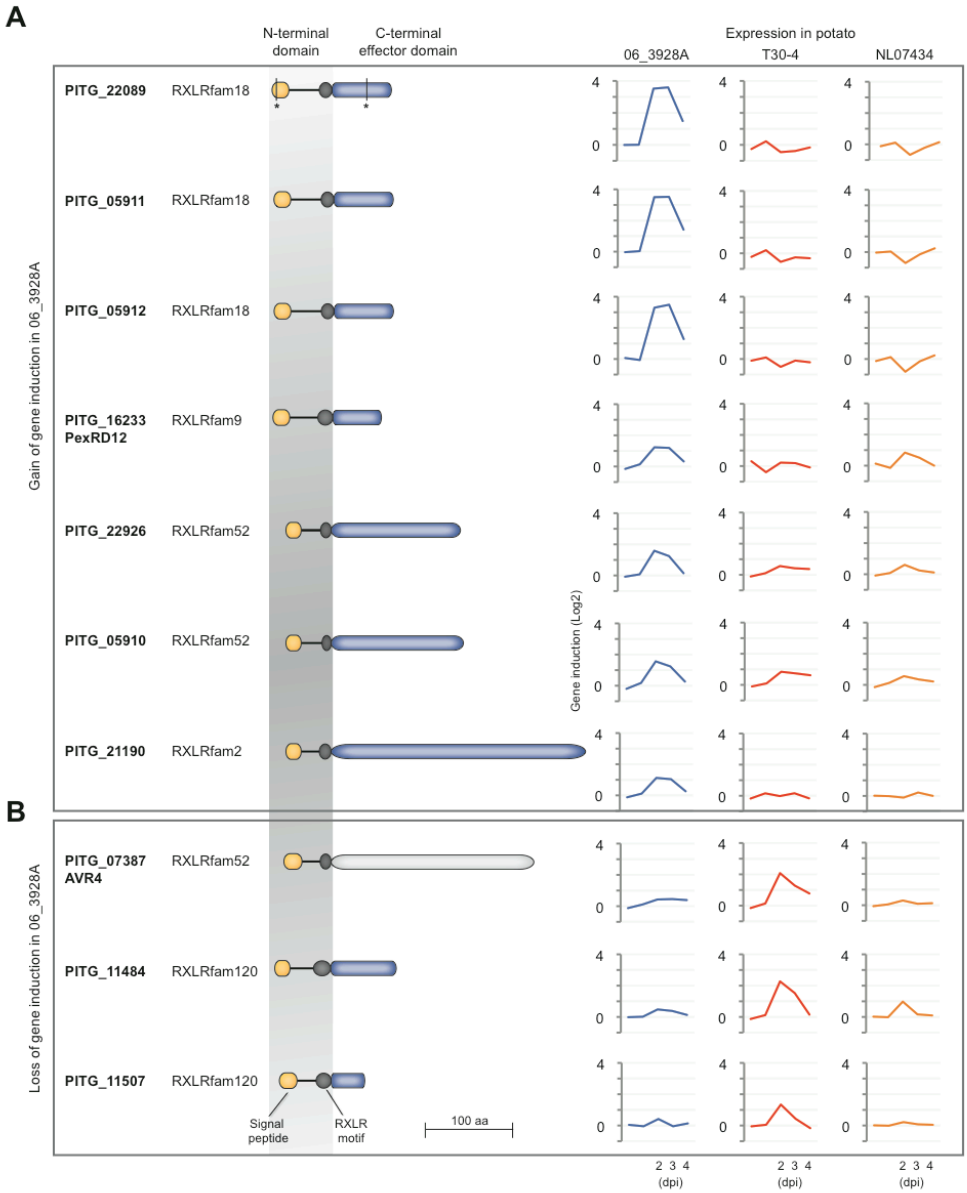


Fig. 6.10. Examples of RXLR effectors showing gene expression polymorphisms in *P. infestans* 06_3928A

Examples of gain (top box) and loss (bottom box) of induction in *P. infestans* 06_3928A RXLR effectors genes showing gene structures (left part) and gene expression patterns (right part) (see chapter 2 section 2.5.1). A). Genes that gain gene induction in 06_3928A isolate. B). Genes that loss gene induction in 06_3928A isolate, but they are induced in at least one of the other two strains T30-4 an/or NL07434. N-terminal (signal peptide) domain, RXLR motif, and C-terminal effector domain are shown in yellow, dark grey and blue respectively. A vertical bar placed in line with an asterisk show a polymorphic amino

acid site in the effector protein of 06_3928A compared to T30-4 strain. The effector domain coloured in light grey is indicative of a change in the ORF compare to T30-4 that resulted in a pseudogene in 06_3928A isolate. The gene expression time course during infection of potato (2-4 dpi) is given for three *P. infestans* strains: 06_3928A (blue), T30-4 (red) and NL07434 (orange).

6.2.2.2. *P. infestans* 06_3928A shows patterns of sustained gene induction and extended biotrophic growth during potato infection

I noted a markedly distinct temporal pattern of gene induction *in planta* in 06_3928A. Whereas in T30-4 and NL07434 gene expression generally declines at 3 dpi, when the pathogen starts shifting to the necrotrophic phase (death of the plant tissue) of the disease, genes that are induced in 06_3928A showed sustained induction over 2 and 3 dpi (Fig. 6.9B-C). These findings prompted to determine the extent to which disease progression differs between 06_3928A and other isolates. Microscopic observations of lesions caused by 06_3928A revealed significantly larger biotrophic zones during infection (Fig. 6.9D and see chapter 2 section 2.5.2). The ability of 06_3928A to establish an extended biotrophic phase during colonization of host plants may explain the enhanced aggressiveness of 13_A2 isolates. Indeed, I noted that the 194 genes encoding secreted proteins and showing an extended induction period in 06_3928A include putative virulence factors such as RXLR effectors, cell wall hydrolases and protease inhibitors (see appendix 4.6).

6.2.2.3. Proposed strategies for the management of epidemics caused by *P. infestans* 13_A2

The genome analysis of the 13_A2 MLG offers opportunities for identifying targets for genetic resistance breeding in plants. We scanned 06_3928A genes that are induced *in planta* for RXLR effectors with avirulence activities. Among these, three genes, *Avrblb1, Avrblb2* and *Avrvnt1* occur as intact coding sequences that are highly induced in potato (Fig. 6.11). To determine the extent to which 13_A2 MLG can infect plants carrying the corresponding *Rpi-blb1, Rpi-blb2* and *Rpi-vnt1.1* resistance genes, I inoculated 06_3928A on tester potato lines expressing each of these three *R* genes. In all cases, 06_3928A was unable

to infect the *R* potatoes and triggered a typical hypersensitive response (Fig. 6.11). These results indicate that the three *R* genes are effective against the 13_A2 MLG and could be used to temper epidemics caused by this aggressive clone of *P. infestans*.

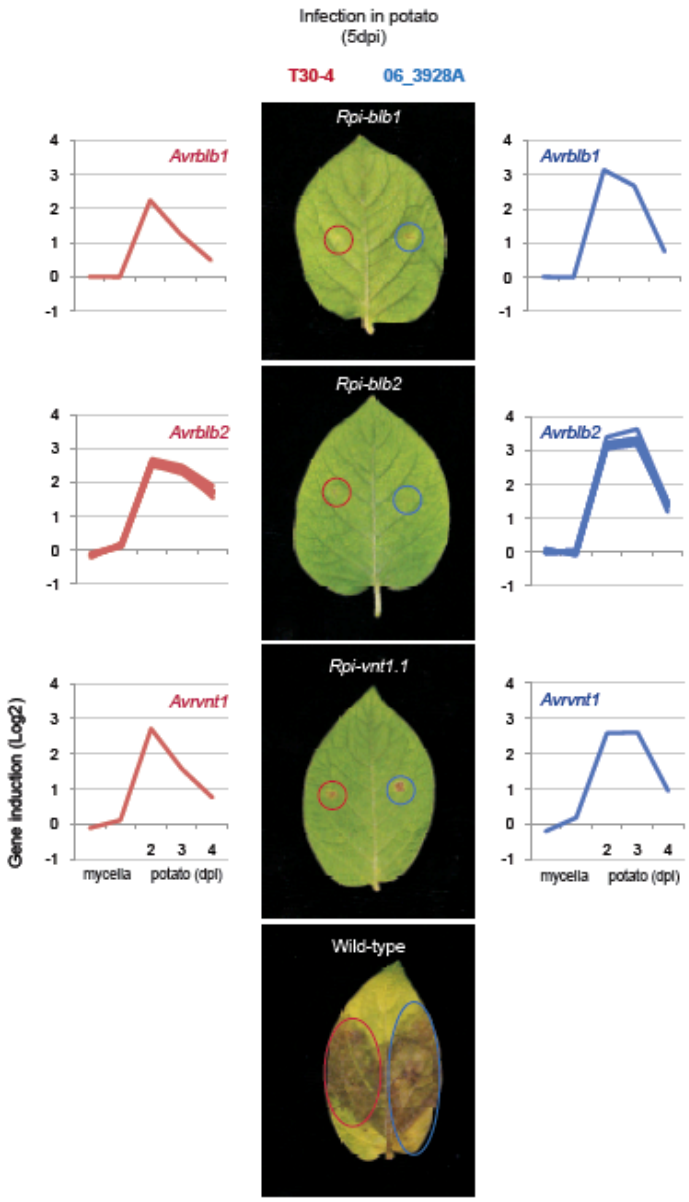


Fig. 6.11. *P. infestans* 06_3928A carries invariant *Avrblb1, Avrblb2* and *Avrvnt1* genes that are induced *in planta*
Gene expression profiles of *Avrblb1*, *Avrblb2* and *Avrvnt1* during a time course infection on potato in *P. infestans* T30-4 (red, left) and 06_3928A (blue, right). Infections of *P. infestans* 06_3928A strain in *Rpi-blb1*, *Rpi-blb2*, *Rpi-vnt1.1* transgenic potato plants and wild type (middle).

6.3. Conclusions

In this chapter, I reported the genome sequencing and gene expression profiling of a clonal lineage 13_A2 of *P. infestans*. I showed that 06_3928A isolate exhibit sequence and gene expression polymorphisms, particularly in RXLR effector genes. In 06_3928A, distinct expression profiles of RXLR effector genes of 06_3928A may collectively explain the enhanced aggressiveness and ability to infect resistant potato varieties. The genome analysis proved particularly valuable in highlighting potential "Achilles' heel" of 13_A2, namely the three RXLR effectors that are sensed by the disease resistance genes *Rpi-blb1, Rpi-blb2* and *Rpi-vnt1.1* (Fig. 6.11). Therefore, deployment of these *R* genes in agriculture, either through classically breeding or transgenic potato varieties, should buffer the spread of the 13_A2 MLG strains or help to manage this aggressive form of the late blight disease. In the future, combining genome analyses with a better understanding of the geographical structure and dynamics of *P. infestans* populations should help to detect and manage emerging aggressive races of this pathogen before they reach epidemic proportions.

CHAPTER 7: Differentially regulated plant genes in the interaction with pseudoflower-forming rust fungus *Puccinia monoica*

7.1. Introduction

Boechera stricta (*Arabis drummondii*) belongs to the Brassicaceae and is mostly present in montane and alpine regions of the western North America. It is infected in late summer by wind-born basidiospores of the rust fungus *Puccinia monoica* produced on the primary host *Trisetum spicatum* (L) Ritcher (Agriculture, 1960; Farr et al., 1989; Roy, 1993a). *P. monoica* inhibits flowering in its host and radically transforms host morphology, producing flower-like structures (pseudoflowers) that mimic true flowers in shape, size, color and nectar production from co-occurring and unrelated yellow-flowered angiosperms such as the buttercup *Ranunculus inamoenus* (Roy, 1993b, 1994). Although pseudoflowers are visually similar to the true flowers from buttercups, they produce a distinct sweet fragrance that allows them to attract insect visitors (Roy, 1994; Roy and Raguso, 1997). The formation of pseudoflowers is critical to the life cycle of this rust fungus. By forming pseudoflowers *P. monoica* attracts flying insect visitors that contribute to the dissemination of spores and sexual reproduction (Roy, 1993a, 1996).

How *P. monoica* produces pseudoflowers after infection of *B. stricta* plant is still unknown. I hypothesised that secreted effector proteins are produced by *P. monoica* to alter various biological processes in *B. stricta* apical meristem cells leading to the development of pseudoflowers. It is known that filamentous plant pathogens can secrete an arsenal of effector molecules to modify host physiology and to successfully colonize its host (Birch et al., 2006; Hogenhout et al., 2009; Kamoun, 2007; Schornack et al., 2009; Stassen and Van den Ackerveken, 2011). To discover and investigate the functions of effector molecules from *P. monoica* will be necessary to generate genome and/or transcriptome sequence data, which is not currently available. Therefore, the

study of pathogen effectors is difficult at present due to limitations with generating sequence information for *P. monoica.*

Another important unknown in this system is what are the effects of pathogen effectors in the infected compared to the uninfected plants? These hypothetical molecular alterations in the host plant *B. stricta* underlying the development of *P. monoica* pseudoflowers have not been described yet. To investigate the transcriptional changes occurring in the *B. stricta* plant during the formation of pseudoflowers, I used a whole-genome microarray of *A. thaliana* and hybridized it with infected pseudoflowers. Here, I highlight plant genes that are differentially regulated during *P. monoica* - *B. stricta* interaction and that could potentially contribute to the formation of pseudoflowers. This study is a first step towards understanding at a molecular level how this rust fungus pathogen manipulates its host plant.

7.2. Results and discussion

7.2.1. Identification of genes significantly regulated in pseudoflowers

To document transcriptional changes in pseudoflowers structures triggered by infection by the rust fungus *Puccinia monoica*, I extracted RNA from field-collected samples: (i) uninfected plant stems and leaves ('SL'), (ii) uninfected plant flowers ('F') and (iii) pseudoflowers from *P. monoica* infected plants ('Pf') (Fig. 7.1, see chapter 2 Table 2.4). NimbleGen microarray services were utilized for cDNA preparations from the extracted RNA, and subsequent chip hybridizations to an *Arabidopsis thaliana* custom array design, a close relative to *B. stricta* (see chapter 2 section 2.6.2). For the analysis of the microarray data. I carried out a t-test to detect genes showing a significant (P-value <0.05) differential expression in three comparisons: 'Pf' *vs* SL (9,173 genes), 'F' *vs* 'SL' (6,851 genes) and 'Pf' *vs* 'F' (9,137 genes) (Fig. 7.2, see chapter 2 section 2.6.2). Then, I used the Rank Products (RP) program to estimate False Discovery Rate (FDR) for differential regulation of individual genes using permutations with no assumption about distribution of the data. RP analysis is recommended for

samples not coming from controlled laboratory conditions (Kammenga et al., 2007). The resulting FDR values are indeed considered more robust than P-values estimated under the assumption of normal t-distribution as in the t-test method. This method is therefore well adapted to the analysis of data from the field-collected samples that are subject to natural environmental variations. Using RP analysis I found significantly differentially regulated (RP-value <0.05) in each of the three comparisons: 'Pf' *vs* 'SL' (1,036 genes), 'F' *vs* 'SL' (910 genes) and 'Pf' *vs* 'F' (687 genes) (Fig. 7.2, see chapter 2 section 2.6.2). To identify genes significantly regulated across the various samples and that overlap between the t-test and RP analysis, I compared the significant gene lists obtained with both tests and found the following number of overlapping genes in each of the three comparisons: 'Pf' *vs* 'SL' (948 genes), 'F' *vs* 'SL' (859 genes) and 'Pf' *vs* 'F' (611 genes) (Fig. 7.3, see chapter 2 section 2.6.2, appendix 5.1 and appendix 5.2).

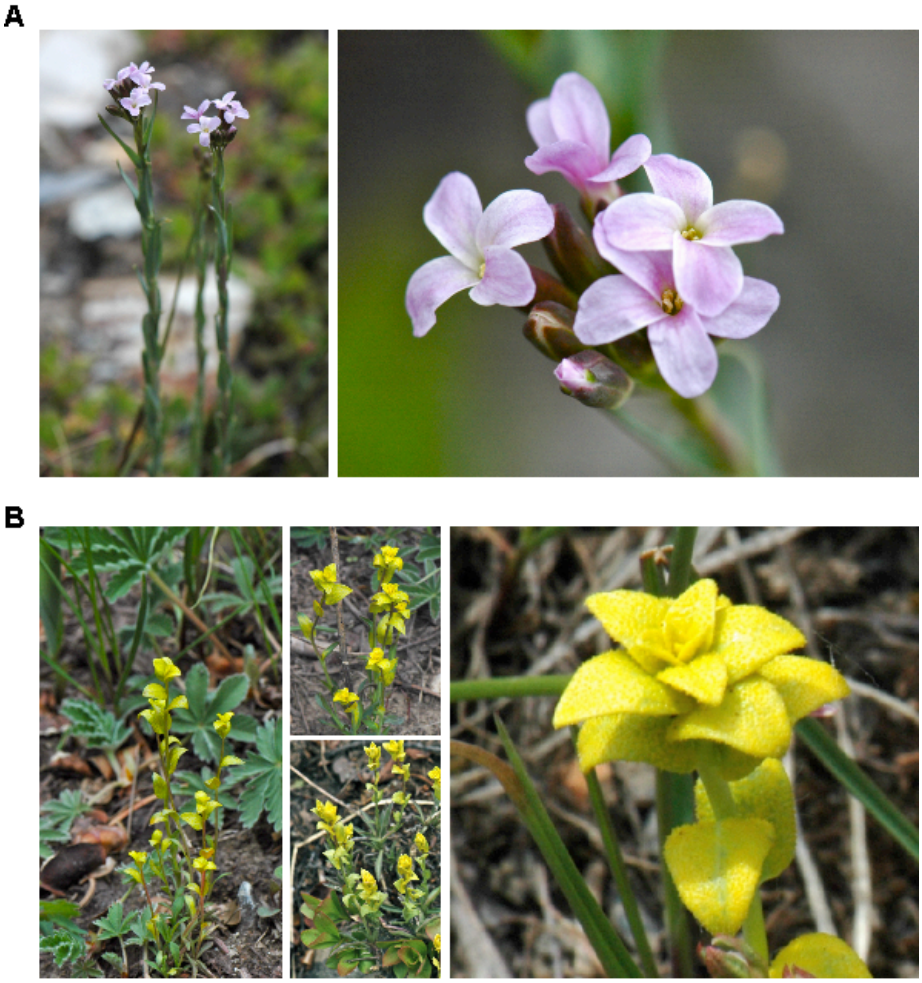


Fig. 7.1. Illustration of floral mimicry produced by the pseudoflower-forming rust fungus *Puccinia monoica*
(A) Picture of uninfected flowering *Boechera stricta* plant (left) and a close up picture of its light pink flowers (right). (B) Pictures of vegetative tissues of *B. stricta* plants that produces pseudoflowers upon infection with *Puccinia monoica* (left) and a close up of a yellow *P. monoica* pseudoflower (right). Professor Sophien Kamoun and I collected the uninfected *B. stricta* (A) and pseudoflowers (B) from near Gunnison, CO, USA and used them for this study.

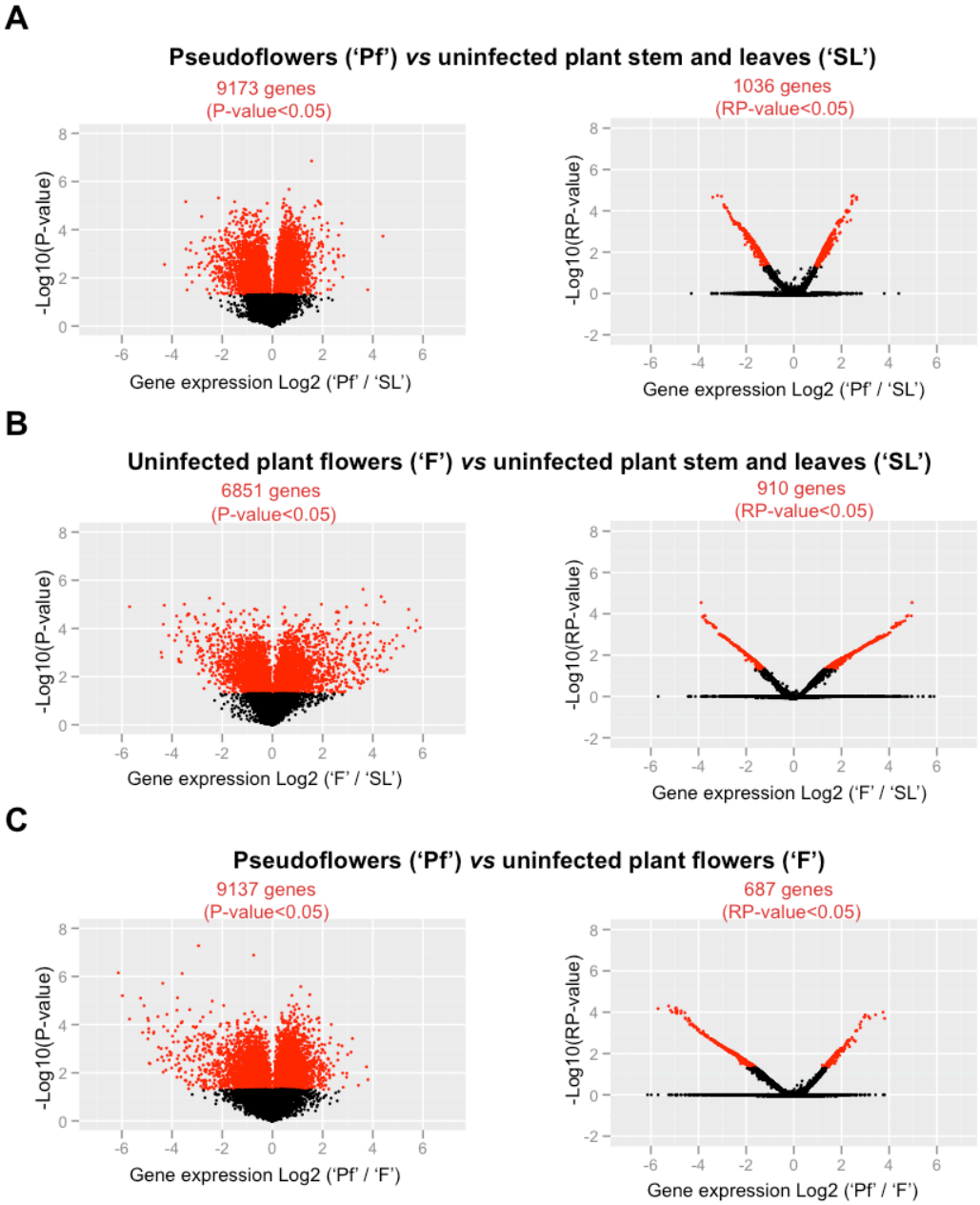


Fig. 7.2. Changes in gene expression in three comparisons using t-test (left) and rank products (RP) (right) analyses
(A) Volcano plots showing changes in gene expression in pseudoflowers from *P. monoica* infected plants ('Pf') *vs* uninfected *Boechera stricta* plant stem and leaves ('SL'). 9,173 (left) and 1,036 (right) differentially expressed genes with P-value <0.05 and with RP-value <0.05, respectively. (B) Volcano plots showing changes in gene expression in uninfected *B. stricta* plant flowers ('F') *vs* 'SL'. 6,851 and 910 differentially expressed genes with P-value <0.05 and with RP-value <0.05, respectively. (C) Volcano plots showing changes in gene expression in 'Pf' *vs* 'F'. 9,137 and 687 differentially expressed genes with P-value <0.05 and with RP-value <0.05, respectively**.** Statistical analyses were performed using a t-test and RP (see chapter 2 section 2.6.2). Individual genes are represented as points. Log2 of fold change in replicate samples from 'Pf 'or 'F' relative to 'SL' and 'Pf' relative to 'F' (x-axis) is plotted against negative Log10 of P-value or RP-value (y-axis). Red points indicate significant genes with a P-value or RP-value criterion

of less than 0.05. Black points indicate no significant genes with a P-value or RP-value greater than 0.05.

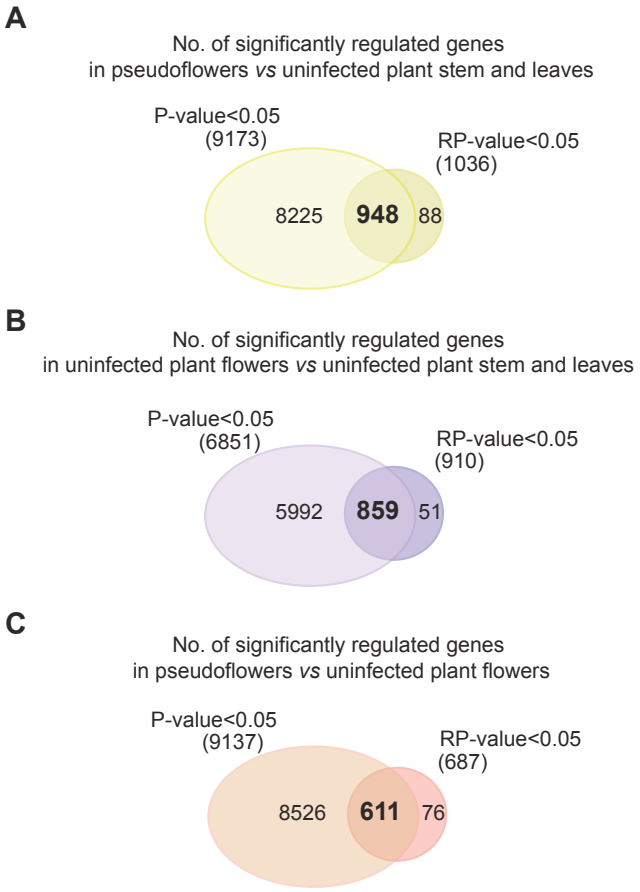


Fig. 7.3. Significantly regulated genes detected with both t-test and rank products (RP) analyses in three comparisons
(A) 948 genes are significantly regulated in pseudoflowers from *P. monoica* infected plants ('Pf') *vs* uninfected *Boechera stricta* plant stem and leaves ('SL') using both t-test and RP analyses. (B) 859 genes are significantly regulated in uninfected *B. stricta* plant flowers ('F') *vs* 'SL' using both t-test and RP analyses. (C) 611 genes are significantly regulated in 'Pf' *vs* 'F' using both t-test and RP analyses.

Table 7.1. Genes significantly up and down-regulated in pseudoflowers ('Pf') or uninfected plant flowers ('F') *vs* uninfected plant stem and leaves ('SL')

Description*	Pseudoflowers ('Pf') *vs* uninfected plant stem and leaves ('SL')	Flowers ('Pf') *vs* uninfected plant stem and leaves ('SL')
No. of significantly regulated genes	948	859
No. of up-regulated genes	454	429
No. of down-regulated genes	494	430

*Genes significantly regulated in both t-test and RP statistical analysis (see chapter 2 section 2.6.2).

RP analysis generates two RP-values for each gene that indicates the probability of being up or down-regulated, respectively (see chapter 2 section 2.6.2) (Breitling et al., 2004). For each of the three lists of genes significantly regulated, I classified genes as up or down-regulated based on the generated RP-values. In 'Pf' *vs* 'SL' comparison, 420 were up and 301 were down-regulated genes with RP-values< 0.05 out of the 948 significant significantly regulated genes. In the 'F' *vs* 'SL' comparison, 395 were up and 237 down-regulated genes out of the 859 significant significantly regulated genes (Table 7.1, appendix 5.1 and appendix 5.2).

7.2.2. Functional classification of genes significantly regulated in pseudoflowers

To identify plant biological processes significantly altered during the formation of pseudoflowers, I performed a Gene Ontology enrichment analysis with the set of 948 and 859 genes differentially regulated in the comparisons: (i) pseudoflowers from *Puccinia monoica* infected plants *vs* uninfected *Boechera stricta* stem and leaves ('SL') and (ii) uninfected *B. stricta* flowers ('F') *vs* uninfected *Boechera stricta* stems and leaves ('SL'); and then focussed on biological processes annotations (see chapter 2 section 2.6.3). These processes are shown as circular nodes in the Fig. 7.4, of size proportional to the P-value of a t-test for enrichment among significantly regulated genes. To document the overall regulation exerted on biological processes enriched among regulated genes, I calculated the average induction fold (as log2 values) for all genes significantly regulated

matching a given gene ontology in both comparisons (Fig. 7.4, see appendix 5.3 and appendix 5.4).

Biological processes that were specifically down-regulated and enriched in the comparison *P. monoica*-induced pseudoflowers ('Pf') *vs* uninfected *B. stricta* stem and leaves ('SL') but up-regulated in uninfected *B. stricta* flowers ('F') vs 'SL' included: (1) reproduction (GO:0000003), (2) floral organ development (GO:0048437) and (3) regulation of transcription (GO:0045449) (see appendix 5.3 and appendix 5.4). In addition, some biological processes that were specifically down-regulated in 'Pf' vs 'SL' included: (1) maintenance of floral meristem identity (GO:001076), anthocyanin biosynthetic process (GO: 00009718) and monoterpenoid biosynthetic process (G0:0016099) (Fig. 7.4, see appendix 5.3). The observation of down-regulation of genes involved in maintenance and development of the floral organ was expected since *P. monoica* infected plants develop an elongated stem that fails to form flowers (Roy, 1993a).

In contrast, biological processes that were specifically up-regulated and enriched in the comparison 'Pf' *vs* 'SL' included: shoot development (GO:0048367), cotyledon development (GO:0048825), leaf development (GO:0048366), leaf morphogenesis (GO:0009965), L-phenylalanine metabolism (GO:0006558), carbohydrate transport (GO: 0034219), wax biosynthesis (GO:0006633) and fatty acid biosynthesis (GO:0010025) (Fig. 7.4, see appendix 5.3). These results suggest that the construction of the pseudoflowers involves an extensive reprogramming of the host including control of shoot and leaf development, synthesis of volatiles and modifications of the cell wall surface. All together, these modifications will result in elongated stems and the formation of clusters of flower-like leaves covered by nectar and wax secretion (Roy, 1993a). In addition, the down-regulation of monoterpenoid biosynthetic process indicates that *P. monoica* is not using the floral organ scent production of the host, but instead new compounds are being synthesized in pseudoflowers. Pseudoflowers distinct fragrance contains phenylacetaldehyde and 2-phenylethanol that are chemically different compounds compare to the terpenoids produced in the uninfected flowers but with similar function as they can efficiently attract pollinators (Raguso and Roy, 1998; Roy, 1993a).

The differential regulated biological processes mentioned above detected in the 'Pf' vs 'SL' comparison are proposed as key processes that could explain the remarkable developmental changes in *P. monoica* induced pseudoflowers. These key biological processes confirm previous postulates suggesting that *P. monoica* does not exploit the flower of the plant, but instead manipulates the host to generate pseudoflowers. These pseudoflowers have different shape, color, nectar and scent compare to uninfected flowers (Roy, 1993a). Pseudoflowers resembles flowers from unrelated co-occurring plant species and function efficiently in the attraction of pollinators acting in benefit of the rust fungus reproduction (Roy, 1993a). Among these key biological processes, I identified 65 gene candidates (35 and 30 were up and down-regulated, respectively) with significant altered gene expression in the 'Pf' *vs* 'SL' comparison, which will be explained in more detail in the sections below (Table 7.2 and Fig. 7.4).

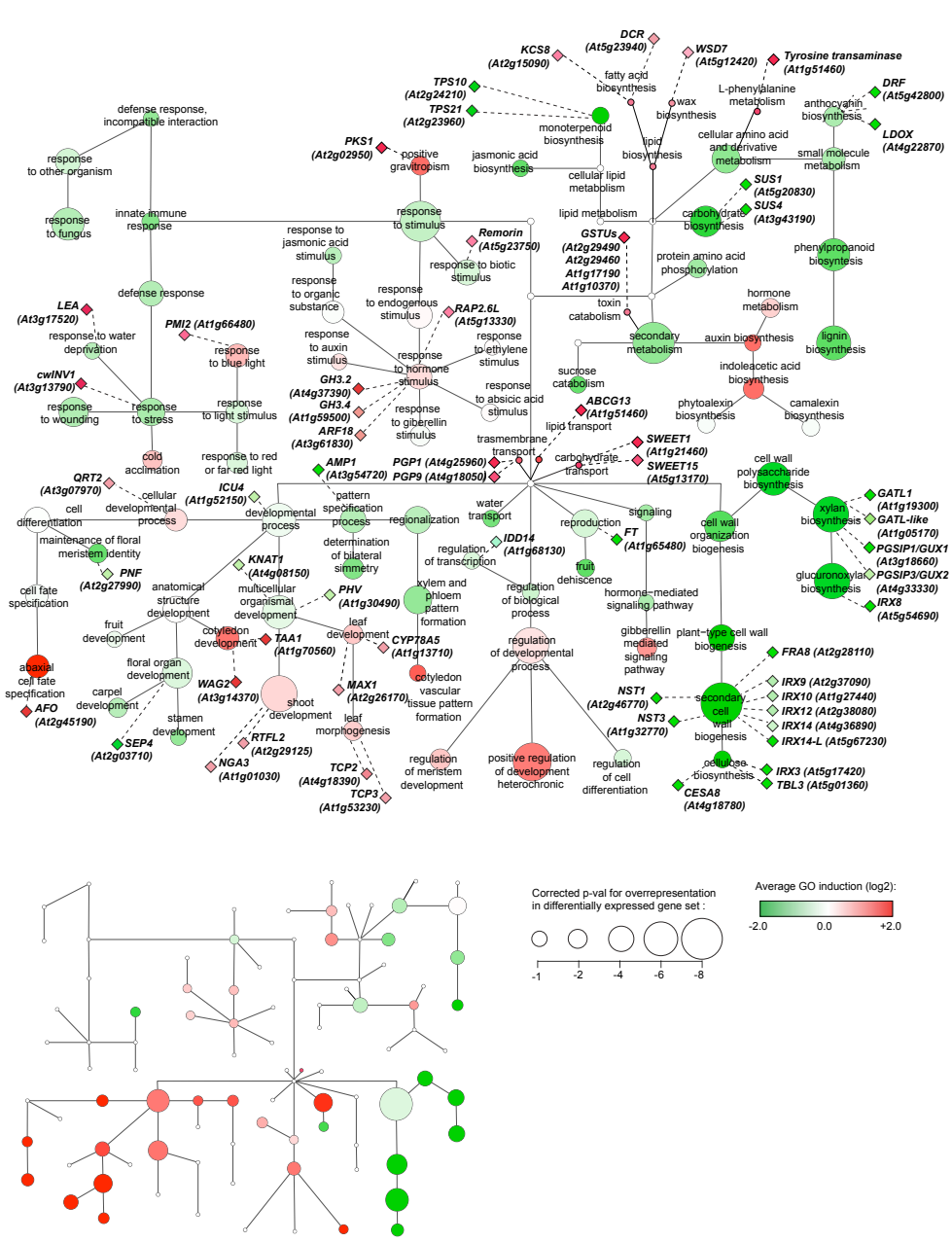


Fig. 7.4. Overview of biological processes altered in *Puccinia monoica*-induced pseudoflowers ('Pf') and *Boechera stricta* flowers ('F') compared to stem and leaves ('SL')

A) Gene Ontology Biological Processes (GOBP) network showing processes enriched among genes with expression altered in *P. monoica*-induced pseudoflowers compared to *B. stricta* stems (node size) with average induction fold for genes involved in each process shown as a color code (from green for average induction folds <0 to red for average induction folds >0). Some nodes and edges have been omitted for clarity. Genes highlighted in the text are indicated with diamonds connected to dashed lines to the processes they are involved in (see chapter 2 section 2.6.3). (B) GOBP network showing processes enriched among genes with expression altered in *B. stricta* flowers compared to stems, with the same network topology as in A.

Table 7.2. Candidate genes with altered gene expression in *Puccinia monoica*-induced pseudoflowers ('Pf') compared to uninfected *Boechera stricta* stems and leaves ('SL')

AGI code	Gene name	Common name	Annotation	Gene expression ('Pf' / 'SL')	Ratio Log2 ('Pf' / 'SL')	P-value[a]	FDR value[b]	GOBP[c]	GOPB Description
At4g25960	*P-GLYCOPROTEIN2*	*PGP1*	Altered morphogenesis	Up-regulated	1.04	2.22E-03	2.59E-02	55085	transmembrane transport
At4g18050	*P-GLYCOPROTEIN9*	*PGP9*	Altered morphogenesis	Up-regulated	2.10	4.04E-03	2.00E-04	55085	transmembrane transport
At1g52150	*INCURVATA4*	*ICU4*	Altered morphogenesis	Down-regulated	-1.12	6.49E-03	3.27E-02	32502	developmental process
At3g54720	*ALTERED MERISTEM PROGRAM1*	*AMP1*	Altered morphogenesis	Down-regulated	-1.07	8.92E-03	4.21E-02	7389	pattern specification process
At1g30490	*PHAVOLUTA*	*PHV*	Altered morphogenesis	Down-regulated	-1.07	2.27E-02	4.22E-02	7275	multicellular organismal development
At1g01030	*NGATHA3*	*NGA3*	Altered morphogenesis	Up-regulated	1.17	1.81E-05	1.11E-02	48367	shoot development
At1g70560	*TRYPTOPHAN AMINOTRANSFERASE OF ARABIDOPSIS1*	*TAA1*	Altered morphogenesis	Up-regulated	1.47	0.016683	4.72E-03	48825	cotyledon development
At3g14370	*KINASE PROTEIN SERINE/THREONINE KINASE ACTVITY*	*WAG2*	Altered morphogenesis	Up-regulated	1.09	0.004397	2.12E-02	48825	cotyledon development
At4g18390	*TEOSINTE BRANCHED, CYCLOIDEA, and PCF1*	*TCP2*	Altered morphogenesis	Up-regulated	1.25	3.78E-03	9.44E-03	9965	leaf morphogenesis
At1g53230	*TEOSINTE BRANCHED, CYCLOIDEA, and PCF2*	*TCP3*	Altered morphogenesis	Up-regulated	1.17	1.32E-03	1.38E-02	9965	leaf morphogenesis
AT2G02950	*PHYTOCHROME KINASE SUBSTRATE1*	*PKS1*	Altered morphogenesis	Up-regulated	1.18	1.15E-02	1.47E-02	9958	gravitropism
At1g66480	*PLASTID MOVEMENT IMPAIRED2*	*PMI2*	Altered morphogenesis	Up-regulated	1.82	6.44E-06	4.15E-04	9637	response to blue light
At2g29125	*ROTUNDIFOLIA-LIKE*	*RTFL2*	Altered morphogenesis	Up-regulated	1.44	2.53E-03	3.39E-03	48856	anatomical structure development
At1g13710	*CYTOCHROME P450 MONOOXYGENASE*	*CYP78A5*	Altered morphogenesis	Up-regulated	1.02	1.49E-02	3.86E-02	48366	leaf development
At2g26170	*MORE AXILLARY GROWTH1*	*MAX1*	Altered morphogenesis	Up-regulated	1.06	4.49E-03	2.60E-02	48366	leaf development
AT2G28110	*FRAGILE FIBER8*	*FRA8*	Cell wall modifications	Down-regulated	-2.40	1.43E-03	3.94E-04	9834	secondary cell wall biogenesis
At5g17420	*IRREGULAR XYLEM3*	*IRX3*	Cell wall modifications	Down-regulated	-1.67	1.07E-02	5.81E-03	30244	cellulose biosynthetic process
At5g54690	*IRREGULAR XYLEM8*	*IRX8*	Cell wall modifications	Down-regulated	-2.35	5.73E-03	4.49E-04	10417	glucuronoxylan biosynthetic process
At2g37090	*IRREGULAR XYLEM9*	*IRX9*	Cell wall modifications	Down-regulated	-1.18	1.42E-02	3.70E-02	9834	secondary cell wall biogenesis
At1g27440	*IRREGULAR XYLEM10*	*IRX10*	Cell wall modifications	Down-regulated	-1.42	2.92E-02	1.83E-02	9834	secondary cell wall biogenesis
At2g38080	*IRREGULAR XYLEM12*	*IRX12*	Cell wall modifications	Down-regulated	-2.39	5.84E-03	4.36E-04	9834	secondary cell wall biogenesis
At4g36890	*IRREGULAR XYLEM14*	*IRX14*	Cell wall modifications	Down-regulated	-1.65	1.15E-03	4.99E-03	9834	secondary cell wall biogenesis
At5g67230	*IRREGULAR XYLEM14-LIKE*	*IRX14-L*	Cell wall modifications	Down-regulated	-1.35	9.76E-04	1.55E-02	9834	secondary cell wall biogenesis
At3g18660	*PLANT GLYCOGENIN-LIKE STARCH INITIATION PROTEIN1*	*PGSIP1/ GUX1*	Cell wall modifications	Down-regulated	-2.38	1.97E-03	4.29E-04	45492	xylan biosynthetic process
At4g33330	*PLANT GLYCOGENIN-LIKE STARCH INITIATION PROTEIN3*	*PGSIP3/ GUX2*	Cell wall modifications	Down-regulated	-1.45	2.34E-03	1.09E-02	45492	xylan biosynthetic process
At1g19300	*GALACTURONOSYLTRANSFERASE-LIKE1*	*GATL1*	Cell wall modifications	Down-regulated	-2.11	1.32E-03	7.78E-04	45492	xylan biosynthetic process
At1g05170	*GALACTURONOSYLTRANSFERASE-LIKE*	*GATL-like*	Cell wall modifications	Down-regulated	-1.73	3.89E-04	2.63E-03	45492	xylan biosynthetic process

[a,b] P-value of chi-square test for the enrichment in genes with the indicated attribute (see chapter 2 section 2.6.3). [c] Gene ontology for biological processes annotated for that gene in *Arabidopsis thaliana*, TAIR version10 (Berardini et al., 2004).

Table 7.2. Candidate genes with altered gene expression in *Puccinia monoica*-induced pseudoflowers ('Pf') compared to uninfected *Boechera stricta* stems and leaves ('SL')

AGI code	Gene name	Common name	Annotation	Gene expression ('Pf' / 'SL')	Ratio Log2 ('Pf' / 'SL')	P-value[a]	FDR value[b]	GOBP[c]	GOPB Description
At2g46770	*NAC (NO APICAL MERISTEM) SECONDARY WALL THICKENING PROMOTING FACTOR 1*	*NST1*	Cell wall modifications	Down-regulated	-2.35	1.27E-03	4.27E-04	9834	secondary cell wall biogenesis
At1g32770	*NAC (NO APICAL MERISTEM) SECONDARY WALL THICKENING PROMOTING FACTOR3*	*NST3*	Cell wall modifications	Down-regulated	-2.60	3.11E-03	2.64E-04	9834	secondary cell wall biogenesis
At5g01360	*TRICHOME BIREFRINGENCE-LIKE3*	*TBL3*	Cell wall modifications	Down-regulated	-2.51	2.43E-03	3.68E-04	30244	cellulose biosynthetic process
At4g18780	*CELLULOSE SYNTHASE 8*	*CESA8*	Cell wall modifications	Down-regulated	-2.93	3.61E-03	5.00E-05	30244	cellulose biosynthetic process
At2g15090	*3-KETOACYL-COA SYNTHASE8*	*KCS8*	Cell surface modifications	Up-regulated	1.31	2.31E-03	7.20E-03	6633	fatty acid biosynthesis
At5g12420	*WAX ESTER SYNTHASE/ACYLCOA : DIACYLGLYCEROL ACETYLTRANSFERASE7*	*WSD7*	Cell surface modifications	Up-regulated	0.97	7.21E-03	4.51E-02	10025	wax biosynthesis
At5g23940	*CUTICULAR RIDGES*	*DCR*	Cell surface modifications	Up-regulated	0.94	8.27E-05	4.48E-02	6633	fatty acid biosynthesis
At1g51460	*ATP-BINDING-CASSETTE (ABC) TRANSPORTERS SUPERFAMILY G 13*	*ABCG13*	Cell surface modifications	Up-regulated	2.80	9.32E-03	0.00E+00	6869	lipid transport
At1g68130	*INDETERMINANT DOMAIN14*	*IDD14*	Regulation of flower development	Down-regulated	-1.18	2.90E-05	2.67E-02	45449	regulation of transcription
At5g20830	*SUCROSE SYNTHASE1*	*SUS1*	Regulation of flower development	Down-regulated	-1.56	9.03E-03	8.25E-03	16051	carbohydrate biosynthetic process
At3g43190	*SUCROSE SYNTHASE4*	*SUS4*	Regulation of flower development	Down-regulated	-2.32	1.11E-02	4.39E-04	16051	carbohydrate biosynthetic process
At1g65480	*FLOWERING LOCUS T*	*FT*	Regulation of flower development	Down-regulated	-1.28	2.24E-02	2.65E-02	3	reproduction
At3g07970	*QUARTER2*	*QRT2*	Regulation of flower development	Up-regulated	0.93	2.35E-03	4.14E-02	48869	cellular developmental process
At2g45190	*ABNORMAL FLORAL ORGANS1*	*AFO*	Repression of flower development and floral transition	Up-regulated	2.23	1.07E-04	9.33E-05	10158	abaxial cell fate specification
At4g08150	*KNOTTED-LIKE1*	*KNAT1*	Repression of flower development and floral transition	Down-regulated	-1.06	1.51E-05	4.54E-02	7275	multicellular organismal development
At2g27990	*POUND-FOOLISH*	*PNF*	Regulation of flower development	Down-regulated	-1.18	2.46E-03	2.93E-02	10076	maintenance of floral meristem identity
At2g03710	*SEPATALLA4*	*SEP4/ AGL3*	Regulation of flower development	Down-regulated	-1.40	8.02E-03	9.58E-03	48437	floral organ development
At5g42800	*DIHYDROFLAVONOL 4-REDUCTASE*	*DRF*	Pigment modifications	Down-regulated	-1.41	1.08E-02	8.18E-03	9718	anthocyanin biosynthetic process
At4g22880	*LEUCOANTHOCYANIDIN DIOXYGENASE*	*LDOX*	Pigment modifications	Down-regulated	-3.34	3.77E-03	0.00E+00	9718	biosynthesis
At1g21460	*SUGAR TRANSPORTER1*	*SWEET1*	Regulation of sugar metabolism	Up-regulated	1.50	1.31E-03	1.99E-03	34219	carbohydrate transmembrane transport

[a,b] P-value of chi-square test for the enrichment in genes with the indicated attribute (see chapter 2 section 2.6.3). [c] Gene ontology for biological processes annotated for that gene in *Arabidopsis thaliana*, TAIR version10 (Berardini et al., 2004).

Table 7.2. Candidate genes with altered gene expression in *Puccinia monoica*-induced pseudoflowers ('Pf') compared to uninfected *Boechera stricta* stems and leaves ('SL')

AGI code	Gene name	Common name	Annotation	Gene expression ('Pf' / 'SL')	Ratio Log2 ('Pf' / 'SL')	P-value[a]	FDR value[b]	GOBP[c]	GOPB Description
At5g13170	*SUGAR TRANSPORTER15*	*SWEET15*	Regulation of sugar metabolism	Up-regulated	1.38	1.68E-02	5.09E-03	34219	carbohydrate transmembrane transport
At3g13790	*CELL WALL INVERTASE1*	*cwINV1*	Regulation of sugar metabolism	Up-regulated	2.44	1.99E-02	4.29E-05	6950	response to stress
At2g24210	*TERPENE SYNTHASE10*	*TPS10*	Changes in volatiles synthesis	Down-regulated	-2.22	2.18E-02	7.44E-04	16099	monoterpenoid biosynthethic process
At5g23960	*TERPENE SYNTHASE21*	*TPS21*	Changes in volatiles synthesis	Down-regulated	-2.65	6.46E-04	1.90E-04	16099	monoterpenoid biosynthethic process
At4g23590	*TYROSINE TRANSAMINASE*	-	Changes in volatiles synthesis	Up-regulated	2.50	4.37E-03	1.82E-05	6519, 6558	cellular amino acid and derivative metabolic process, L-phenylalanine metabolic process
At4g37390	*IAA AMINO ACID SYNTHASE, AUXIN-RESPONSIVE GH3 FAMILY PROTEIN*	*GH3.2*	Regulation of plant hormones	Up-regulated	4.40	1.86E-04	0.00E+00	9725	response to hormone stimulus
At1g59500	*IAA AMINO ACID SYNTHASE, AUXIN-RESPONSIVE GH3 FAMILY PROTEIN*	*GH3.4*	Regulation of plant hormones	Up-regulated	2.64	1.69E-04	2.86E-05	9725	response to hormone stimulus
At2g38870	*SERINE PROTEASE INHIBITOR*	-	Delayed senescence	Up-regulated	1.09	0.024873	2.77E-02	10951	negative regulator of endopeptidase activity
At5g50260	*CYSTEINE PROTEINASE*	-	Delayed senescence	Down-regulated	-4.30	0.002777	0.00E+00	4197	cysteine-type endopeptidase activity
At3g61830	*AUXIN RESPONSE FACTOR18*	*ARF18*	Activation of defense responses	Up-regulated	1.16	6.18E-04	1.23E-02	9725	response to hormone stimulus
At2g29490	*GLUTATHIONE S-TRANSFERASE TAU1*	*ATGSTU1*	Activation of defense responses	Up-regulated	1.46	1.46E-02	5.21E-03	8152	toxin catabolic process
At2g29480	*GLUTATHIONE S-TRANSFERASE TAU2*	*ATGSTU2*	Activation of defense responses	Up-regulated	0.99	1.55E-02	4.36E-02	8152	toxin catabolic process
At2g29460	*GLUTATHIONE S-TRANSFERASE TAU4*	*ATGSTU4*	Activation of defense responses	Up-regulated	1.10	1.10E-02	2.47E-02	8152	toxin catabolic process
At1g10370	*GLUTATHIONE S-TRANSFERASE TAU17*	*ATGSTU17*	Activation of defense responses	Up-regulated	1.04	9.66E-04	2.46E-02	8152	toxin catabolic process
At1g17190	*GLUTATHIONE S-TRANSFERASE TAU26*	*ATGSTU26*	Activation of defense responses	Up-regulated	0.99	8.01E-03	3.89E-02	8152	toxin catabolic process
At5g13330	*RELATED TO AP2 6L*	*RAP2.6L*	Activation of defense responses	Up-regulated	1.49	4.82E-03	3.31E-03	9607	response to biotic stimulus
At5g23750	*REMORIN FAMILY PROTEIN*	*Remorin*	Activation of defense responses	Up-regulated	1.31	1.02E-03	5.30E-03	9607	response to biotic stimulus
At3g17520	*LATE EMBRYOGENESIS ABUNDANT PROTEIN4*	*LEA4*	Activation of defense responses	Up-regulated	1.93	1.28E-02	3.21E-04	9414	response to water deprivation
At2g46150	*LATE EMBRYOGENESIS ABUNDANT FAMILY PROTEIN*	*LEA family*	Activation of defense responses	Up-regulated	1.71	1.21E-03	8.87E-04	9414	response to water deprivation
At1g65690	*LATE EMBRYOGENESIS ABUNDANT FAMILY PROTEIN*	*LEA family*	Activation of defense responses	Up-regulated	1.41	7.00E-03	5.24E-03	9414	response to water deprivation

[a,b] P-value of chi-square test for the enrichment in genes with the indicated attribute (see chapter 2 section 2.6.3). [c] Gene ontology for biological processes annotated for that gene in *Arabidopsis thaliana*, TAIR version10 (Berardini et al., 2004).

7.2.3. Description of candidate genes showing altered gene expression in *P. monoica*-induced pseudoflowers ('Pf') compared to uninfected *Boechera stricta* stem and leaves ('SL')

7.2.3.1. Altered morphogenesis in pseudoflowers

Pseudoflowers are modified leaves of different shape and size compared to the uninfected host leaves (Fig 7.1). I investigated the presence of significant regulated genes that could participate in the altered morphology of the host plant leaves and identified two *P-GLYCOPROTEINS* genes up-regulated in pseudoflowers. *PGP1* (*At4g25960*) and *PGP9* (*At4g18050*) were up-regulated in pseudoflowers (Fig. 7.4A and Table 7.2). Plant phosphoglycoproteins (PGPs) are B-type ATP binding cassette (ABCB) transporters that function in auxin transport and also in a photropin-regulated pathway (Blakeslee et al., 2007; Blakeslee et al., 2004). ATP binding cassette (ABC) transporters play critical roles in plant growth and development associated with their auxin transport activities (Geisler et al., 2005; Geisler and Murphy, 2006; Sidler et al., 1998). *P-GLYCOPROTEIN1 (PGP1)* gene functions in hypocotyl cell elongation in the light (Sidler et al., 1998). The up-regulation of *PGP1* and *PGP9* genes, observed only in pseudoflowers, suggests a possible function in stem elongation during the formation of pseudoflowers (Roy, 1993a).

Several genes involved in shoot development were differentially regulated in pseudoflowers (Fig. 7.4). The *INCURVATA4 (ICU4, At1g52150)* and *PHAVOLUTA (PHV) (At1g30490)* genes were down-regulated in pseudoflowers, whereas the *NGATHA3* (*NGA3) (At1g01030)* gene was upregulated (Fig. 7.4A and Table 7.2). *ICU4* encodes a HD-ZIP III transcriptional factor ATHB15, required for shoot apical meristem pattering and stem vascular differentiation (Ochando et al., 2006). Impaired shoot apical meristem is inferred from abnormal arrangement of leaves with paired leaves born in the stem and axillary shoots in *icu4* mutants (Ochando et al., 2006). In addition *icu4* mutants show enlarged metaxylem tracheids (unopened ends in the xylem), extra layers of procambial cells (cells in the xylem that retain their meristematic activity) and reduction in the

number of vascular bundles as well as poor lignification of the interfascicular fibers indicating a role for *ICU4* in shoot vascular bundles pattering (Ochando et al., 2006). *PHV, PHABULOSA (PHB)* and *REVOLUTA (RV)* are HD-ZIP family proteins involved in radial pattering in the leaf primordium (Emery et al., 2003; McConnell et al., 2001). *phb-6 phv-5 rev-9* mutant plants show leaf-like organs and failed to form primary apical meristem (Emery et al., 2003; McConnell et al., 2001). Finally, *NGA3* is part of a small B3 DNA binding domain protein family widely expressed in roots, stem, leaves and inflorescence tissues in *Arabidopsis thaliana* (Alvarez et al., 2006; Schmid et al., 2005; Trigueros et al., 2009). Overexpression of *NGA3* in transgenic plants resulted in apical dominance and altered flower phyllotaxy with abnormal arrangement of leaves in the axis of the stem, abnormal leaf morphology with longer, narrow and darker color rosette leaves, and flattened stem (Trigueros et al., 2009). In summary, down-regulation of transcriptional regulators of the development of the leaf *ICU4* and *NGA3* could contribute to the altered morphology of leaves and stems in pseudoflowers (Fig. 7.1B) (Roy, 1993a).

Pseudoflowers consist of clusters of elongated stems that bolt from the infected rosettes and that almost never reach flowering. Regulation of host hormones involved in plant organogenesis could participate in the formation of these dense flower-like clusters. *PROTEIN SERINE/THREONINE AGC3 KINASE (WAG2) (At3g14370)* and *TRYPTOPHAN AMINOTRANSFERASE OF ARABIDOPSIS1 (TAA1) (At1g70560)*, which play important roles in auxin-transport and auxin-dependent developmental processes in the cotyledons, were up-regulated in pseudoflowers (Cheng et al., 2008; Dhonukshe et al., 2010; Stepanova et al., 2008) (Fig. 7.4A and Table 7.2) (Cheng et al., 2008; Stepanova et al., 2008). Up-regulation of *WAG2* and *TAA1* genes could participate in the redirecting of auxin to the apical tissues to promote bolting and growth of leaf organs in the infected rosettes (Fig. 7.1B).

In contrast to the uninfected *B. stricta* plants, stems from infected *P. monoica* pseudoflowers do not exhibit primary or secondary cauline leaf branching (leaves growing on stems) (Fig. 7.1) (Roy, 1993a). I found that the *TEOSINTE BRANCHED1, CYCLOIDEA,* and *PCF (TCP) TCP2 (At4g18390)* and *TCP3*

(At1g53230) genes that act as suppressors of shoot lateral organ morphogenesis were up-regulated in pseudoflowers (Fig. 7.4A and Table 7.2) (Aida et al., 1997; Cubas et al., 1999; Koyama et al., 2007; Koyama et al., 2010). The up-regulation of *TCP2* and *TCP3* genes could participate in the maintenance of apical dominance in each stem of the flower-like cluster by suppressing lateral shoot development, as there is no branching observed in shoots bearing pseudoflowers (Fig. 7.1B). *MORE AXILLARY GROWTH1 (MAX1/ CYP711A1) (Atg2g26170)*, a gene that represses vegetative axillary branching by controlling auxin transport in *A. thaliana* was also up-regulated in pseudoflowers (Fig. 7.4A and Table 7.2) (Bennett et al., 2006; Lazar and Goodman, 2006). *PHYTOCHROME KINASE SUBSTRATE1 (PKS1)* gene *(At2g02950)* encoding a cytoplasmic protein that interacts with the phytochrome phytA and the most abundant red light photochrome phyB was up-regulated in pseudoflowers (Fankhauser et al., 1999; Lariguet et al., 2006; Neff et al., 2000) (Fig. 7.4A and Table 7.2). Previous reports show that *PKS1* overexpressing plants exhibit longer hypocotyls in red light due to negative regulation of phyB (Clack et al., 1994; Fankhauser et al., 1999; Whitelam et al., 1998). Up-regulation of *TCP2*, *TCP3*, *MAX1* and *PKS1* could act as signals in the inactivation of secondary shoot meristems in the infected plants, which is consistent with the observed absence of shoot branching in pseudoflowers (Fig. 7.1B).

It is possible to speculate that in order to produce pseudoflowers *P. monoica* must first induce the dedifferentiation of host leaf cells followed by a reprogramming that changes leaf morphology and development. The *TCP2* and *TCP3* genes that are up-regulated in pseudoflowers also participate in the maintenance of undifferentiated fates in the shoot apical meristem (SAM) and in the production of differentiated cells in leaves (Palatnik et al., 2003) (Fig. 7.4A and Table 7.2). In addition, in pseudoflowers I found down-regulation of *ALTERED MERISTEM PROGRAM 1 (AMP1)* gene *(At3g54720*) that in contrast to *TCP2* and *TCP3* gene promotes cell differentiation (Conway and Poethig, 1997; Vidaurre et al., 2007) (Fig. 7.4A and Table 7.2). Moreover, mutations in *AMP1* increases cotyledon number, rate of leaf initiation, produces a general reduction in the size of leaves, inflorescence stems, floral organs and cause apical dominance (Conway and Poethig, 1997). Therefore, both up-regulation of

TCPs and down-regulation of *AMP1* could function in the dedifferentiation process of leaf cells in infected leaf cells (Fig. 7.1B).

Pseudoflowers exhibit modified leaves in size and shape compared to uninfected *B. stricta* leaves (Fig. 7.1). I found that in pseudoflowers, the *ROTUNDIFOLIA like2* gene (*RTFL2*) gene *(At2g29125)* that regulates the number of cells in the leaf organs in *A. thaliana* was up-regulated (Fig. 7.4A and Table 7.2) (Narita et al., 2004; Wen et al., 2004). Another pseudoflower up-regulated gene in pseudoflowers that increases numbers of cells causing overgrowth of various plant organs, including leaves, is the cytochrome P450 monooxygenase *CYP78A5/KLU* gene *(At1g13710)* (Fig. 7.4A and Table 7.2) (Eriksson et al., 2010; Zondlo and Irish, 1999). *CYP78A5/KLU* is important in the coordination of growth of individual flowers, and flowers within the inflorescence contributing to uniformity of size and symmetry, which is an important determinant of a plant's attractiveness to pollinators (Anastasiou et al., 2007; Eriksson et al., 2010; Moller, 1995). Up-regulation of *RTFL2* and *CYP78A5/KLU* genes could contribute to the development of leaf organ primordia, particularly leaf cell size in pseudoflowers (Fig. 7.4A). In addition, *CYP78A5* gene could also contribute to the symmetry of the flower-like leaf clusters, and in that way ensures their attractiveness to visiting insects (Fig. 7.1B).

7.2.3.2. Cell wall modifications

Puccinia monoica pseudoflowers have thinner stems compared to uninfected *B. stricta* plants. This suggests alteration in cell wall composition of the stem cells of infected plants (Fig. 7.1). Plant cell wall, in particularly, secondary cell walls are constituted in majority by glucuronoxylan (GX), along with cellulose and lignin. In Arabidopsis, several Glycosyltransferases (GTs) are involved in GX biosynthesis: *FRAGILE FIBER*8 *(FRA8), IRREGULAR XYLEM3 (IRX3), IRREGULAR XYLEM8 (IRX8), IRREGULAR XYLEM9 (IRX9), IRREGULAR XYLEM10 (IRX10), IRREGULAR XYLEM12 (IRX12), IRREGULAR XYLEM14 (IRX14)* and *IRREGULAR XYLEM14-LIKE (IRX14-L)* (Brown et al., 2009; Keppler and Showalter, 2010; Lee et al., 2010; Wu et al., 2010; Wu et al., 2009; Zhong et al.,

2005). Mutations in the *IRX3* gene cause dramatic reduction in cellulose content (Brown et al., 2005). Arabidopsis *fra8, irx8* and *irx9* and *irx10* mutants show reduction in xylose, a main component of xylan, with decreased fiber wall cell thickness and stem strength (Brown et al., 2005; Pena et al., 2007; Wu et al., 2009; Zhong et al., 2005). Plants homozygous for *irx14* and heterozygous for *irx14-L* mutations exhibit smaller leaves, stems, and siliques that rarely contain any viable seeds (Keppler and Showalter, 2010). *IRX12* gene is proposed to be involved in lignin but not cellulose or xylan synthesis due to minor changes in the sugar composition and cellulose content observed in the mutants (Brown et al., 2005). I found that *FRA8 (At2g28110)*, *IRX3 (At5g17420), IRX8 (At5g54690), IRX9 (At2g37090), IRX10* (At1g27440)*, IRX12* (*AT2g38080*), *IRX14 (At4g36890)* and *IRX14-L (At5g67230)* genes were down-regulated in pseudoflowers (Fig. 7.4A and Table 7.2). The down-regulation of *the FRA8*, *IRX3, IRX8, IRX9, IRX10, IRX12*, *IRX14* and *IRX14-L* genes in pseudoflowers might be involved in the alteration of the cell wall structure resulting in the decreased stem strength and smaller leaves observed in pseudoflowers compared to uninfected plants (Fig. 7.1). Two other genes involved GX biosynthesis that are associated with secondary wall thickening in fibers and vessels that were down-regulated in pseudoflowers: *GALACTURONOSYLTRANSFERASE-LIKE 1 (GATL1) (At1g19300)* and *GALACTOSYLTRANSFERASE-LIKE (GATL-like) (At1g05170) GATL1* (Fig. 7.4A and Table 7.2) (Brown et al., 2007; Kong et al., 2011; Lee et al., 2007). I also found down-regulation in pseudoflowers of two other gene members of the GT family, that participate in GX biosynthesis are: the *PLANT GLYCOGENIN-LIKE STARCH INITIATION PROTEIN1/GLUCURONIC ACID SUBSTITUTION OF XYLAN1 (PGSIP1/GUX1) (At3g18660)* and the *PLANT GLYCOGENIN-LIKE STARCH INITIATION PROTEIN3/GLUCURONIC ACID SUBSTITUTION OF XYLAN2 (PGSIP3/GUX2) (At4g33330)* (Fig. 7.4A and Table 7.2). Double mutant plants *PGSIP1/GUX1* and *PGSIP3/GUX2* show highly decreased content of glucuronic acid in secondary cell walls and substantially reduced xylan glucuronosyltransferase activity (Mortimer et al., 2010; Oikawa et al., 2010). Also the stems of these mutants are weakened, but the xylem vessels are not collapsed. Interestingly, the xylan of these plants is composed of a single monosaccharide that requires fewer enzymes for hydrolysis (Mortimer et al., 2010). The down-regulation of *GALT1, GATL1-like, PGSIP1 and PGSIP3* genes

could also be involved in stem weakening in pseudoflowers as described above for other genes that form part of the GX biosynthesis pathway. In addition, it is possible that modification of the composition of the cell wall to a single monosaccharide by *PGSIP1 and PGSIP3* genes facilitates transport and acquisition of nutrients from the plant to the fungus.

Other genes involved in normal stem development were down-regulated in pseudoflowers. *NAC SECONDARY WALL THICKENING PROMOTING FACTOR1 (NST1) (At2g46770)* and *NAC SECONDARY WALL THICKENING PROMOTING FACTOR3 (NST3) (At1g32770)* that act as suppressors of secondary wall thickenings in between vascular bundles of inflorescence stems in *A. thaliana* were down-regulated in pseudoflowers (Fig. 7.4A and Table 7.2) (Mitsuda et al., 2007). Also, six *TRICHOME BIREFRINGENCE (TBR)* homologs genes of *TBL3 (At5g01360): TBL19 (At5g15900), TBL28 (At2g40150), TBL29 (At3g55990), TBL31 (At1g73140), TBL33 (At2g40320)* and *TBL40 (At2g31110)* were down-regulated in pseudoflowers (Fig. 7.4A and Table 7.2). Mutations in the *TBL3* gene cause reduction in the stem diameter in *A. thaliana* (Bischoff et al., 2010). In addition, the *CELLULOSE SYNTHASE 8* (*CESA8*) gene *(At4g18780*) that is known to be strongly coexpressed with the homolog *TBL3* was also down-regulated in pseudoflowers (Fig. 7.4A and Table 7.2) (Bischoff et al., 2010). Down-regulation of *NST1*, *NST3* and the *TBL3* homologs could be involved in the reduction in thickening of the stems in pseudoflowers (Fig. 7.1B).

7.2.3.3. Cell surface modifications

Pseudoflowers are composed of flower-like leaves with a glossy aspect due to the secretion of cuticular waxes (Fig. 7.1B). They mimic the unrelated yellow and glossy true flowers of *Ranunculus* species (PARKIN, 1935; Roy, 1993a). A homolog of the *WAX ESTER SYNTHASE/ACYLCOA: DIACYLGLYCEROL ACETYLTRANSFERASE1 (WSD1), WDS7* gene *(At5g12420)* that encodes a wax synthase required for stem wax ester biosynthesis in *A. thaliana* stem was up-regulated in pseudoflowers (Fig. 7.4A and Table 7.2) (Kalscheuer and Steinbuchel, 2003; Li et al., 2008a). *WSD1* gene expression is mainly detected in

flowers, top part of stems and leaves, which is consistent with its role in cuticular wax production. Another gene involved in the synthesis of cuticular waxes, particularly in the first step of fatty acid elongation, the 3-*KETOACYL-COA SYNTHASE8 (KCS8)* gene *(At2g15090)* was up-regulated in pseudoflowers (Fig. 7.4A and Table 7.2) (Joubes et al., 2008). A third gene involved in cuticular waxes that was up-regulated in pseudoflowers is *CUTICULAR RIDGES (DCR)* previously known as *PERMEABLE LEAVES3 (PEL3)* gene *(At5g23940)* encoding a putative acyltransferase of the *A. thaliana* BAHD family required for the incorporation of the most abundant flower cutin monomer (Fig 7.4A and Table 7.2) (Marks et al., 2009; Panikashvili et al., 2009; Tanaka et al., 2004). The expression of *DCR* gene is not restricted to inflorescence; it is also present in young emerging leaves and the elongating part of stems that suggests an additional role for cutin polymer formation in vegetative organs (Panikashvili et al., 2009). Up-regulation of *DCR* genes in pseudoflowers could be involved in the production of cuticular waxes, which is consistent with the glossy phenotype of infected flower-like leaves (Fig. 7.1B). In addition, a gene involved in transport of wax in *A. thaliana,* the *ATP-BINDING-CASSETTE (ABC) TRANSPORTERS SUPERFAMILY G GENE ABCG13 (At1g51460)* was up-regulated in pseudoflowers (Fig. 7.4A and Table 7.2) (Bird et al., 2007; Luo et al., 2007; Panikashvili et al., 2007; Panikashvili et al., 2011; Pighin et al., 2004). The reported expression in leaves of *KCS8* and *DCR* is consistent with the expression found in flower-like leaves of infected plants. However, *ABCG13 transporter* is only known to be expressed in true flowers (Panikashvili et al., 2011). This observation indicates that wax transporters from true flowers can be present in pseudoflowers, and that *ABCG13* might also function in the production of cuticular wax in pseudoflowers. In summary, the up-regulation in pseudoflowers of *KCS8, DCR* and *WSD7* involved in wax biosynthesis and *ABCG13* involved in wax transport suggest changes in wax production and allocation in pseudoflowers and perhaps beneficial roles in shininess to attract pollinators. Altered wax production could also result in better adhesion of rust spores during infection and subsequent fertilization. Waxy cuticle compounds are known to facilitate germination of fungal spores, which require a highly hydrophobic surface for adhesion and the amount of these compounds determines rust fungus infection (Staples et al., 1985).

7.2.3.4. Regulation of flower organ development

Puccinia monoica causes flower-like leaves or pseudoflowers to form on systemically infected *Boechera stricta* host plant (Roy, 1993a). Therefore, it is possible that *P. monoica* induced the inhibition of floral signals and floral organ development. I found several genes that are known to be involved in the floral transition to be down-regulated in pseudoflowers. The mobile floral activator signal protein produced by the *FLOWERING LOCUS T (FT) (At1g65480)* was down-regulated in pseudoflowers (Fig. 7.4A and Table 7.2) (Corbesier et al., 2007). FT signal moves from the induced leaf through the phloem to the shoot apex where it interacts with FLOWERING LOCUS D (FD) bZIP transcription factor to initiate transcription of floral specification genes (Abe et al., 2005; Corbesier et al., 2007; Giakountis and Coupland, 2008). Down-regulated of FT in pseudoflowers suggests interference with activation and transmission of the floral signal that might have contributed to the inhibition of floral organs in the infected plants (Fig. 7.1B). Signals to initiate flowering are also associated with sugar contents in *A. thaliana* (Eimert et al., 1995). *INDETERMINATE DOMAIN* transcription factor14 (*IDD14) gene (At1g68130),* a homolog of *IDD8* gene that regulates photoperiodic flowering by modulating sugar transport and metabolism, was down-regulated in pseudoflowers (Fig. 7.4A and Table 7.2) (Seo et al., 2011). In addition, the *SUCROSE SYNTHASE1 (SUS1) (AT3g43190)* and the *SUCROSE SYNTHASE4 (SUS4) (At5g20830)* genes that are co-regulated by *IDD8* were down-regulated in pseudoflowers (Fig. 7.4A and Table 7.2) (Seo et al., 2011). Down-regulation of *IDD14, SUS1* and *SUS4* suggests manipulation of host sugar metabolism by the rust pathogen to prevent floral transition in infected plants.

Five other genes involved in the development of floral organs were differentially expressed in pseudoflowers. 1) *QUARTER2 (QRT2) (AT3g07970)*, a gene that encodes a polygalacturonase (PG) involved in cell division was up-regulated in pseudoflowers (Ogawa et al., 2009) (Fig. 7.4A and Table 7.2). Plant overexpressing *QRT2* have flowers that do not open, atypical petals, and anthers that fail to dehisce (release the organ content) normally. 2) *ABNORMAL FLORAL ORGANS 1 (AFO) (At2g45190)* that encodes a member of the YABBY family of

transcriptional regulators required for normal flower development in *A. thaliana* was up-regulated in pseudoflowers (Kumaran et al., 1999) (Fig. 7.4A and Table 7.2). *afo* mutant flowers have defects in all four floral whorls that are evident from an early stage (Kumaran et al., 1999). 3) *KNAT1/BREVIPEDICELLUS (BP) KNAT1/BP (At4g08150), a* transcriptional regulator member of the *CLASS1 KNOTTED1-LIKE HOMEOBOX (KNOX)* family was down-regulated in pseudoflowers (Fig. 7.4A and Table 7.2) (Scofield et al., 2008). Loss of *KNAT1/BP* results in reduced growth of floral pedicels, internodes and the style during reproductive growth (Douglas et al., 2002; Scofield et al., 2008; Venglat et al., 2002). 4) *POUND-FOOLISH (PNF)* gene *(At2g27990),* a paralog *of BEL1-like homeobox gene (BLH) of PENNYWISE (PNY) (At5g02030)* that control inflorescence patterning events including floral specification and internode pattering was down-regulated in pseudoflowers (Kanrar et al., 2008; Kanrar et al., 2006; Smith et al., 2004; Smith and Hake, 2003) (Fig. 7.4A and Table 7.2). Arabidopsis *pny pnf* double mutants initiate compact shoots that fail to response to flowering inductive signals and to form flowers (Smith et al., 2004). 5) The MADS-BOX *SEPATALLA 4/AGAMOUS-LIKE 3 (SEP4/ AGL3) (At2g03710)* that play central roles in flower meristem and flower organ identity was down-regulated in pseudoflowers (Fig. 7.4A and Table 7.2) (Ditta et al., 2004; Huang et al., 1995). *sep4* single mutants do no exhibit visible phenotypes, but mutations in the four members of the *SEP* gene family *sep1 sep2 sep3 sep4* show a conversion of floral organs to leaf-like organs, which suggest *SEP4* gene is probably functionally redundant in *A. thaliana* (Ditta et al., 2004). However, putative redundancy of *SEP4* is questionable in other plant species where homologous SEP4 proteins show differences in protein-protein interactions (de Folter et al., 2005; Immink et al., 2003; Malcomber and Kellogg, 2005). These results suggest that regulation of several genes involved in sepals, anthers and other parts of the floral organs (up-regulation of *QRT2* and down-regulation of *AFO, KNAT1, PNF and SEP4*) may coordinate inhibition of flower formation in the infected plants (Fig. 7.1B). Prevention of flowering is very successful in infected host plants and obviously negatively impacts plant fitness, which indicates that *P. monoica* greatly affects host populations (Roy, 1993a).

7.2.3.5. Pigment modifications

Most of the plant pigments ranging from red to purple colors are anthocyanins, a group of flavonoids that are crucial for flower coloration, attracting insects for pollination and seed dispersal (Clegg and Durbin, 2000). I found that two genes encoding enzymes participating in the biosynthesis of anthocyanin were down-regulated in pseudoflowers. Dihydroflavonol 4-reductase (DFR) enzyme (At5g42800) reduces dihydroflavonol to leucocyanidin, and leucoanthocyanidin dioxygenase (LDOX/TDSA) enzyme (At4g22880) uses leucocyanidin to produce anthocyanin (Abrahams et al., 2003; Shirley et al., 1995) (Fig. 7.4A, Table 7.2 and Fig. 7.5). In addition, I found that *S-like ribonucleases (RNases) (RSN1)* gene *(At2g02990)* member of the widespread ribonuclease T_2 family known to inhibit the production of anthocyanin was up-regulated in pseudoflowers (Bariola et al., 1999). Altogether, these results suggest that production of anthocyanin may be shutdown in pseudoflowers.

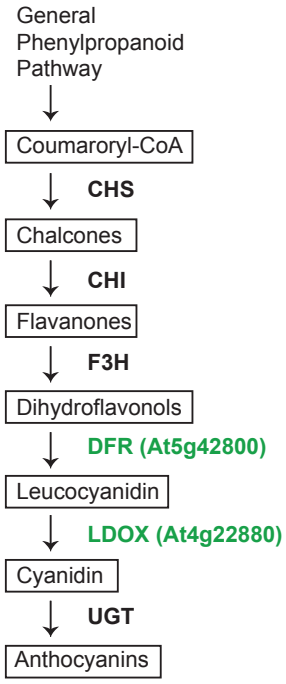


Fig. 7.5. Scheme showing down-regulated genes in pseudoflowers involved in the flavonoid pathway leading to synthesis of anthocyanins
The enzymes involved in the pathway are shown as follows: CHS, chalcone synthase; CHI, chalcone isomerase; F3'H, flavonoid-3'-hydroxylase; F3'5'H, flavonoid-3', 5'-hydroxylase; F3H, flavanone-3b-hydroxylase; DFR, dihydroflavonol-4-reductase; LDOX, leucoanthocyanidin dioxygenase and UFGT, UDP-Glc:flavonoid-3-O-glucosyltransferase. Individual enzymes labeled in green indicate those that are encode by genes down-regulated in pseudoflowers within the flavonoid pathway.

7.2.3.6. Regulation of sugar metabolism

Puccinia monoica like many other biotrophic pathogens is thought to acquire nutrients from the host plant to ensure colonization and reproduction (Divon and Fluhr, 2007). Previous studies show that sugar transfer occurred from plant leaves to powdery mildews (Aked and Hall, 1993; Sutton et al., 2007; Sutton et al., 1999). I found that two plant SWEET sugar transporters *AtSWEET1 (At1g21460)* and *AtSWEET15 (At5g13170)* which are exploited by pathogens for acquiring sugars from the host were up-regulated in pseudoflowers (Fig. 7.4A and Table 7.2). The up-regulation of *AtSWEET1* and *AtSWEET15* in pseudoflowers, suggest that sugar transporters might be co-opted during infection by *P. monoica* for nutritional gain.

Pseudoflowers also mimic flowers by producing sugar nectar that helps to attract flower-visiting insects (Roy, 1993a, b, 1994). *CELL WALL INVERTASE1 (CWINV1),* a homologous gene of (CWINV4) with a putative conserved role in nectar secretion within the Brassicaceae was up-regulated in pseudoflowers (Fig. 7.4A and Table 7.2) (Kram and Carter, 2009; Kram et al., 2009; Ruhlmann et al., 2010). *AtcwINV4* is preferentially expressed in flowers, unlike *AtcwINV1* that is highly expressed in both flowers and leaves (Sherson et al., 2003). Up-regulation of *AtcwINV1* gene in pseudoflowers could contribute to the production of sugars in leaves of infected plants. Sugar accumulation over the pseudoflower surface should benefit the rust pathogen by prolonging insect visits and increasing the likelihood of fungus fertilization (Roy, 1993a).

7.2.3.7. Changes in volatiles synthesis

Terpenes are the largest and most diverse class of specialized metabolites of volatile organic compounds (VOC) emitted by plants. *Arabidopsis thaliana* flowers emits a mixture of volatiles dominated by monoterpenes and sesquiterpenes (Chen et al., 2003). I found down-regulation in pseudoflowers of two terpene synthases genes that are involved in the biosynthesis of terpenes in Arabidopsis (Fig. 7.4A, Table 7.2 and Fig. 7.7). 1) *Terpene synthase 10* (*TPS10) (At2g24210)* is expressed in flowers and leaves and mediates the production of β-myrcene and (E)-β-ocimene (Bohlmann et al., 2000; Chen et al., 2003). 2) *Terpene synthase 21 (TPS21) (At5g23960)* is expressed almost exclusively in flowers and capable of producing five sesquiterpenes [(—)-(E)-β-caryophyllene and α-humulene in major amounts and (—)-α-copaene and β-elemene in lower amounts] (Chen et al., 2003). In addition, I found up-regulation of *TYROSINE TRANSAMINASE* gene *(At4g23590)*, which participates in the phenylalanine degradation pathway and the production of the volatile compounds phenylacetaldehyde and phenylethyl ethanol (Fig 7.5A and Table 7.2). This finding is consistent with a previous study that showed that phenylacetaldehyde and phenylethyl ethanol are the most dominant volatiles in various *Puccinia*-induced pseudoflowers (Raguso and Roy, 1998). In contrast, terpenes are not detectable in pseudoflowers (Raguso and Roy, 1998). Moreover,

phenylacetaldehyde and phenylethyl ethanol were suggested to play roles in favouring reproduction and protecting flowers by attracting different sets of pollinating and predatory insects, respectively (Raguso et al., 2003; Zhu et al., 2005). I hypothesize that these compounds (phenylacetaldehyde and phenylethyl ethanol) give a distinct fragrance to the pseudoflowers and help to the sexual reproduction of the rust fungus (Roy, 1993a, 1996; Roy and Raguso, 1997).

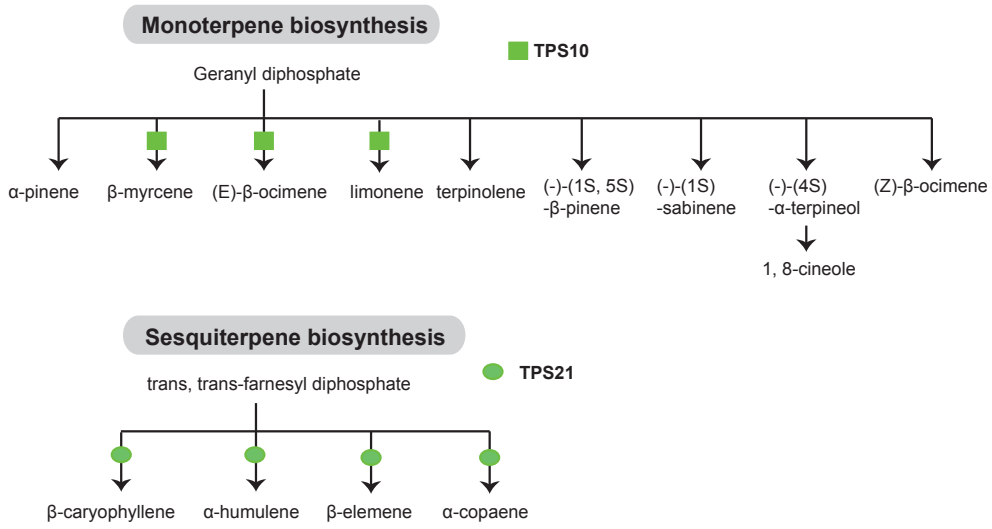


Fig. 7.6. Simplified scheme showing down-regulated genes in pseudoflowers involved in the terpenoid biosynthetic pathway
Individual enzymes labeled in green indicate those that are encoded by genes down-regulated in pseudoflowers within the terpenoid (monoterpenes and sesquiterpenes) biosynthetic pathway.

7.2.3.8. Regulation of hormones

The phytohormone indole-3-acetic acid (IAA) commonly known as auxin is a key regulator of cell expansion and division. IAA induces the production of expansins and cell wall-loosening proteins and makes plants vulnerable to pathogens. GH3 genes encode IAA-amido synthetase enzymes that help to maintain auxin homeostasis by conjugating excess IAA to amino acids (Staswick et al., 2005). Previous studies show that the GH3-2 gene confers broad-spectrum resistance to plants against bacterial and fungal pathogens by suppressing pathogen induced IAA accumulation (Fu et al., 2011). GH3-mediated auxin homeostasis

activates the reallocation of plant metabolic resources to facilitate resistance, which is linked to growth regulation. Therefore, GH3-mediated growth suppression is considered a fitness cost of the induced resistance (Park et al., 2007). I detected up-regulation of two IAA-amido synthetase genes in pseudoflowers: *GH3.2 (At4g37390)* and *GH3.4 (At1g59500),* known to be involved in the production of IAA conjugates to regulate the level of active auxin inside the plant (Fig. 7.4A and Table 7.2). In addition, I found up-regulation of an *AUXIN RESPONSE FACTOR18* gene *(ARF18) (At3g61830)* that belongs to a family of transcription factors that regulate the expression of auxin responsive genes (Fig. 7.4A and Table 7.2). Plants optimize their growth in response abiotic or biotic stimulus. I propose that upon pathogen infection, the induced GH3 genes mediate not only auxin homeostasis but also growth suppression in pseudoflowers. This interpretation is correlated with the observed reduced overall growth in pseudoflowers compared to flowering uninfected plants (Fig. 7.1B).

7.2.3.9. Delayed leaf senescence

Pseudoflowers are formed of flower-like leaves that are covered by sugary fluid containing nectar with spermatia (spores). Because this fluid is spread over the whole infected plant and not concentrated in a nectary, pseudoflowers have longer period of pollinator visits compared to uninfected co-occurring flowering plants (Roy, 1993a). If senescence occurs in uninfected host flowers but it is delayed in flower-like leaves of infected plants, it is possible to assume that pseudoflowers could gain more number of pollinators and also maintain the extended visits. Senescence is the process that leads to death of a particular organ or whole plant and involve a variety of proteases, among which cysteine proteases are the most common proteolytic enzymes (Gepstein, 2004; Morris et al., 1996). I found down-regulation in pseudoflowers of a gene (*At5g50260*) encoding for a cysteine proteinase with putative endopeptidase activity (Fig. 7.4A and Table 7.2). In addition, protease inhibitors are thought to delay visible symptoms of senescence in plants (Pak and van Doorn, 2005). I found up-regulation in pseudoflowers of a gene *(At2g38870)* encoding a serine-protease inhibitor annotated as a negative regulator of endopeptidase activity. Down-

regulation of cysteine proteinase and up-regulation of the serine-protease inhibitor in pseudoflowers might suggest the reduction of the protein degradation and increased longevity of the infected plants. Longevity of the infected plants could benefit the pathogen as the plant keep supplying nutrients, which ensures pathogen production of a fluid rich in fungal spores and therefore maintenance of the extended visits of pollinators.

7.2.3.10. Activation of defense responses

Plant glutathione S-transferases (GSTs) are multifunctional proteins that detoxify both xenobiotic and endogenous compounds that accumulate during oxidative stress (Marrs, 1996). GST properties include: (i) vacuolar sequestration of anthocyanins (Kitamura et al., 2004), (ii) binding to auxin proteins (Smith et al., 2003), (iii) binding to cytokinin proteins (Gonneau et al., 2001) and (iv) function in signalling (Ghanta et al., 2011). The majority of the plant GSTs belongs to the tau (GSTU) class, which are plant-specific (Wagner et al., 2002). I found up-regulation of five GSTU genes in pseudoflowers: *AtGSTU1 (At2g29490), AtGSTU2 (At2g29480), AtGSTU4 (At2g29460), AtGSTU17 (At1g10370)* and *AtGSTU26 (At1g17190)* (Fig. 7.4A and Table 7.2). Up-regulation of GSTU genes could result in several effects: 1) GSTUs could act help the plant tolerate biotic stress caused by the rust infection, 2) GSTUs could have a cooperative participation in the binding and inactivation of auxin, together with the other up-regulated auxin-inactivating enzymes *GH3.2 (At4g37390)* and *GH3.4 (At1g59500).*

I also found up-regulation of genes involved in defense and pathogen infection. 1) Up-regulation of a gene encoding the *ADG2-like DEFENSE RESPONSE PROTEIN1 (ALD1)* gene *(At2g13810)* that generates an amino acid-derived molecule important in the activation of defense signalling (Fig. 7.4A and Table 7.2) (Song et al., 2004). 2) Up-regulation of a gene encoding the *TREHALOSE PHOSPHATASE SYNTHASE 11 (AtTPS11) (At2g18700)* that is a plant stress protector and a multifunctional sugar in fungi (Fig. 7.4A and Table 7.2) (Fernandez et al., 2010). 3) Up-regulation of a gene encoding the transcription

factor *APETALA2 (AP2) 6L* (*RAP2.6L) (At5g13330)* that enhances performance under salt and drought stress (Krishnaswamy et al., 2011) and confers resistance against bacterial pathogens (Fig.7.4A and Table 7.2) (Sun et al., 2010).

Remorins are thought to be involved in cellular signal transduction processes (Lefebvre et al., 2010). I found that the remorin gene *(At5g23750)* was up-regulated in pseudoflowers (Fig. 7.4A and Table 7.2). RPM1 interacting protein4 (RIN4), a protein that associates with remorin, is involved in RPM1-mediated resistance in Arabidopsis and was found to be a virulence target of the cognate AvrRpm1 effector of *Pseudomonas syringae* (Liu et al., 2009). Also remorins are found to interfere with viral cell-to-cell viral movement in plants due to the presence of a hydrophobic N-terminal region (Raffaele et al., 2007). I hypothesize that *B. stricta* induces remorins to interfere with the transfer of nutrients to *P. monoica*. Alternatively, host remorins are up-regulated to interfere with R gene mediated resistance to this rust pathogen.

In addition to the biotic stress responses, I identified genes involved in abiotic stress. *Late embryogenesis abundant (LEA)* genes, *LEA4 (At3g17520*) and other two *LEA-like* genes (*At2g46150 and At1g65690*) that play crucial roles in tolerance to water deficit tolerance in *A. thaliana* were up-regulated in pseudoflowers (Fig. 7.4A and Table 7.2) (Olvera-Carrillo et al., 2010). *LEA* genes encode for hydrophilin related proteins that have a high content of water-interacting residues and facilitate collection of water molecules when cells experience changes in water status (Colmenero-Flores et al., 1999). Overexpression of genes encoding LEA proteins enhances tolerance to drought, freezing, salinity and water stresses in transgenic plants (Lal et al., 2008; Puhakainen et al., 2004; Xiao et al., 2007). Up-regulation of *LEA* genes in pseudoflowers could contribute to protecting against abiotic stress.

7.3. Conclusions

In this chapter, using whole-genome expression profiling, I identified and described a large number of genes that show altered gene expression in

Puccinia monoica-induced pseudoflowers ('Pf') compared to *Boechera stricta* stem and leaves ('SL'). Overall plant development is affected in pseudoflowers. I found up-regulation of genes causing hypocotyl elongation, which correlates with the elongated stem. Also, I found up-regulation of genes involved in increased initiation of leaves in the stems, with correlates with the cluster of infected leaves observed in pseudoflowers (Roy, 1993a). Down-regulation of cell wall linked genes points to weakening of the wall facilitating the modification of the host cell by the rust pathogen. Altered cell walls also might contribute to the reduction in diameter and thickening of the stem in pseudoflowers. Leaf morphology is greatly affected in pseudoflowers with up-regulation of genes that control lateral organ development, with increased number of cells in the leaf and organ size. Cuticular wax production and transport is increased in pseudoflowers with the up-regulation of genes involved in wax synthesis and secretion. Down-regulation of the floral signal in pseudoflowers interferes with floral transition. The influorescence architecture is drastically affected in pseudoflowers due to regulation of genes that control floral meristem and floral organ identity. There is probably increased sugar metabolism and transport in pseudoflowers; this is suggested by the up-regulation of genes involved catalysis and transport of sugars. Moreover this finding can also be correlated with the sweet-smelling odour and elevated sugar content found in the surface of pseudoflowers compared to uninfected plants (Roy, 1993a). Infected plants presented a change in volatile compounds synthesis. Analysis of infected leaves revealed the up-regulation of an enzyme that contributes to the degradation of L-phenylalanine and produces phenylacetaldehyde and 2-phenylethanol. These two volatile compounds were chemically detected previously in pseudoflowers and also attributed to its distinct fragrance (Raguso and Roy, 1998). These indicate that there is biosynthesis of novel volatiles in pseudoflowers. In addition, natural floral pigments are shutdown in pseudoflowers and that is consistent with the down-regulation of genes involved in anthocyanin biosynthesis. Also, there is up-regulation of late embryogenesis genes in pseudoflowers that enhanced the tolerance to various abiotic stresses in the infected plant. Some genes involved in defense responses to biotic stress were up-regulation in pseudoflowers. In the near future, the goal is to sequence the transcriptome of *P. monoica* to discover and identify putative secreted effector molecules from the *P. monoica* that could

modulate *B. stricta* and that could associated with the dramatic phenotype in pseudoflowers.

CHAPTER 8: General Discussion and Outlook

8.1. Signal peptides in host-translocated effectors

Effector proteins must be secreted to reach their cellular targets in the apoplast or the cytoplasm of the host cell (Kamoun, 2006). In oomycetes, as in other eukaryotic organisms, the majority of the secreted proteins are thought to be secreted through the general secretory pathway, via short N-terminal amino acid signal peptide sequences (Torto et al., 2003). *Phytophthora infestans* genome contains two types of host-translocated effectors. RXLR effector proteins have a N-terminal RXLR motif that function in translocation. 86% of the annotated RXLR effectors are predicted to carry signal peptides. There are at least 79 RXLR genes that are induced during potato infection, and they include all known *P. infestans* effectors with an avirulence activity. The induction pattern of these 79 RXLR genes suggests that they might function during pathogenesis of *P. infestans.* In chapter 3, I report the functional validation of the signal peptides of four *in planta*-induced RXLR effector genes of *P. infestans* (*PexRD6/ipiO* (*Avrblb1*), *PexRD39* (*Avrblb2*), *PexRD40* (*Avrblb2*) and *PexRD8*) using the yeast signal sequence trap method (SST). *PexRD6/ipiO (Avrblb1), PexRD39 (Avrblb2)* and *PexRD40 (Avrblb2)* are avirulence proteins that are recognized by *Rpi-blb1* and *Rpi-blb2* respectively, resulting in the induction of hypersensitive cell death and immunity (Oh et al., 2009; Vleeshouwers et al., 2008). These four AVRs are secreted and then translocated to the host cytoplasm via the RXLR motif where they are recognized by the cognate R protein. This model is assumed for all RXLR-containing effector proteins of *P. infestans* and other haustoria-forming oomycete pathogens (Morgan and Kamoun, 2007; Schornack et al., 2009). The data I obtained that the signal peptides of these proteins are functional in yeast supports this model. I also functionally validated the signal peptide of the effector PexRD8 that suppresses the hypersensitive cell death produced by PAMP-like protein *P. infestans* INF1 (Oh et al., 2009). Secretion of INF1 has been previously functionally validated using proteomics and it is described as the major secreted elicitin in *P. infestans* (Kamoun et al., 1997). The mechanism by which INF1 is suppressed by PexRD8 is unknown, but it is possible to speculate that a

PexRD8-interactor protein translocated to the host cytoplasm could mediate cell death as described in another INF1-suppressing *P. infestans* RXLR effector AVR3a (Bos et al., 2010; Bos et al., 2006).

RXLR proteins generally contain signal peptides (only 14% do not have signal peptides), even in proteins where the RXLR motif deviates from the consensus. For example ATR5 from the haustoria forming-oomycete *H. arabidopsidis* has no clear RXLR motif but still contains the EER sequence, a second motif present next to the RXLR motif. ATR5 carries an intact signal peptide and is translocated and also recognized intracellularly; triggering immunity in the host (Bailey et al., 2011). This suggests that (i) the presence of signal peptides is crucial for the identification of effectors with putative roles in pathogenicity; (ii) other means could be used by these effectors in order to be translocated inside the host cell after they are secreted. For example, although the RXLR-EER twin peptide motif has been shown to be required for translocation, it is possible that the EER motif alone could be sufficient signal for the translocation (Dou et al., 2008b; Grouffaud et al., 2008; Whisson et al., 2007).

Besides RXLRs, in *P. infestans* there is another class of ancient host translocated effectors termed crinkler (CRN) that elicit necrosis *in planta* (Haas et al., 2009). CRN effectors are also modular proteins that carry a N-terminal signal peptide followed by the translocation motif LFLAQ and an adjacent diversified DWL domain. CRNs also have a putative motif HVLVXXP that is junction point in the diversity of domains observed in CRNs, which are thought to evolve by recombination. CRNs are shown to target the host nucleus and also expression of some members can induce cell within plant cells (Haas et al., 2009; Schornack et al., 2010). Only 60% CRNs are predicted to carry signal peptides indicating that the frequency of signal peptides in CRNs is lower compared to RXLRs. It is possible that the loss of signal peptides is more likely to happen in CRN genes compared to RXLRs, because CRNs can shuffle and fuse to N-terminal sequences that lack signal peptides. Within the set of CRNs that carry signal peptides, 9% (24 proteins) were predicted to be secreted with HMM scores>0.90 (see appendix 1.3). The majority of these predicted secreted CRNs have lower HMM scores compare to RXLRs. I found that 14 out the 24 secreted CRNs have

HMM scores <0.980 (see appendix 1.3) compared to the 449 out the 483 secreted RXLRs that have HMM scores >0.999 (see chapter 3 Fig. 3.2). The lower HMM scores in CRN signal peptides suggest that there are differences in the sequence that are detected by the signalPv2.0 algorithm. Another difference of CRNs with RXLRs is related to their expression patterns *in planta*. CRNs genes are highly expressed in mycelia, but only 12 genes are induced during the biotrophic phase of infection on potato (see appendix 1.3) compared to 79 induced *in planta* RXLR genes (see chapter 3 Fig. 3.3). Although some of CRNs are induced *in planta*, this raises questions about the extent to which CRN genes are implicated during biotrophy as is predicted in the RXLR genes. However, it is likely that there are technical problems using microarrays for the accurate measurements of CRN gene expression given the repetitive and chimeric nature of these genes. Therefore, for this family, it will be more accurate to determine gene expression using specific oligonucleotide primers that hybridize to particular CRN domains.

In conclusion, functional validation of *in planta*-induced secretory host translocated RXLR proteins has assisted in the discovery of a large set of potential candidate effectors. However, more experiments are needed to test if the CRN signal peptides with lower HMM scores are secreted in yeast and *Phytophthora*.

8.2. Widely occurring apoplastic effector families and their functions in oomycetes

Effectors not only target the host intracellular space but also the extracellular space, and these are called apoplastic effectors. Apoplastic effectors include secreted hydrolytic enzymes that probably degrade plant tissue; enzyme inhibitors to protect against host defence enzymes; and necrotizing toxins and PcF-like small cysteine-rich proteins (Kamoun, 2006). Elicitins as other oomycete effectors are modular proteins that carry N-terminal signal peptides and a C-terminal conserved eliciting domain that can trigger defenses in a variety of plants (Kamoun, 2006; Nurnberger and Brunner, 2002; Vleeshouwers et al.,

2006). Because various plants can respond with an immune response to elicitins, and the fact that they are structurally conserved proteins in oomycetes (Baxter et al., 2010; Haas et al., 2009; Kemen et al., 2011; Tyler et al., 2006), indicates that elicitins have features of PAMPs (Chaparro-Garcia et al., 2011; Kanzaki et al., 2008; Nurnberger and Brunner, 2002; Vleeshouwers et al., 2006). Elicitin genes are generally expressed across many developmental stages and they can be down-regulated during the biotrophic phase of infection (Haas et al., 2009; Jiang et al., 2006b; Qutob et al., 2003). Protease inhibitors of both classes of Kazal-like serine protease and cystatin-like cysteine protease inhibitor are also conserved across several oomycetes species like elicitins (see chapter 4).

In contrast to elicitins, protease inhibitors are induced during the biotrophic phase of infection (chapter 4 Table 4.1). Protease inhibitors are predicted to interact with extracellular enzymes rather than with plant receptors and exhibit a dynamic evolutionary history (Kamoun, 2006; Schornack et al., 2009; Song et al., 2009; Tian et al., 2005; Tian et al., 2004; Tian and Kamoun, 2005; Tian et al., 2007). Two examples are the *P. infestans* cystatins EPIC1 and EPIC2 that bind and inhibit several tomato apoplastic proteases (Song et al., 2009; Tian et al., 2007). Unlike other cystatins that do not have inhibitory activities *epiC1* and *epiC2* are induced *in planta* and lack orthologs in *Phytophthora* sojae and *Phytophthora ramorum*, and even in more closely related species to *P. infestans* such *Phytophthora phaseoli* (see chapter 4 Table 4.1, also see chapter 5, PITG_09169, PITG_09175 and PITG_09169 in Fig. 5.5) (Raffaele et al., 2010a; Song et al., 2009; Tian et al., 2007). In other species like *Phytophthora mirabilis*, *epiC1* gene is under positive selection compared to *P. infestans* (see chapter 5, PITG_09169, PITG_09175 and PITG_09169 in Fig. 5.5) (Jing Song, unpublished data) (Raffaele et al., 2010a). This suggests that enzyme inhibitors are target to selection pressures to adapt their inhibition activities to various host proteases. Also, host proteases are subject to variation, it might be that target proteases across the different *Phytophthora* hosts are differentially inhibited by EPICs (Jing Song and Joe Win, unpublished data).

Cystatin-like cysteine protease inhibitors from oomycetes like in animals and plants have three conserved domains named NT (NT), Loop1 and Loop (L2) (see

chapter 4 Fig. 4.5) (Song et al., 2009; Tian et al., 2007). Interestingly, cysteine protease inhibitors of parasitic nematodes have an additional conserved SND domain before the loop1 that inhibits asparaginyl endopeptidase enzymes that control antigen processing in the host (Alvarez-Fernandez et al., 1999; Gregory and Maizels, 2008). I found a putative second motif RXC (an Arg, a variable amino acid Ile/Val/Leu/Met/Pro, and Cysteine) before Loop1 (L1) that is only present in oomycetes and not in plant or animals cystatins (see chapter 4 Fig. 4.5). This putative conserved motif may contribute to activity against host enzymes. Further experiments will help to determine if RXC has a functional role.

Kazal-type serine protease (EPI) inhibitors *epi1* and *epi10* genes are induced *in planta* (see chapter 4 Table 4.1) and have been predicted to inhibit the plant subtilisin A (Tian et al., 2005; Tian et al., 2004). These two protease inhibitors are divergent in sequence but present both atypical Kazal-like domains with two disulfide bridges (Tian et al., 2005). Although the typical Kazal-like domains with three disulfide bridges are structurally conserved across various oomycete species, the above described atypical Kazal-like domains are present only in *Phytophthora* species. Are these atypical Kazal-like domains a specialized structural variation of serine protease inhibitors in *Phytophthora?* It would be useful to carry out predictions using the Laskowski algorithm (Tian et al., 2004; Tian and Kamoun, 2005) of the 15 Kazal inhibitors containing these atypical Kazal-like domains predicted in *P. infestans* as well as evaluate their inhibition activity. This information will help to point to putative targets and the functional relevance of these atypical domains in *Phytophthora*.

Both atypical and typical Kazal-like domains in the oomycete Kazal inhibitors contain a P1 residue that contributes to specificity (Lu et al., 2001). At least half of Kazal-like domains in *P. infestans* (30 out of 60) have aspartic acid (Asp) P1 residue that is uncommon in natural Kazal inhibitors of animals and apicomplexans (see chapter 4 Fig. 4.4) (Tian et al., 2004). In animals, P1 aspartic residues are present in inhibitors of cysteine proteases with caspase activity and involved in the initiation of cell death (Schaller, 2004). Interestingly, there are plant proteases that can cleaves animal caspases substrates, suggesting that Asp specific plant proteases could be involved in the regulation of

programmed cell death (PCD) in plants (Schaller, 2004). It was discovered that oat contain proteases named saspases that are involved in pathogen-induced programmed cell death. These saspases have caspase activity and resemble subtilisin-like serine proteases (Coffeen and Wolpert, 2004). It was already hypothesized that *Phytophthora* EPIs that carry aspartate as the P1 residue might target plant saspases and suppress host cell death (Tian et al., 2004). The finding that the majority of Kazal-like domains in *P. infestans* confirm previous observations and support the above hypothesis.

Although the P1 residue with the amino acid Asp was the most abundant in *P. infestans,* pathogen of *Solanum* species, this P1 amino acid residue is variable in other oomycete Kazal-like inhibitors (see chapter 4 Fig. 4.4). This suggests that target specificity may not be as marked in other oomycetes. I found that oomycetes with broad host range like *Pythium ultimum* have P1 residues such as Ala, Glu and Met in addition to Asp (see chapter 4 Fig. 4.4). It is possible that a wider repertoire of amino acids in the P1 residue might benefit the pathogen and result in a powerful counter defense and the inhibition of a broader range of host proteases.

8.3. How do effectors evolve?

Given that the phenotype of the effectors extends to plant cells, they are expected to be the direct target of the evolutionary forces that drive the interplay between pathogen and host. During this interplay, effectors will face at least three hypothetical scenarios over time: neutral (no selection), adaptive and/or relaxed selection (leading to pseudogenisation) (Kamoun, unpublished). The model consists of (i) neutral selection in cases when pathogen effector is recognized with no significant differences; (ii) adaptive/purifying selection in cases when the pathogen effector adapts to avoid recognition of the target and suppress defenses in the host; (iii) relaxed selection when the target is absent.

I found that several Avr effectors in the emergent clonal lineages of *P. infestans* 13_A2 genotype have evolved to overcome recognition by the cognate R genes

(see chapter 6). To overcome resistance, these effectors were subject to selective pressures as explained in the above model. For example, I found that the *Avr2* gene was highly polymorphic in the sequenced 13_A2 isolate 06_3928A (see chapter 6). It was quite challenging to identify this new variant, as it was only detected after de novo assembly of 06_3928A Illumina reads that did not map back to the genome of reference strain *P. infestans* T30-4 (see chapter 2 section 2.4.4 and chapter 6). The new variant of *Avr2* evades recognition by the cognate *R2* resistance gene and explains the virulence of 06_3928A on *R2* plants (Gilroy et al., 2011). Another example of an effectors gene that has been under selective pressure in *P. infestans* 06_3928A is the *Avr4* effector gene (van Poppel et al., 2008). I found that *Avr4* contains a frameshift mutation and the gene was also not induced *in planta*. *Avr4* also evades recognition by the cognate *R4* resistance gene and explains the virulence of 06_3928A on *R4* plants (David Cooke, unpublished) (see chapter 6). Some other *Avr* effectors of the *P. infestans* isolate 06_3928A presented neutral (no) selection. I found that 06_3928A carries intact *Avrblb1, Avrblb2* and *Avrvnt1* effector genes that are induced *in planta*. *Avrblb1, Avrblb2* and *Avrvnt1* genes are recognized in potato lines that carry the corresponding R immune receptor genes *Rpi-blb1*, *Rpi-blb2*, *Rpi-vnt1.1* (see chapter 6) (David Cooke, unpublished).

Effectors from closely related *Phytophthora* species were detected to evolve rapidly due to selective pressures like host adaption (see chapter 5). The hypothetical scenarios mentioned above could be applied to many effectors from *Phytophthora* clade 1c species, such as *P. infestans*, *P. ipomoeae*, *P. mirabilis* and *P. phaseoli*. These species infect unrelated host plant species consistent with evolution by host jumps (Grunwald and Flier, 2005; Raffaele et al., 2010a). Previous analyses of the *P. infestans* genome architecture showed an uneven distribution with gene sparse regions being highly populated with effectors (Haas et al., 2009; Raffaele et al., 2010b). Comparative analyses of the sequenced genomes of the *Phytophthora* clade 1c species showed that these gene-sparse regions are enriched in effectors with fast evolving features and genes that are induced *in planta* (Raffaele et al., 2010a). Gene-sparse regions of the *Phytophthora* clade 1c species were suggested to experience accelerated rates of evolution following host jumps (see chapter 5) (Raffaele et al., 2010a).

8.4. Flower mimicry by plant pathogens

Plant pathogens can produce mimics that resemble host components in both form and function (Elde and Malik, 2009). *Puccinia monoica* is a rust fungus that infects *Boechera stricta* and inhibits host flowering and interestingly has the ability to modify host plant leaves to produce "pseudoflowers" to promote its own reproduction (Roy, 1993a). Pseudoflowers are described as the most dramatic form of mimicry in plant-parasitic pathogens; since they resemble host components in both form and function (Ngugi and Scherm, 2006). Pseudoflowers mimic true flowers in shape, color, scent, and production of sugar nectar from co-occurring and unrelated flowering plant species (Roy, 1993a). In chapter 7, I identified biological processes in the host that are significantly perturbed (differentially regulated) by *P. monoica* in infected *B. stricta* plants. These results suggest that formation of pseudoflowers involves extensive reprogramming of the host including alteration of flower, shoot and leaf development, cell wall and cell surface modifications, and volatiles synthesis. Which factors are involved in the modification of these host processes and the production of pseudoflowers? I proposed that *P. monoica* secretes effectors that alter these biological processes leading to the development of pseudoflowers. It is possible that this pathogen evolved a battery of effectors that modify the host processes and mimic host components. In the future, the identification of these effectors would be important in understanding how the pathogen triggers flower mimicry.

How can *P. monoica* effectors mediate these extensive reprogramming of the host? Effectors may bind host proteins and control transcriptional regulation of plant genes. This could be the case for plant SWEET sugar transporters that are induced during infection by pathogenic bacteria, including fungi (Chen et al., 2010). It has been demonstrated that induction of *SWEET* genes was caused by the direct biding of a type III secretion effector to the promoter of the SWEET genes (Chen et al., 2010). The induction of SWEET is proposed to release nutrients that could be used for the pathogen (see chapter 7 Fig. 7.4). To study this interaction in detail it will be necessary to obtain the genome and transcriptome sequences of the pathogen to reveal the effector repertoire and their functions.

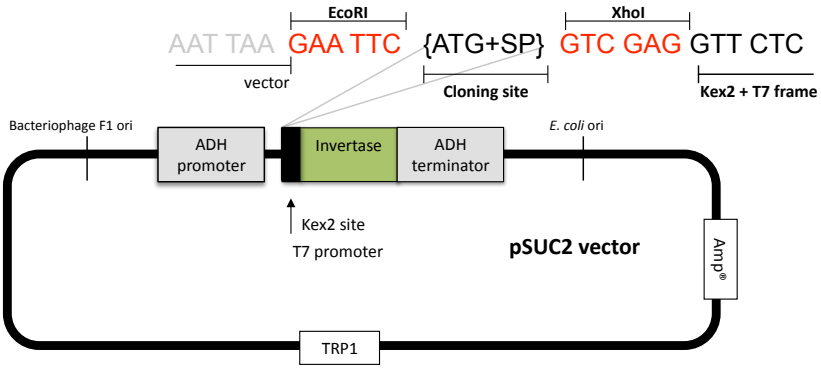
APPENDICES

APPENDIX 1: Signal Sequence Trap (SST) system

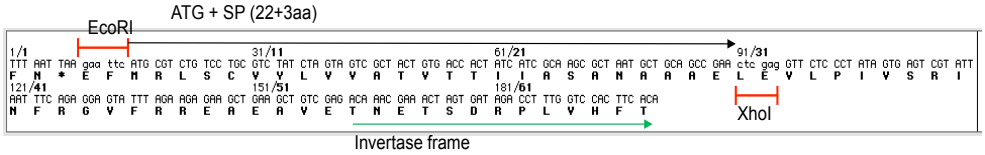
Appendix 1.1. List of 79 *Phytophthora infestans* secreted RXLR effectors that are induction during infection on potato

Appendix 1.2. Example of signal peptide sequences fused to pSUC2 vector

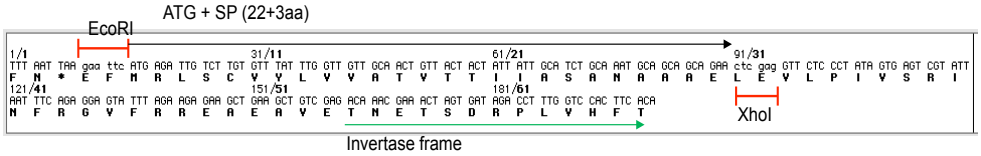
Signal peptides of four RXLR effectors PexRD8 with upstream EcoR1 and upstream XhoI sites fused to in frame invertase mutant gene. PexRD8 sequence was codon optimized for expression in yeast (see chapter 2 section 2.1.1)



PexRD8

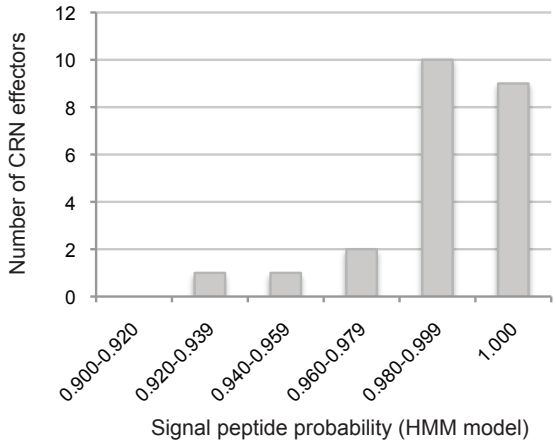


PexRD8 codon optimized to be expressed in yeast

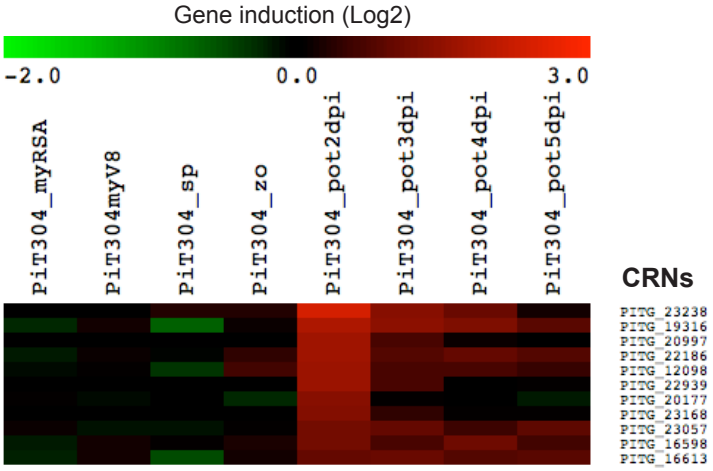


Codons of the ATG+SP of interest are added upstream the XhoI site in frame with the invertase mutant gene
>EcoRI_ATG+SP+3aa_Xho1_Kex2site-T7promoter_Invertaseframe
EFMRLSCVYLVVATVTTIIASANAAAELEVLPIVSRINFRGVFRREAEAVETNETSDRPLVHFT

Appendix 1.3. Distribution of signal peptide probabilities in CRN effectors predicted to be secreted in *P. infestans*



Appendix 1.4. *P. infestans* CRNs effector genes of that are induced *in planta*



x

APPENDIX 2: Kazal-like and cystatin-like protease inhibitors domains from oomycetes pathogens

Appendix 2.1. List of 140 Kazal-like domains predicted in 64 serine protease inhibitors from seven pathogenic oomycete species

Specie	Kazal-like domain*	P1 residue	Type domain	Domain sequence
Phytophthora infestans	Pi_EPI1_d1	D	atypical	CPEYCLDVYDPVGDGEGNTYSNECYMKRAKCHNETTPPAWKDLVL
Phytophthora infestans	Pi_EPI1_d2	D	typical	CSTVCPDVELPVCGSNRVRYGNPCELRIAACEHPELNIVEDSGKAC
Phytophthora infestans	Pi_EPI1-LIKE1_d1	D	atypical	CPQICLDYYTPVADEEGNFYSNECYMKRAKCEKNSARTNSSSIND
Phytophthora infestans	Pi_EPI1-LIKE1_d2	D	typical	CPDSCPDIALPVCVSDGIKYSNPCELKIAACKHPERKIVEFSYSSTC
Phytophthora infestans	Pi_EPI2_d1	D	atypical	CPKYCLDIDDPVGDEEGNMYSNECYMKRAKCAKNKPTDPPFWKNF
Phytophthora infestans	Pi_EPI2_d2	D	typical	CSSGCPDVELPVCGSDGVRYGNPCELKIAACEHPELNIVEAVGMGC
Phytophthora infestans	Pi_EPI2-LIKE1_d1	A	atypical	CPNIMCPAVYQPVSDENGVMYPNKCSMEAAKCKGPRENPLDEYKRI
Phytophthora infestans	Pi_EPI2-LIKE1_d2	D	typical	CASACPDVVLRVCGSDGVWYSNPCELKIAACKNPEQNIVEEEGAC
Phytophthora infestans	Pi_EPI2-LIKE2	H	atypical	CNFACFHVMRPVKDENGVMYPNECEMRRARCRKNEQNVDVQGEQE
Phytophthora infestans	Pi_EPI3	E	typical	CADMLCPEVHDPVCGTDKVTYPNECDLGLAQCAHPERNITVFARSTC
Phytophthora infestans	Pi_EPI4_d1	T	typical	CDAICPTDYEPVCGSDGVTYANDCAFGIALCKTATLSLLAVGEC
Phytophthora infestans	Pi_EPI4_d2	D	atypical	CPDACVDVYDPVSDESGKTYSNECYMRMAKCKDKKKDVDILAEYK
Phytophthora infestans	Pi_EPI4_d3	D	typical	CAAACPDIYSPVCGSDGVTYSSPCHLKLASCKKPKIKLVQDSADSC
Phytophthora infestans	Pi_EPI5	R	typical	CDDNCQRDLMPVCGSDGATYGNDCLLDFAHCENSTITKLHDGKC
Phytophthora infestans	Pi_EPI5-LIKE	D	typical	CDDNGERDFTPVCGPDGITYGNVDFAHCESSAITKKPDGDC
Phytophthora infestans	Pi_EPI6_d1	Q	atypical	CNFVCIQVMSPVTDENGVTYSNECMMHAAKCKDNGKREDPLEEYK
Phytophthora infestans	Pi_EPI6_d2	D	atypical	CPNIMCLDVYEPVTDENGVTYPNKCSMEAAKCKGPRENVLDEYKRI
Phytophthora infestans	Pi_EPI6_d3	D	typical	CASACPDVELPVCGSDGVRYSNPCELKIAACKNPEQNIVEEDGAC
Phytophthora infestans	Pi_EPI6-LIKE1_d1	Q	atypical	CNFVCIQVMRPVTDENGVTYSNKCMMRAAKCKGNGKREDPLEEYK
Phytophthora infestans	Pi_EPI6-LIKE1_d2	D	atypical	CPNIMCLDVYGPVTDENGVAYPNKCSMEAAKCKGPRENVLDEYKRI
Phytophthora infestans	Pi_EPI6-LIKE1_d3	D	typical	CASACPDVELPVCGSDGVRYSNPCELKIAACKNPEQNIVEEDGAC
Phytophthora infestans	Pi_EPI6-LIKE2_d1	A	atypical	CNFVCIQVMRPVTDENGVTYSNECMMRAAKCKGKGKREDPLEEYK
Phytophthora infestans	Pi_EPI6-LIKE2_d2	A	atypical	CPNIMCPAVYQPVTDENGVTYPNKCSMEAAKCKGPRENVLDEYKRI
Phytophthora infestans	Pi_EPI6-LIKE2_d3	D	typical	CASACPDVELPVCGSDGVRYSNPCELKIAACKNPEQNIVEKDGAC
Phytophthora infestans	Pi_EPI6-LIKE3_d1	K	atypical	CNFACIKMMSPVTDENGVTYSNECMMRAAKCKGNWNQDPLEEYKR
Phytophthora infestans	Pi_EPI6-LIKE3_d2	D	atypical	CPNIVCLDVYEPVTDENGVTYPNQCSMDVEKCKGPREDVYDEYKRI
Phytophthora infestans	Pi_EPI6-LIKE3_d3	D	typical	CATACPDVKFYVCGSDGVWYSNPCELKIAACENPEQNIVEKDGAC
Phytophthora infestans	Pi_EPI7	D	typical	CSQVCPDVYEPVCGTDSVTYSNSCELGIASCKSPEKNIAKKINGRC
Phytophthora infestans	Pi_EPI7-LIKE	D	typical	CPDACPDVYTPVCGSDGNTYSNSCFLGIASCKNPDKHIAQASEGSC
Phytophthora infestans	Pi_EPI8_d1	D	typical	CSFGCPDVYEPVCGSNGKTYSNSCYLRLESCQNNNEITEAGNGEC
Phytophthora infestans	Pi_EPI8_d2	D	atypical	CPACLDVYEPVTDENGNVYSNECYMKMAKCKGADDDASMRSDSP
Phytophthora infestans	Pi_EPI9	R	typical	CPTRCTRDYRPICGSDGITYANKCLFKVGQCLDPSLKKFHKGKC
Phytophthora infestans	Pi_EPI9-LIKE	R	typical	CSGLCTRDLMRVCGSNGVTYDNECWFEVVQCEGPGIKLKNKGRC

Appendix 2.1. List of 140 Kazal-like domains predicted in 64 serine protease inhibitors from seven pathogenic oomycete species

Specie	Kazal-like domain*	P1 residue	Type domain	Domain sequence
Phytophthora infestans	Pi_EPI10_d1	D	typical	CSFGCLDVYKPVCGSNGETYSNSCYLRLASCKSNNGITEAGDGEC
Phytophthora infestans	Pi_EPI10_d2	D	atypical	CPDMCLDVYDPVSDENGKEYSNQCYMEMAKCKGTGYDDNKRSGNP
Phytophthora infestans	Pi_EPI10_d3	D	typical	CGDMLCPDNYAPVCGSDGETYPNECDLGITSCNHPEQNITMVGEGT
Phytophthora infestans	Pi_EPI11_d1	D	typical	CPSLCTDLFAPVCGSDGVTYSNDCYLLLADCESAARITKVSDGKC
Phytophthora infestans	Pi_EPI11_d2	K	typical	CSGVCPKILKPVCGSDGVTYPNECLLGVADCECSDDITKAYDGEC
Phytophthora infestans	Pi_EPI11_d3	E	typical	CNDVCPENFQPVCGSDGVTYSNDCTLEYAQCTSGGVITKVSEGEC
Phytophthora infestans	Pi_EPI11_d4	E	typical	CSEVCPEVFEPVCGSDGVTYSDSCFLGIATCKDPSIVLAHDGAC
Phytophthora infestans	Pi_EPI11_d5	E	typical	CPDVCIEIFRPVCGSDGVTYANSCFLGIASCHDPSITLAHNGAC
Phytophthora infestans	Pi_EPI11_d6	E	typical	CPDVCIEIFRPVCGSDGVTYANSCFLGLASCEDPRIAQAHEGPC
Phytophthora infestans	Pi_EPI11_d7	A	typical	CPDICPAIYAPVCGSDGVTYSNECLLNIASCNHPELKLTKASDGAC
Phytophthora infestans	Pi_EPI12	S	typical	CDKRNCESHKGRVCSNGNQTYATLCDLTSVMCNHPTRGVSLAYDGPC
Phytophthora infestans	Pi_EPI12-LIKE	D	typical	CNKKNCKDHVGPVCGNDNVTYASLCDLTSVMCEHPERRVGMGYDGPC
Phytophthora infestans	Pi_EPI15	D	typical	CDQVCPDVNERVCGTDGVTHTNSCYLGVASCKNPDKNIALVSNGAC
Phytophthora infestans	Pi_EPI16	D	atypical	CPDACLDVYSPVIGDDGISYPNECSMQMAKCKKSGKKDDWYASYK
Phytophthora infestans	Pi_EPI16-LIKE	Q	atypical	CAGACMQVDAPVLGDDGIWYTNACEMRMAKCEKSGKKARTQREAL
Phytophthora infestans	Pi_EPI17_d1	D	typical	CDTECPDDFNPVCGSDHVTYTNDCAFTVAQCNATELVVANSGEC
Phytophthora infestans	Pi_EPI17_d2	M	atypical	CPDACTMEYSPVTDENGKKYSNECAMRLAKCKGEAGEEKKIVTFA
Phytophthora infestans	Pi_EPI18_d1	L	typical	CKLNCQLISSPVCGSDNVSYANSCFLKEARCSTGNTDLHVIFRGLC
Phytophthora infestans	Pi_EPI18_d2	Q	typical	CPATCTQTYSPVCASNGQLYGNECRFRQAKCSRLGLLAVNLEPRTLAEC
Phytophthora infestans	Pi_EPI19_d1	A	typical	CSKAFECDAVSHAPVCGSDGTTYANSCAFASVFCSSEHDADTLFIQALGEC
Phytophthora infestans	Pi_EPI19_d2	R	typical	CNPMCERVYDPVCGSDGITYANLCLLEYAECRNPNVKMFGPGKC
Phytophthora infestans	Pi_EPI19_d3	Y	typical	CIPEPCPYTFAPVCGSDGQTHDNLCLFANAKCQQPTLTVIHEGEC
Phytophthora infestans	Pi_KAZAL-LIKE1	P	typical	CADTPCLPEHAPVCGSNGVTYENECELGQANCNNAGLNVTQVSYGAC
Phytophthora infestans	Pi_KAZAL-LIKE2	P	typical	CADTPCLPEHAPVCGSNSVTYENECELDQANCNNAGLNVTQVSYGAC
Phytophthora infestans	Pi_KAZAL-LIKE3	P	typical	CADTPCLPEHAPVCGSNGVTYENESELDQANCNNAGLNVTQVSYGAC
Phytophthora infestans	Pi_KAZAL-LIKE4	P	typical	CADTPCLPEHAPVCGSNGVTYENECELDQANCNNAGLNVTQVSYGAC
Phytophthora infestans	Pi_KAZAL-LIKE5	P	typical	CADTPCLPEHAPVCGSNGVTYENECELDQANCNNAGLNVTQVSYGAC
Pythium ultimum	Pu_PYU1_T012157	T	typical	CDLGCGTHWSPICASDGVTYRNACTLEEAYCEDHDVRPLHNGEC
Pythium ultimum	Pu_PYU1_T012160_d1	G	typical	CEAIECSGDWQSDNPVCGSNGVRYESLCAFELVKCENPSLGDVHVAPC
Pythium ultimum	Pu_PYU1_T012160_d2	D	typical	CELSCEDLWSPICGSDDVTYRNPCHLEEAFCRNHQVEPTYYGVC
Pythium ultimum	Pu_PYU1_T012156_NS	S	typical	CARDCGSNRAPICASDGVTYANSCLFDQAHCVNNELLPMHYGDC
Pythium ultimum	Pu_PYU1_T012159_d1	V	typical	CHPDSCIVSVPQLLCGSDGVTYRSICELELAQCTRPDLKIASMGAC
Pythium ultimum	Pu_PYU1_T012159_d2	E	typical	CSEQEACEESSYPICGSDGVTYQNACYFDRAYCKNNDLVPMGYGTC

Appendix 2.1. List of 140 Kazal-like domains predicted in 64 serine protease inhibitors from seven pathogenic oomycete species

Specie	Kazal-like domain*	P1 residue	Type domain	Domain sequence
Pythium ultimum	Pu_PYU1_T012158_d1	V	typical	CHPDSCIVSVPQLLCGSDGVTYRSICELELAQCTRPDLKIASMGAC
Pythium ultimum	Pu_PYU1_T012158_d2	E	typical	CSEQEACEESSYPICGSDGVTYQNACYFDRAYCKNNDLVPMGYGTC
Pythium ultimum	Pu_PYU1_T012161_d1	A	typical	CAPESCTAAAQKLLCGSDGVTYTSACELELAQCSHPTLQLASVGAC
Pythium ultimum	Pu_PYU1_T012161_d2	D	typical	CETVKCGDHANPICASNGVTYQNACDFDRAYCKNKELAPVSYGAC
Pythium ultimum	Pu_PYU1_T000142_d1	E	typical	CAAACPENYKPLCGSDGKTYSNECMLEYAKCSTNSTTLTVASDGEC
Pythium ultimum	Pu_PYU1_T000142_d2	S	typical	CLQEIACLSVIDYVCGSDGKTYNNACELRKAKCQNPSLTQVSTGEC
Pythium ultimum	Pu_PYU1_T000142_d3	K	typical	CTTTMCTKIYLPVCGSDDKTYSNECEFKNAQCKATSPLTLKANVSC
Pythium ultimum	Pu_PYU1_T000142_d4	T	typical	CSTVCTTEFNPVCGSNGITYNNACLLKNAQCTNSTVTKAADGAC
Pythium ultimum	Pu_PYU1_T010209_d1	A	typical	CSDACGALYQPVCGSDGKTYPNECTLSVANCKSPELKLTVKSPGAC
Pythium ultimum	Pu_PYU1_T010209_d2	K	typical	CKQTCSKIRKPVCGSDGTMYSNLCILKNAQCDNSEIMQMAEDKC
Pythium ultimum	Pu_PYU1_T010209_d3	M	typical	CTTMCTMELDPVCGSDGKTYSNPCALKNAQCENPKSNIVVKAAGEC
Pythium ultimum	Pu_PYU1_T010209_d4	A	typical	CPSMCTADYTPVCGSDGKTYSNKCQLSIAKCKNPTSNISLKSEGEC
Pythium ultimum	Pu_PYU1_T010209_d5	K	typical	CEMACTKQYAPVCGSNGKTYTNSCALKLANCKSSKKEITIRSEGAC
Pythium ultimum	Pu_PYU1_T014337_d1	V	typical	CDRTCEVTDAAVCGNDDVTYANYCFFSVAACKNKTLALAYTSPC
Pythium ultimum	Pu_PYU1_T014337_d2	L	typical	CDRFCTLEYEPVCGSDGVTYGNACAFDEANCRAGGGLAVKAVGTC
Pythium ultimum	Pu_PYU1_T013339_d1	A	typical	CKAVKCDARANTPVCGSDGKSYANDCLFEFARCNDAALTLVAKTSC
Pythium ultimum	Pu_PYU1_T013339_d2	R	typical	CNTDCTRELDQMCGSDGKTYNNQCLFDNAKCLNPALVVVKNDAC
Pythium ultimum	Pu_PYU1_T009699_d1	D	typical	CDNRSVCTDKDPNVCGSDGDTYVNKCTFESAYCDEPNELFFIVSDGAC
Pythium ultimum	Pu_PYU1_T009699_d2	L	typical	CQIKCALDGVPVCGTDGKPYINDCHLLAAKCKFPNLAKAYNGAC
Pythium ultimum	Pu_PYU1_T009699_d3	R	typical	CNPICARVYEPVCGSNSVTYANQCLLDYAACKNPRVTKLSNGKC
Pythium ultimum	Pu_PYU1_T009699_d4	S	typical	CVPVACTSEEDPVCASNGASYLNTCMFENAQCQFPELSILHEGEC
Pythium ultimum	Pu_PYU1_T000511_d1_NS	M	typical	CSRLCPMIDSPVCGSDGVSYANACYFDEAQCNNPGLSIAVHALC
Pythium ultimum	Pu_PYU1_T000511_d2_NS	D	typical	CDAIVCADIDDPVCTTSGTMKNACFLKREQCKHPYVELLRRGSC
Pythium ultimum	Pu_PYU1_T000511_d3_NS	Q	typical	CPASCGQEYAPVCASSGVIYGNECLFRQAKCARAFASNFVARDLSYC
Pythium ultimum	Pu_PYU1_T009700_d1	M	typical	CAIGDKRQCIMIYAPVCASNGQTYGNVCQFSSAYCTLPEAEREGLKIVHDGEC
Pythium ultimum	Pu_PYU1_T009700_d2	D	typical	CALFKCSDTGDGVCASDGKTYVNACLVRAAGCANPGLFVVSDKPC
Pythium ultimum	Pu_PYU1_T009700_d3	P	typical	CMVPQCAPIDKPICASNGKTYMNRCLFSYDECKNPSLRVAHSGAC
Pythium ultimum	Pu_PYU1_T005024_d1_NS	L	typical	CTIRDCKLTHDDVNVCGSDGKTYLNECLFRNAQCRSNDALTRNTNWNGYRC
Pythium ultimum	Pu_PYU1_T005024_d2_NS	E	typical	CGHTITCKEIGKYVCGSDGNVYFGYCNLYVAQCVDPSVEEIEC
Pythium ultimum	Pu_PYU1_T014335_d1	A	typical	CTLLTECPADPDPSESVCGDDYNAYPSACSLLLTHCQHPGAVGPYPLEGAVPPTC
Pythium ultimum	Pu_PYU1_T014335_d2	M	typical	CAFVCPMFYAPVCTDDGHVYENKCVYASARCRDTALTEANADNC
Saprolegnia parasitica	Sp_SPRG_16334_d1	M	typical	CPITVCPMYYQPVCGSDRVTYSNKCELEVAACKTPGLIMANATVC
Saprolegnia parasitica	Sp_SPRG_16334_d2	E	typical	CPRYCLEIYRPVCGSDGKTYSNECELNIAACKNPSLTRVRDGPC

Appendix 2.1. List of 140 Kazal-like domains predicted in 64 serine protease inhibitors from seven pathogenic oomycete species

Specie	Kazal-like domain*	P1 residue	Type domain	Domain sequence
Saprolegnia parasitica	Sp_SPRG_16334_d3	K	typical	CPDACHKMYEPVCGSNWVTYANKCLLEAAQCRNPSILLAATGDC
Saprolegnia parasitica	Sp_SPRG_16334_d4	R	typical	CNYACMRSYDPICASDKRTYSNWCEFSKAVCKQPELTFRSIGVC
Saprolegnia parasitica	Sp_SPRG_05363_d1	D	typical	CPTVCIDLFDEVCGSDGKTYTNECKLDIAACADPTIKLVSKGAC
Saprolegnia parasitica	Sp_SPRG_05363_d2	E	typical	CPIRGCIEILSPVCGSDGVNYDNECFLRKAKCTKPELTLVSNTSC
Saprolegnia parasitica	Sp_SPRG_11788_d1	K	typical	CEKACTKDMKPVCGSDGVTYNNECLLQNAQCTNATMTAVPC
Saprolegnia parasitica	Sp_SPRG_11788_d2	K	typical	CAMLCDKMYAPVCGSDNNTYNSECELKNKACNNPTLKVAKKGEC
Saprolegnia parasitica	Sp_SPRG_11788_d3	D	typical	CPKVCNDVLDEVCGSDGKTYNNNCELLKAACAKPSEKLTVVSTGAC
Saprolegnia parasitica	Sp_SPRG_10958_d1	K	typical	CDDASPCPKGGSPVCATNGVTYTNACALAKANCIDANLVLASNGVC
Saprolegnia parasitica	Sp_SPRG_10958_d2	V	typical	CAMECPVSYDPQCGSNGMTYANVCEFKKAHCTNPTVTLDHTGRC
Saprolegnia parasitica	Sp_SPRG_10958_d3	M	typical	CPSICTMEYAPVCGTDGTTYSNGCKLEIARCRGGPKSTLRIAHVGPC
Saprolegnia parasitica	Sp_SPRG_10958_d4	D	typical	CPTACNDKYAPVCGSDGHTYVNACNFEKVHCGNDDMHIVHRGAC
Saprolegnia parasitica	Sp_SPRG_10958_d5	R	typical	CRDRPCNRMFKPVCGSDNKTYNNMCLFENAQCANRGLALLHDGSC
Saprolegnia parasitica	Sp_SPRG_13295_d1	K	typical	CEKACTKDMKPVCGSDGVTYNNECLLQNAQCTNATMTAVPC
Saprolegnia parasitica	Sp_SPRG_13295_d2	D	typical	CPKVCNDVLDEVCGSDGKTYNNNCELLKAACAKPSEKLTVVSTGAC
Saprolegnia parasitica	Sp_SPRG_09559_d1	P	typical	CAVACPPLKQTYCALESPSATTFVGTYSSQCNCLVAKCGNKKVQC
Saprolegnia parasitica	Sp_SPRG_09559_d2	P	typical	CTRKIAKCKPNEVKPVCGSNGVTYDNLCFLKAARCVKPDIQFIAPGKC
Saprolegnia parasitica	Sp_SPRG_09559_d3	L	typical	CGIAKCPLTKTKVCASMDGGKTELKYQNQCFLDAATCVNPLIKKMAAC
Saprolegnia parasitica	Sp_SPRG_09563_d1	S	typical	CAANCTSSKQAVYCALSNSTTAYDVSYSNQCECLAAKCKNRKVHC
Saprolegnia parasitica	Sp_SPRG_09563_d2	P	typical	CLSKLARCKPNQVAPVCGSNGVTYDHLCYLKAARCLNPSIEFLSPGKC
Saprolegnia parasitica	Sp_SPRG_09563_d3	K	typical	CGLGKCSKNKEKVKVCATLKGATIKFKNECLLTAATCVNAAVTPVDC
Saprolegnia parasitica	Sp_SPRG_16956_d1	S	typical	CAANCTSSKQVVYCALSNSTTAYDVSFSNQCECLAAKC
Saprolegnia parasitica	Sp_SPRG_16956_d2	P	typical	CLSKLAKCKPNQVAPVCGSNGATYDHLCYLKAARCLNPSIEFLSPGKC
Saprolegnia parasitica	Sp_SPRG_16956_d3	K	typical	CGLGKCSKTKGKEKVCATLKGATIKFKNECLLTAATCVNAAVTPVDC
Hyaloperonospora parasitica	Hpa_804983_d1	F	typical	CAIRCAFTGDRVCGSNNVTYPNLCLLTLANCANPGEDITVASEGEC
Hyaloperonospora parasitica	Hpa_804983_d2	M	typical	CADACPMIFAPVCGSDSITYGNGCLLGIAHCESKGTITQTSEGQC
Hyaloperonospora parasitica	Hpa_804983_d3	Q	typical	CPDLCVQTYEPVCGSDGVTHNNICMLRAVACYDPSITLAYEGAC
Hyaloperonospora parasitica	Hpa_804983_d4	A	typical	CPDVCLAVFAPVCGSNDVTYGNECELGIFPTVLAQTNSLIQTAKQRLYI
Albugo laibachii	Al_Nc14C621G12264_1_NS	K	typical	CNLIKCDKSPFQRVCGSDGFEYSSRCQAERMQCFEASLSWIDEPC
Albugo laibachii	Al_Nc14C621G12264_2_NS	M	typical	CTECNMYCLDNYDPVCGYSRGIEKSYPNECARKNEICEDPTIKPC
Albugo laibachii	Al_Nc14C188G8390	Q	typical	CNPQPECTQEADPVCGSNYYTYANRCFLANDRCTYPHLSVRADGIC
Albugo laibachii	Al_Nc14C84G5389	Q	typical	CNPQPECTQEADPVCGSNYYTYANRCFLANDRCTYPHLSVRADGIC
Albugo laibachii	Al_Nc14C177G8157_d1	Y	typical	CESICAYDYSGPACGSDGHTYPNKCMITCLDSGTKYFHRGYC
Albugo laibachii	Al_Nc14C177G8157_d2	R	typical	CNVECSRDKELICGIDGQTYINYCHYAVTYCDKRLATLPFLSGEC

Appendix 2.1. List of 140 Kazal-like domains predicted in 64 serine protease inhibitors from seven pathogenic oomycete species

Specie	Kazal-like domain*	P1 residue	Type domain	Domain sequence
Albugo laibachii	Al_Nc14C76G5100_d1_NS	D	typical	CMNECIDSDSNSQLCGTNGITYANLCELKKTGCTGTQIALKHFGVC
Albugo laibachii	Al_Nc14C76G5100_d2_NS	N	typical	CAIAMCSNNVEPVCDLYPSKLTTYQNSCHFRAARCQALHGEKGELLNGPGEENGKLRKREKDAKGKKC
Plasmopara halstedii	Ph_CB174657	R	typical	CTIQCTREYVPVCDSNGQLHANLCLFDVAVCLNPQLTQEKC
Aphanomyces euteiches	Ae_11AL6547_d1	E	typical	CIQSCHEVYQPVCGSDGQTYSNECSLKRESCLKGVKVEMKSPGRC
Aphanomyces euteiches	Ae_11AL6547_d2	E	typical	CPRACIEIFQPVCGTDGNTYANKCTLKQDACARKVSIQVAHEGDC
Aphanomyces euteiches	Ae_11AL6547_d3	K	typical	CPKGCPKIYHPVCGTDGKTYANECTLHLHACENKVDVAVAHDGKC
Aphanomyces euteiches	Ae_11AL6547_d4	E	typical	CRKGCPEIYHAVCGTDGKTYENECTLQRVACENKIDVAVAHDGDC
Aphanomyces euteiches	Ae_11AL6547_d5	E	typical	CPVACIEILRPVCGSDGKTYDNECFLKRDACSKNVHVQVAHEGMC

*NS, Not secreted.

Appendix 2.2. List of 34 cystatin-like domains predicted in 28 cysteine protease inhibitors from seven pathogenic oomycetes species

Specie	Cystatin-like domain*	Domain sequence
Phytophthora infestans	Pi_EPIC1	QVDGGYSKKEVTPEDMELLQKAQSNVSAYNSDVTSRICYLKVDSLETQVVSGENYKFHVSGCSVNSDNELGGCANQNCESSKYDIVIYSQSWTNTLEVTSITPVK
Phytophthora infestans	Pi_EPIC2A	QMNGYTKKEVTPEDMELLQKAQSNVSAYNRDVTSRICYLKVDSLETQVVSGESYKFHVSGCGVNSDKELGGCANQNCESSKYDIVIYSQSWTNTLEVTSITPAN
Phytophthora infestans	Pi_EPIC2B	QLNGYSKKEVTPEDTELLQKAQSNVSAYNSDVTSRICYLKVDSLETQVVSGENYKFHVSGCSVNSDKELGGCANQNCESSKYDIVIYSQSWTNTLKVTSITPAN
Phytophthora infestans	Pi_EPIC3	TILGGYTQKNATSDDIELLTQATSSANMYNKNVDTRICLIAIENLETQTVAGTNYKFQVAGCPVETDDELGACDDRNCDYSSYNIVIFSQPWSDTIEVTSITPAEYQ
Phytophthora infestans	Pi_EPIC4	GMTGSWHPADVTEDNTKLLGTALSGSSFSKSVGDKRVCYSEVTSLETQVVAGTNYRFHISGCDVTDSDGECSTSALSSCELSGFVVQIFEQSWTNTLKVTNIKAEEAA
Phytophthora infestans	Pi_EPIC-LIKE	QMDGGYSKKEVTSEAMELLQKARSNVSAYNSDVTSRICYLKVDSLETQVVSGENYKFHVTGCSVNSDNELGGCTYWNSVPFGSTLDL
Phytophthora infestans	Pi_EPIC5_NS	LQGSDLAVTNILSVRSQVVAGTNYEFEVEGSSASHNDATRFVVKVFDQPWTNTTQLTSLATATAPQ
Phytophthora infestans	Pi_EPIC6	LVGQWMPATKNTATENLLAEALQKKNPSLKSQMCFTEVAAIEQQIVNGIHFRYHVRGCETATPGRCNSGTCATEKKFDVELFVQPWADIVQVMSAVDVQ
Pythium ultimum	Pu_PYU1_T011854	TTFGAWKDEDLTDSVVSTIVDALSNATNYSPTIIKPICALQINSAQSQLVSGTNYKYEVEGCAINFNDELGACRNRDCAKAVYEVVVYSQTWTDTLQVSSITLVE
Pythium ultimum	Pu_PYU1_T012817_NS	MQTGGWAKADVTETNTKILLGAMTGGAGAYGDAVKNTRVCFTKVTDVEQQVVAGMNYRFHIAGCTVSATKLAGDCAAHSETKCVNPKEVFEQNWTSTLQVTAITDAAGK
Pythium ultimum	Pu_PYU1_T012816	AEVGGWTSVPVTANATSLLDKALQNESNYRDTVTARVCVFEVHNLSEQVVAGTNYKYEVQACLVSATVSAGLCAVKTLTTNASCADYTQIFEQVWTNTLEVTSIEKSDSS
Pythium ultimum	Pu_PYU1_T012805_NS	AVTRGWSLVAISNTSMDLLDKTLKNESSYQYADIAMRLCLATTPNEVYQQVVSGVIHEFRGPACQVNTTEEAGACASPPETYAMCAEYAIRIYEQVWTNTTRVMSIELSSGL
Pythium ultimum	Pu_PYU1_T011856_NS	YSMGVWLNATANTPTLAVLDQALRDFPAATTPGGSDLALQLPSLTSPICFQEVVAIEQQIVNGINFRFHVTGCPWLNVVNGEAASARTTGKCVDDCGSSAESYQVTVFCQPWTNTAHVLNLVKEAQR
Pythium ultimum	Pu_PYU1_T012815	AVDTAWKRIEVTEDATTRLETALLNESQYREDVKERVCVEVVERLYELVVANDRKYQYYAYACQVESAAQSGSCTHSRETFYQCAMFDIRIYEQAWTRDVEVQSIEFSHGL
Saprolegnia parasitica	Sp_SPRG_19559	HITGGYGPTKSRVSLDAKTDFFAAVGDDAHYAPETNGVRVCATTFVSVSQQVVAGINYKFRVKGCAVDRAANAVQDCACPTDAPRQTYEISIFVQGWTDTYAVNSITNVTDA
Saprolegnia parasitica	Sp_SPRG_04120	PMLGGWHNTTNVTEGLVETYYSAVASPASYAADATLFVCATSLTSVSAQVVSGMNYIFHVEGCAVHAATDSGADCTCPAPSTAYDVAITDAPWMQMLSVTSITPV
Saprolegnia parasitica	Sp_SPRG_02768_1	GLVGGWSQTSIEDAKPALYGAFQNASSAFVCVTGISSVQKQVVAGINYKFHVVGCPVNNKAKTLESCPATFCPPTDKDQINYEIDVFAPLSSNAFELKAVAMEDAPPE
Saprolegnia parasitica	Sp_SPRG_02768_2	VLVGGWKEGDIDDAADDLYNGLSQETSYKNHNTAHVCVTSIEHVHQQVVSGMNYRFDVLGCQVPNAVAATRGCSCASSSAFKIGLYAQSWTHTYEVLSVESAAPL
Saprolegnia parasitica	Sp_SPRG_02768_3	AGSWKHAEMNSEAKDDFYNALTNDTSHAHVCVSSFLSVASQVVAGTQYRFNVEGCASYLISIYVQPWTRTYEVIHVYEESQLLQLVTQWISANDRNQFGDAKDT
Saprolegnia parasitica	Sp_SPRG_04117	QEGAWSINVKMTNSLVTKYFDVISASSSYLNATSDKICTTTIATVDTQLVSGTNYRYHVSGCAINTVPAANRTCSCANKTVRAYAVSIYEPWINTRFITGVEVEQSA

Appendix 2.2. List of 34 cystatin-like domains predicted in 28 cysteine protease inhibitors from seven pathogenic oomycetes species

Specie	Cystatin-like domain*	Domain sequence
Saprolegnia parasitica	Sp_SPRG_02767	TISGGGFRTTRPASSAASRRQRLRPPTTSRSSSTRHAPRPTT WAPPSPAKQVVAGQNLVYSVKGCALPKASPESSLTSCNTTCA TKDTSSYQVKVFASLMGGFEVSGSLQAKVDG
Saprolegnia parasitica	Sp_SPRG_13039_1	QMGGWAPVTDPSVQTALTAIVSDPANYPNTTTRLCATDVAW ATQQVVAGVQYHVGVHGCATTATSNCSCNGQRHAYMVTVL EAINEPVRITDVVSTVET
Saprolegnia parasitica	Sp_SPRG_13039_2	GLVGGVTAPAPATADDKLLYVRAVTKDANFASASVPRVCPVQ FVSVAKQVVSGIKYIFIVRGCPLVPAGPLNNVFDCACEAPKTY RVEIYEDATRRIQVTKAVVQT
Hyaloperonospora arabidopsidis	Hpa_806306	VIVGGYSTPRTMTLNEVAFLTTTACHPSLYTAGVTSRICFTEFG SIQSQAVSGTNDMFMVKGCPVNRDEHLGYCRDGVCSTTSTY EVIIYSQVWTNTVNVTSVREVNAG
Hyaloperonospora arabidopsidis	Hpa_806307	HDLGLNAYDQARDVTLNEVAFLTTTACHPSLYNADVTSRVCF TEFTTVTTQTSGGGTYYKFQVKGCPVDTEKQLGYCREGACST TSLYEVAIYSQPRTSAVFLTSIKEVV
Hyaloperonospora arabidopsidis	Hpa_801477	KLLGGWQPAEVTDANVKLLNQALSGKRYSTRVGDTRVCYSD VLSVETQVVAGTNYRFRISGCDVTTSDGECLEDTLKDCAPSD FQVVVFEELSTGAPEVTDIQKVAEGGTED
Hyaloperonospora arabidopsidis	Hpa_806312	AKNGSPRPMAVNDVAFLTTTACHPSLYGAGVTNRICFTDFLTV KTLDGGVLNRFQVRGCPVNTEIELGYCRDGCPTTSAYEVVIY SEPWSALSNVPYITEIVQG
Albugo laibachii	Al_Nc14C291G10244_1	EALGGWKEEKVDADSEGRLVSVLSAQTETTAPRICVNKVILV KKQVVAGMNYQYTIEGCDQESKSGMQKCVNCVNRKTYDVV IYERLGENVKELISFEEVKSESKPD
Albugo laibachii	Al_Nc14C291G10244_2	SYTGGNPLFDENKGKAIDYYEYMLGRFPTRPWKEMAVHLQS NVTNGGLQLMAEDKKQSTRTIVLLTSFVVAVLVAMAAMVIFV RLQRNQRRHTYESISDSVHN
Albugo laibachii	Al_Nc14C202G8728_1	EALGGWKEEKVDADSEGRLVSVLSAQTETTAPRICVNKVILV KKQVVAGMNYQYTIEGCDQESKSGMQKCVNCVNRKTYDVV IYERLGENVKELISFEEVKSESKPD
Albugo laibachii	Al_Nc14C202G8728_2	SYTGGNPLFDENKGKAIDYYEYMLGRFPTRPWKEMAVHLQS NVTNGGLQLMAEDKKQSTRTIVLLTSFVVAVLVAMAAMVIFV RLQRNQRRHTYESISDSVHN
Plasmopara halstedii	Ph_CB174713	GMTGSWSPAEITSNATDLLTTALKGDRYDSSVGEKRVCYTEV TSLETQVVAGTNYRFHMDGCEVTNSEGVCSESTLTSCDPSG FVVQIFEQTWTSTLKVTCIKPEESS
Aphanomyces euteiches	Ae_2AL5945_1	SPVGGWSNASLDDAKAAYYEAAALDDSYPTSNTKRVCATTF NSAQQQVVAGINYKISLAGCSVKSVNDTANGCQCASGVDQ YTVIVYKRLQDTPL
Aphanomyces euteiches	Ae_2AL5945_2	LAVGGFSAQRDVTADDKAIFANSTSSDSNYYSAALPRVCATD FVSVSTQVVAGTNYLFTVKGCQLDKADSNSVKDCAATCASK AKTSFQVKIYRDLQQSTK

*NS, Not secreted.

APPENDIX 3: *Phytophthora infestans* tribes enriched in genes that reside in the gene-sparse regions and fast evolving

Appendix 3.1. List of *Phytophthora infestans* genes contained in tribes that are enriched in genes that reside in the gene-sparse regions and fast evolving (See attached CD)

APPENDIX 4: Effector features in the sequenced *Phytophthora infestans* isolate 06_3928A

Appendix 4.1. Features of RXLRs in the sequenced *P. infestans* isolate 06_3928A

Gene ID	Annotation	Secreted	Core ortho	RXLR family	Inter. dist.	Cov	CNV	No. of SNPs	No. of Nonsyn SNPs	No. of Syn SNPs	dN/dS	dN	dS	Induced in potato (dpi) by *P.infestans*		
														06_3928A	T30-4	NL07434
PITG_11947	PexRD33	Yes	Yes	RxLRsng164	GSR	100%	0.44	8	6	2	NA	NA	NA	2 and 3	NA	2 and 3
PITG_23230		Yes	No	RxLRfam9	Not	100%	-0.75	0	0	0	NA	NA	NA	2 and 3	NA	2
PITG_14783		Yes	No	RxLRfam6	GSR	100%	3.46	6	6	0	NA	0.02	0.00	2 and 3	2	NA
PITG_22798		Yes	No	RxLRsng157	GSR	100%	-0.33	5	5	0	NA	0.01	0.00	2	2	NA
PITG_14787		Yes	No	RxLRfam6	GSR	100%	4.89	5	5	0	NA	0.02	0.00	2 and 3	2	NA
PITG_12731		Yes	No	RxLRfam1	GSR	100%	-0.15	3	3	0	NA	0.00	0.00	2 and 3	3	2
PITG_14788		Yes	No	RxLRfam8	GSR	100%	0.10	1	1	0	NA	0.00	0.00	2	2	NA
PITG_17309		Yes	No	RxLRfam1	InBtw	100%	-0.31	1	1	0	NA	0.00	0.00	2 and 3	NA	2
PITG_15255		Yes	No	RxLRfam4	GSR	100%	-0.59	1	1	0	NA	0.00	0.00	2	NA	NA
PITG_16195		Yes	No	RxLRfam1	GSR	100%	-0.12	1	0	1	NA	0.00	0.00	2	2 and 3	2
PITG_15039		Yes	No	RxLRfam1	GSR	100%	-0.22	1	1	0	NA	0.00	0.00	2 and 3	2 and 3	2
PITG_14984	PexRD42	Yes	No	RxLRfam6	InBtw	100%	0.19	1	1	0	NA	0.00	0.00	2	2 and 3	NA
PITG_22804		Yes	No	RxLRfam27	GSR	100%	-0.48	0	0	0	NA	0.00	0.00	2 and 3	2	2
PITG_22757		Yes	No	RxLRsng203	GSR	100%	-0.02	0	0	0	NA	0.00	0.00	2 and 3	2	NA
PITG_22724		Yes	No	RxLRfam67	GSR	100%	1.29	0	0	0	NA	0.00	0.00	2	2	NA
PITG_22648		Yes	No	NA	Not	100%	0.55	0	0	0	NA	0.00	0.00	2 and 3	2	2
PITG_22256		Yes	No	RxLRsng187	Not	57%	NA	0	0	0	NA	0.00	0.00	2	2	NA
PITG_21778		Yes	No	RxLRfam6	Not	100%	0.97	0	0	0	NA	0.00	0.00	2	2	NA
PITG_21388	*Avrblb1*, PexRD6	Yes	No	RxLRfam54	Not	100%	0.09	0	0	0	NA	0.00	0.00	2 and 3	2	2
PITG_19942		Yes	No	RxLRsng237	GSR	100%	-0.18	0	0	0	NA	0.00	0.00	2	2	NA
PITG_18609		Yes	No	RxLRfam26	GSR	98%	-0.69	0	0	0	NA	0.00	0.00	2	2	NA
PITG_14368	Pex147-2	Yes	No	RxLRfam58	GSR	100%	-0.06	0	0	0	NA	0.00	0.00	2 and 3	2	2
PITG_13093		Yes	No	RxLRfam38	InBtw	100%	-0.11	0	0	0	NA	0.00	0.00	2 and 3	2	2
PITG_10232		Yes	No	RxLRfam69	GSR	100%	1.04	0	0	0	NA	0.00	0.00	2 and 3	2	2
PITG_08174		Yes	No	RxLRfam19	InBtw	100%	-0.09	0	0	0	NA	0.00	0.00	2 and 3	2	NA
PITG_07594		Yes	No	RxLRfam26	GSR	100%	0.08	0	0	0	NA	0.00	0.00	2	2	NA
PITG_06099	PexRD50	Yes	No	RxLRfam36	GSR	100%	-0.06	0	0	0	NA	0.00	0.00	2 and 3	2	2
PITG_05750	PexRD49	Yes	No	RxLRfam29	InBtw	100%	-0.03	0	0	0	NA	0.00	0.00	2 and 3	2	2
PITG_04314	PexRD24	Yes	No	RxLRfam49	GSR	100%	-0.30	0	0	0	NA	0.00	0.00	2 and 3	2	2
PITG_04196		Yes	No	RxLRfam47	GSR	100%	-0.14	0	0	0	NA	0.00	0.00	2	2	NA
PITG_01934		Yes	No	RxLRfam6	GSR	100%	-0.03	0	0	0	NA	0.00	0.00	2 and 3	2	NA
PITG_00774		Yes	No	RxLRsng199	GSR	100%	-0.19	0	0	0	NA	0.00	0.00	2 and 3	2	NA
PITG_00582		Yes	No	RxLRsng212	GSR	100%	0.10	0	0	0	NA	0.00	0.00	2 and 3	2	2
PITG_23206	PexRD10	Yes	No	RxLRsng192	GDR	100%	-0.52	0	0	0	NA	0.00	0.00	2	NA	NA
PITG_23193		Yes	No	RxLRfam5	InBtw	94%	-0.90	0	0	0	NA	0.00	0.00	2	NA	NA
PITG_23131		Yes	No	RxLRfam128	GSR	100%	-0.13	0	0	0	NA	0.00	0.00	2 and 3	NA	2
PITG_22926		Yes	No	RxLRfam52	GSR	99%	-0.67	0	0	0	NA	0.00	0.00	2 and 3	NA	NA
PITG_22675		Yes	Yes	RxLRfam73	InBtw	100%	-0.11	0	0	0	NA	0.00	0.00	2	NA	NA
PITG_21190		Yes	No	RxLRfam2	Not	100%	-0.04	0	0	0	NA	0.00	0.00	2 and 3	NA	NA
PITG_20336		Yes	No	RxLRfam9	Not	96%	-0.89	0	0	0	NA	0.00	0.00	2 and 3	NA	2
PITG_16427		Yes	No	RxLRfam9	InBtw	100%	-0.65	0	0	0	NA	0.00	0.00	2 and 3	NA	2
PITG_16233	PexRD12 paralog	Yes	No	RxLRfam9	InBtw	100%	-0.54	0	0	0	NA	0.00	0.00	2 and 3	NA	NA
PITG_15930		Yes	No	RxLRfam2	InBtw	100%	-0.41	0	0	0	NA	0.00	0.00	2 and 3	NA	2
PITG_15753		Yes	No	RxLRfam38	GSR	100%	-0.01	0	0	0	NA	0.00	0.00	2 and 3	NA	2
PITG_15679		Yes	No	RxLRfam23	GSR	100%	0.20	0	0	0	NA	0.00	0.00	2 and 3	NA	2
PITG_14360		Yes	No	RxLRfam72	GSR	100%	0.14	0	0	0	NA	0.00	0.00	2	NA	2
PITG_09739		Yes	No	RxLRfam6	GSR	100%	-0.67	0	0	0	NA	0.00	0.00	2	NA	NA
PITG_09660		Yes	No	NA	GSR	100%	0.14	0	0	0	NA	0.00	0.00	2	NA	2

Appendix 4.1. Features of RXLRs in the sequenced *P. infestans* isolate 06_3928A

Gene ID	Annotation	Secreted	Core ortho	RXLR family	Inter. dist.	Cov	CNV	No. of SNPs	No. of Nonsyn SNPs	No. of Syn SNPs	dN/dS	dN	dS	Induced in potato (dpi) by *P.infestans*		
														06_3928A	T30-4	NL07434
PITG_07630		Yes	No	RxLRfam1	GSR	100%	-0.15	0	0	0	NA	0.00	0.00	2	NA	2
PITG_07587		Yes	No	RxLRfam26	InBtw	100%	-0.28	0	0	0	NA	0.00	0.00	2	NA	NA
PITG_06094		Yes	No	RxLRfam36	GSR	100%	-0.05	0	0	0	NA	0.00	0.00	2 and 3	NA	NA
PITG_05912	homolog of *PsAvr*	Yes	No	RxLRfam18	GSR	83%	-0.20	0	0	0	NA	0.00	0.00	2 and 3	NA	NA
PITG_05911	homolog of *PsAvr*	Yes	No	RxLRfam18	InBtw	83%	0.35	0	0	0	NA	0.00	0.00	2 and 3	NA	NA
PITG_05910		Yes	No	RxLRfam52	InBtw	100%	0.24	0	0	0	NA	0.00	0.00	2 and 3	NA	NA
PITG_05846		Yes	No	RxLRfam23	GSR	100%	-0.09	0	0	0	NA	0.00	0.00	2 and 3	NA	2
PITG_04339	PexRD20	Yes	No	RxLRfam81	InBtw	100%	-0.09	0	0	0	NA	0.00	0.00	2	NA	NA
PITG_02779		Yes	No	RxLRfam80	GSR	100%	0.05	0	0	0	NA	0.00	0.00	2	NA	NA
PITG_23226		Yes	No	RxLRfam100	Not	94%	-0.58	0	0	0	NA	0.00	0.00	2 and 3	2 and 3	2
PITG_23015		Yes	No	RxLRfam100	GSR	100%	0.29	0	0	0	NA	0.00	0.00	2 and 3	2 and 3	2
PITG_22922		Yes	Yes	RxLRfam2	InBtw	100%	-0.02	0	0	0	NA	0.00	0.00	2 and 3	2 and 3	2 and 3
PITG_22547		Yes	No	RxLRfam97	Not	99%	-0.45	0	0	0	NA	0.00	0.00	2 and 3	2 and 3	2
PITG_21740		Yes	No	RxLRfam1	Not	100%	-0.53	0	0	0	NA	0.00	0.00	2 and 3	2 and 3	2
PITG_20303	*Avrblb2* paralog	Yes	No	RxLRfam5	Not	78%	-0.88	0	0	0	NA	0.00	0.00	2 and 3	2 and 3	2
PITG_20301	*Avrblb2* paralog	Yes	No	RxLRfam5	GSR	70%	-0.82	0	0	0	NA	0.00	0.00	2 and 3	2 and 3	2 and 3
PITG_20300	*Avrblb2*, PexRD39	Yes	No	RxLRfam5	GSR	100%	-0.76	0	0	0	NA	0.00	0.00	2 and 3	2 and 3	2 and 3
PITG_18683		Yes	No	RxLRfam5	GSR	77%	-0.83	0	0	0	NA	0.00	0.00	2 and 3	2 and 3	2
PITG_18670		Yes	No	RxLRfam5	InBtw	93%	-0.81	0	0	0	NA	0.00	0.00	2 and 3	2 and 3	2
PITG_17063	PexRD44	Yes	No	RxLRfam45	Not	100%	-0.32	0	0	0	NA	0.00	0.00	2 and 3	2 and 3	2
PITG_16705		Yes	No	RxLRfam1	GSR	100%	-0.05	0	0	0	NA	0.00	0.00	2 and 3	2 and 3	2 and 3
PITG_16294	*Avrvnt1*	Yes	No	RxLRfam97	GSR	100%	0.50	0	0	0	NA	0.00	0.00	2 and 3	2 and 3	2
PITG_14983		Yes	No	RxLRfam6	GSR	100%	0.09	0	0	0	NA	0.00	0.00	2	2 and 3	NA
PITG_14371	*Avr3a*, PexRD7	Yes	No	RxLRfam58	GSR	100%	-0.26	0	0	0	NA	0.00	0.00	2 and 3	2 and 3	2
PITG_12737		Yes	No	RxLRfam43	GSR	100%	-0.03	0	0	0	NA	0.00	0.00	2 and 3	2 and 3	2
PITG_10654		Yes	No	RxLRfam46	GSR	100%	-0.21	0	0	0	NA	0.00	0.00	2 and 3	2 and 3	2 and 3
PITG_09732		Yes	No	RxLRfam1	GSR	100%	-0.24	0	0	0	NA	0.00	0.00	2 and 3	2 and 3	2
PITG_09216		Yes	No	RxLRfam55	GSR	100%	0.43	0	0	0	NA	0.00	0.00	2 and 3	2 and 3	2 and 3
PITG_07555		Yes	No	RxLRsng247	GSR	100%	-0.13	0	0	0	NA	0.00	0.00	2 and 3	2 and 3	NA
PITG_07550		Yes	No	RxLRfam117	GSR	100%	-0.36	0	0	0	NA	0.00	0.00	2 and 3	2 and 3	2
PITG_06478	PexRD18	Yes	No	RxLRfam16	GSR	100%	-0.07	0	0	0	NA	0.00	0.00	2 and 3	2 and 3	2 and 3
PITG_06087	PexRD16	Yes	Yes	RxLRfam87	InBtw	100%	-0.20	0	0	0	NA	0.00	0.00	2 and 3	2 and 3	2 and 3
PITG_04266		Yes	No	RxLRsng248	InBtw	100%	-0.23	0	0	0	NA	0.00	0.00	2 and 3	2 and 3	2 and 3
PITG_04090	*Avrblb2* paralog	Yes	No	RxLRfam5	GSR	100%	0.20	0	0	0	NA	0.00	0.00	2 and 3	2 and 3	2
PITG_04085	*Avrblb2* paralog	Yes	No	RxLRfam5	InBtw	100%	0.33	0	0	0	NA	0.00	0.00	2 and 3	2 and 3	2 and 3
PITG_02860		Yes	No	RxLRfam80	GSR	100%	-0.24	0	0	0	NA	0.00	0.00	2 and 3	2 and 3	2 and 3
PITG_04049		Yes	No	RxLRfam67	InBtw	100%	-0.37	3	2	1	0.79	0.01	0.01	2 and 3	2 and 3	NA
PITG_15278		Yes	No	RxLRfam1	InBtw	100%	0.17	2	1	1	0.43	0.00	0.00	2 and 3	2 and 3	2 and 3
PITG_22089		Yes	No	RxLRfam18	Not	100%	0.19	4	2	2	0.38	0.01	0.02	2 and 3	NA	NA
PITG_23239		Yes	No	RxLRfam67	Not	96%	-0.60	2	1	1	0.38	0.00	0.01	2	2	NA
PITG_22870	*Avr2*	Yes	No	RxLRfam7	GSR	81%	-0.50	4	2	2	0.34	0.01	0.03	2 and 3	2 and 3	2
PITG_21362		Yes	No	RxLRfam57	GSR	93%	-0.47	3	1	2	0.21	0.00	0.01	2 and 3	NA	2
PITG_10540	PexRD5	Yes	No	RxLRfam57	InBtw	100%	0.05	3	1	2	0.21	0.00	0.01	2 and 3	2 and 3	2
PITG_16844		Yes	No	RxLRfam1	GSR	99%	-0.71	3	1	2	0.19	0.00	0.01	2	NA	NA
PITG_23035		Yes	Yes	RxLRfam1	InBtw	100%	-0.21	9	2	7	0.08	0.00	0.01	2 and 3	NA	2
PITG_14443		Yes	No	RxLRfam69	InBtw	59%	-0.49	2	0	2	0.00	0.00	0.02	2 and 3	2	2
PITG_15110		Yes	No	RxLRfam1	InBtw	100%	-0.02	2	0	2	0.00	0.00	0.00	2 and 3	2 and 3	2
PITG_22604		Yes	No	RxLRfam5	Not	100%	-0.02	1	0	1	0.00	0.00	0.01	2 and 3	2	2
PITG_04089	PexRD41	Yes	No	RxLRfam5	GSR	94%	NA	1	0	1	0.00	0.00	0.01	2 and 3	2	2

Appendix 4.1. Features of RXLRs in the sequenced *P. infestans* isolate 06_3928A

Gene ID	Annotation	Secreted	Core ortho	RXLR family	Inter. dist.	Cov	CNV	No. of SNPs	No. of Nonsyn SNPs	No. of Syn SNPs	dN/dS	dN	dS	Induced in potato (dpi) by *P.infestans*		
														06_3928A	T30-4	NL07434
PITG_17316		Yes	No	RxLRfam1	InBtw	100%	-0.39	1	0	1	0.00	0.00	0.00	2 and 3	NA	2
PITG_13612		Yes	No	RxLRfam6	GSR	28%	-0.27	1	0	1	0.00	0.00	0.03	2	NA	NA
PITG_22740		Yes	No	RxLRfam1	InBtw	100%	-0.17	12	7	5	NA	NA	NA	NA	NA	NA
PITG_02918		Yes	No	RxLRfam112	GSR	100%	0.21	7	5	2	NA	NA	NA	NA	NA	NA
PITG_23000		Yes	No	RxLRsng171	InBtw	100%	0.20	6	5	1	NA	NA	NA	NA	NA	NA
PITG_09109		Yes	No	RxLRfam1	GSR	100%	0.52	5	5	0	NA	0.01	0.00	NA	NA	NA
PITG_22929		Yes	No	RxLRsng221	GSR	100%	-0.13	4	4	0	NA	0.02	0.00	NA	NA	NA
PITG_11836		Yes	No	NA	GSR	100%	0.93	4	4	0	NA	0.02	0.00	NA	NA	NA
PITG_22978		Yes	No	RxLRsng233	GSR	100%	-0.08	3	3	0	NA	0.01	0.00	NA	NA	NA
PITG_18318		Yes	No	RxLRfam17	GSR	100%	0.10	3	3	0	NA	0.01	0.00	NA	NA	NA
PITG_12721		Yes	No	RxLRfam4	InBtw	47%	0.15	3	3	0	NA	0.03	0.00	NA	NA	NA
PITG_10465		Yes	No	NA	GSR	100%	0.52	3	3	0	NA	0.01	0.00	NA	NA	NA
PITG_07947	PexRD26	Yes	No	RxLRfam38	GSR	100%	-0.03	3	3	0	NA	0.01	0.00	NA	NA	NA
PITG_04203		Yes	No	RxLRfam48	InBtw	100%	-0.34	3	3	0	NA	0.01	0.00	NA	NA	NA
PITG_22375		Yes	No	RxLRfam10	Not	94%	-0.63	2	2	0	NA	0.01	0.00	NA	NA	NA
PITG_15728		Yes	No	RxLRfam23	GSR	100%	2.03	2	2	0	NA	0.00	0.00	NA	NA	NA
PITG_15177		Yes	No	RxLRfam95	GSR	90%	-0.51	2	2	0	NA	0.01	0.00	NA	NA	NA
PITG_14662		Yes	No	RxLRfam1	GSR	99%	-0.32	2	2	0	NA	0.01	0.00	NA	NA	NA
PITG_14432		Yes	No	RxLRfam13	GSR	100%	0.42	2	2	0	NA	0.01	0.00	NA	NA	NA
PITG_12952		Yes	No	RxLRfam46	GSR	100%	0.11	2	2	0	NA	0.01	0.00	NA	NA	NA
PITG_12816		Yes	No	RxLRfam43	GSR	100%	-0.42	2	2	0	NA	0.01	0.00	NA	NA	NA
PITG_10396		Yes	No	RxLRfam10	InBtw	100%	0.10	2	2	0	NA	0.01	0.00	NA	NA	NA
PITG_07435		Yes	No	RxLRfam52	GDR	100%	-0.34	2	2	0	NA	0.01	0.00	NA	NA	NA
PITG_22986		Yes	No	RxLRfam99	GSR	100%	-0.18	1	1	0	NA	0.00	0.00	NA	NA	NA
PITG_22880		Yes	No	RxLRfam1	GSR	100%	0.60	1	1	0	NA	0.00	0.00	NA	NA	NA
PITG_22871		Yes	No	RxLRfam21	GSR	100%	-0.07	1	1	0	NA	0.00	0.00	NA	NA	NA
PITG_22816		Yes	No	RxLRsng178	GDR	100%	-0.12	1	1	0	NA	0.01	0.00	NA	NA	NA
PITG_19800		Yes	No	RxLRfam50	Not	100%	1.86	1	1	0	NA	0.00	0.00	NA	NA	NA
PITG_18510		Yes	No	RxLRfam45	InBtw	100%	-0.19	1	1	0	NA	0.00	0.00	NA	NA	NA
PITG_18325		Yes	No	RxLRfam17	GSR	100%	0.22	1	1	0	NA	0.00	0.00	NA	NA	NA
PITG_18163		Yes	No	NA	InBtw	100%	-0.27	1	1	0	NA	0.00	0.00	NA	NA	NA
PITG_17218		Yes	No	RxLRfam1	GSR	100%	-0.28	1	1	0	NA	0.00	0.00	NA	NA	NA
PITG_16738		Yes	No	RxLRfam8	GSR	100%	-0.25	1	1	0	NA	0.00	0.00	NA	NA	NA
PITG_15337		Yes	No	RxLRfam24	GSR	100%	0.47	1	1	0	NA	0.00	0.00	NA	NA	NA
PITG_15318		Yes	No	RxLRfam59	InBtw	100%	-0.46	1	1	0	NA	0.00	0.00	NA	NA	NA
PITG_15297		Yes	No	RxLRfam59	GSR	100%	-0.45	1	1	0	NA	0.00	0.00	NA	NA	NA
PITG_13940		Yes	No	RxLRfam32	InBtw	100%	-0.06	1	1	0	NA	0.00	0.00	NA	NA	NA
PITG_13628	PexRD27	Yes	No	RxLRfam6	GSR	100%	0.05	1	1	0	NA	0.00	0.00	NA	NA	NA
PITG_13119		Yes	No	RxLRfam16	GSR	100%	-0.08	1	1	0	NA	0.00	0.00	NA	NA	NA
PITG_12791		Yes	No	RxLRfam1	InBtw	100%	-0.18	1	1	0	NA	0.00	0.00	NA	NA	NA
PITG_10348		Yes	No	RxLRfam93	GSR	100%	-0.38	1	1	0	NA	0.00	0.00	NA	NA	NA
PITG_09586		Yes	No	RxLRfam2	InBtw	100%	-0.18	1	1	0	NA	0.00	0.00	NA	NA	NA
PITG_08903		Yes	No	RxLRfam54	GSR	100%	-0.23	1	1	0	NA	0.00	0.00	NA	NA	NA
PITG_05918		Yes	No	RxLRfam18	GSR	100%	0.30	1	1	0	NA	0.00	0.00	NA	NA	NA
PITG_05841		Yes	No	RxLRfam23	GSR	100%	-0.19	1	1	0	NA	0.00	0.00	NA	NA	NA
PITG_04052		Yes	No	RxLRfam1	InBtw	100%	-0.48	1	1	0	NA	0.00	0.00	NA	NA	NA
PITG_07387	*Avr4*	Yes	No	RxLRfam52	GSR	89%	-0.25	1	1	0	NA	0.00	0.00	NA	2 and 3	NA
PITG_22727		Yes	No	RxLRfam5	GSR	100%	0.59	0	0	0	NA	0.00	0.00	NA	2	NA
PITG_19617		Yes	No	RxLRfam7	GSR	100%	-0.14	0	0	0	NA	0.00	0.00	NA	2	NA

Appendix 4.1. Features of RXLRs in the sequenced *P. infestans* isolate 06_3928A

Gene ID	Annotation	Secreted	Core ortho	RXLR family	Inter. dist.	Cov	CNV	No. of SNPs	No. of Nonsyn SNPs	No. of Syn SNPs	dN/dS	dN	dS	Induced in potato (dpi) by *P.infestans*		
														06_3928A	T30-4	NL07434
PITG_16726		Yes	No	RxLRfam1	InBtw	100%	0.02	0	0	0	NA	0.00	0.00	NA	2	NA
PITG_16282		Yes	No	RxLRfam18	GSR	100%	1.29	0	0	0	NA	0.00	0.00	NA	2	NA
PITG_11507		Yes	No	RxLRfam120	GSR	100%	-0.03	0	0	0	NA	0.00	0.00	NA	2	NA
PITG_04097		Yes	No	RxLRfam5	GSR	100%	0.78	0	0	0	NA	0.00	0.00	NA	2	NA
PITG_23216		Yes	No	RxLRfam93	Not	100%	0.81	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_23215		Yes	No	RxLRfam125	Not	100%	-0.09	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_23185		Yes	No	RxLRfam5	GSR	82%	-0.85	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_23154		Yes	No	RxLRsng155	GSR	100%	-0.15	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_23135		Yes	No	RxLRfam5	GSR	100%	-0.80	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_23132	PexRD36	Yes	No	RxLRfam88	InBtw	100%	0.24	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_23125		Yes	No	RxLRfam28	InBtw	100%	-0.02	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_23120		Yes	No	RxLRfam39	InBtw	100%	0.12	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_23074		Yes	No	RxLRfam9	InBtw	100%	-0.55	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_23069		Yes	No	RxLRfam9	InBtw	100%	-0.49	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_23061		Yes	No	RxLRfam16	GSR	100%	-0.14	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_23036		Yes	No	RxLRfam1	InBtw	100%	-0.19	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_23026		Yes	No	RxLRsng242	InBtw	98%	0.13	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_23016		Yes	No	RxLRfam58	GSR	100%	0.46	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_23008		Yes	No	RxLRfam32	GSR	100%	0.76	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_22999		Yes	No	RxLRfam126	GSR	100%	0.07	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_22998		Yes	No	RxLRfam126	GSR	100%	0.21	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_22990		Yes	Yes	RxLRfam34	GSR	100%	-0.33	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_22987		Yes	No	RxLRfam99	GSR	100%	0.34	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_22935		Yes	No	RxLRfam6	InBtw	100%	-0.50	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_22933		Yes	No	RxLRfam98	GSR	100%	1.16	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_22932		Yes	No	RxLRsng170	InBtw	100%	-0.31	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_22925		Yes	No	RxLRsng191	GSR	100%	-0.25	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_22900		Yes	No	RxLRfam91	GSR	100%	-0.14	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_22896		Yes	No	RxLRfam56	GSR	100%	-0.56	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_22894		Yes	No	RxLRfam56	InBtw	100%	0.32	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_22891		Yes	No	RxLRsng241	InBtw	100%	-0.07	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_22890		Yes	No	RxLRfam20	GSR	100%	0.16	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_22879		Yes	No	RxLRfam1	GSR	100%	0.94	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_22853		Yes	No	RxLRfam49	GSR	100%	-0.37	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_22844		Yes	No	RxLRfam95	GDR	100%	-0.32	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_22813		Yes	No	RxLRsng240	GSR	100%	0.00	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_22802		Yes	No	RxLRsng222	GSR	100%	-0.20	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_22766		Yes	No	RxLRsng235	GSR	100%	0.08	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_22730		Yes	No	RxLRfam43	InBtw	100%	-0.20	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_22729		Yes	No	RxLRfam43	InBtw	100%	-0.26	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_22725		Yes	No	RxLRfam5	GSR	100%	0.37	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_22712		Yes	No	RxLRsng163	Not	99%	-0.33	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_22683		Yes	No	RxLRsng209	InBtw	100%	2.34	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_22676		Yes	No	RxLRfam125	InBtw	100%	-0.38	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_22118		Yes	No	RxLRfam1	Not	100%	0.09	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_21949		Yes	No	RxLRfam32	Not	100%	0.04	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_21739		Yes	No	RxLRfam84	Not	100%	-0.39	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_21422		Yes	No	RxLRfam6	Not	100%	0.02	0	0	0	NA	0.00	0.00	NA	NA	2
PITG_21303		Yes	No	RxLRfam40	Not	29%	-0.89	0	0	0	NA	0.00	0.00	NA	NA	NA

Appendix 4.1. Features of RXLRs in the sequenced *P. infestans* isolate 06_3928A

Gene ID	Annotation	Secreted	Core ortho	RXLR family	Inter. dist.	Cov	CNV	No. of SNPs	No. of Nonsyn SNPs	No. of Syn SNPs	dN/dS	dN	dS	Induced in potato (dpi) by *P.infestans*		
														06_3928A	T30-4	NL07434
PITG_21288		Yes	No	RxLRfam1	Not	100%	0.31	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_21238		Yes	No	RxLRfam66	Not	100%	0.08	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_21107		Yes	No	RxLRfam3	Not	100%	1.22	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_20972		Yes	No	RxLRfam109	Not	100%	0.05	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_20934		Yes	No	RxLRfam9	GSR	100%	-0.43	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_20365		Yes	No	RxLRfam39	GSR	100%	-0.75	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_20144	PexRD2	Yes	No	RxLRfam95	GSR	100%	-0.28	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_20052		Yes	No	RxLRfam1	GSR	100%	0.24	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_19998		Yes	No	RxLRfam2	GSR	100%	-0.04	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_19996		Yes	No	RxLRfam2	InBtw	100%	0.21	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_19994		Yes	No	RxLRfam1	GSR	100%	0.02	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_19992		Yes	No	RxLRfam1	GSR	100%	-0.15	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_19831		Yes	No	RxLRfam40	GSR	100%	1.02	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_19808		Yes	No	NA	GSR	100%	0.15	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_19803		Yes	No	RxLRfam37	GSR	100%	-0.54	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_19655		Yes	No	RxLRfam1	GSR	100%	-0.12	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_19528		Yes	No	RxLRfam25	GSR	100%	0.56	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_19523		Yes	No	RxLRfam1	GSR	100%	0.05	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_19309		Yes	No	RxLRfam1	GSR	100%	-0.27	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_19308		Yes	No	RxLRfam1	GSR	100%	0.16	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_19307		Yes	No	RxLRfam1	InBtw	100%	0.20	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_19302		Yes	No	RxLRfam1	GSR	100%	-0.27	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_19232		Yes	No	RxLRfam1	GSR	100%	0.10	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_18986		Yes	No	RxLRfam4	GSR	100%	-0.16	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_18981		Yes	No	RxLRfam10	GSR	100%	-0.30	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_18956		Yes	No	RxLRfam4	Not	100%	0.97	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_18908		Yes	No	RxLRfam54	GSR	100%	-0.33	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_18880		Yes	No	RxLRfam97	GSR	100%	-0.19	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_18487		Yes	No	RxLRfam45	InBtw	100%	0.05	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_18405		Yes	No	RxLRfam27	GSR	100%	0.62	0	0	0	NA	0.00	0.00	3	NA	2
PITG_18153		Yes	No	RxLRfam39	GSR	100%	0.20	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_17871		Yes	No	RxLRfam1	GSR	100%	-0.03	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_17838	PexRD8 paralog	Yes	No	RxLRfam3	GSR	100%	-0.58	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_17670		Yes	No	RxLRfam15	GSR	100%	0.43	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_17310		Yes	No	RxLRfam58	InBtw	100%	-0.07	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_17214		Yes	No	RxLRfam45	InBtw	99%	-0.07	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_16836		Yes	No	RxLRfam117	GSR	100%	0.95	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_16737	PexRD15	Yes	No	RxLRfam8	GSR	100%	0.37	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_16708		Yes	No	RxLRfam1	GSR	100%	-0.08	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_16541		Yes	No	RxLRfam115	GSR	100%	0.60	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_16529		Yes	No	RxLRfam38	GSR	100%	-0.10	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_16515		Yes	No	RxLRfam38	GSR	100%	-0.17	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_16428		Yes	No	RxLRfam9	GSR	100%	-0.12	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_16402	PexRD31	Yes	No	RxLRfam9	GSR	100%	-0.27	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_16283		Yes	No	RxLRfam1	GSR	100%	1.61	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_16248		Yes	No	RxLRfam9	GSR	100%	-0.36	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_16243		Yes	No	RxLRfam9	GSR	57%	-0.86	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_16193		Yes	No	RxLRfam1	GSR	100%	-0.11	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_16188		Yes	No	RxLRfam82	GSR	100%	-0.13	0	0	0	NA	0.00	0.00	NA	NA	NA

Appendix 4.1. Features of RXLRs in the sequenced *P. infestans* isolate 06_3928A

Gene ID	Annotation	Secreted	Core ortho	RXLR family	Inter. dist.	Cov	CNV	No. of SNPs	No. of Nonsyn SNPs	No. of Syn SNPs	dN/dS	dN	dS	Induced in potato (dpi) by *P.infestans*		
														06_3928A	T30-4	NL07434
PITG_16180		Yes	No	RxLRfam4	InBtw	29%	-0.89	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_15940		Yes	No	RxLRfam2	GSR	100%	-0.22	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_15764		Yes	No	RxLRfam16	GSR	100%	0.02	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_15763		Yes	No	RxLRfam16	InBtw	100%	-0.01	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_15757		Yes	No	RxLRfam38	GSR	100%	0.03	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_15556		Yes	No	RxLRfam10	GSR	100%	-0.26	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_15424		Yes	No	RxLRfam8	InBtw	100%	0.04	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_15341		Yes	No	RxLRfam6	GSR	100%	0.01	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_15304		Yes	No	RxLRfam17	GSR	100%	0.02	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_15303		Yes	No	RxLRfam17	GSR	100%	0.70	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_15166		Yes	No	RxLRfam43	GSR	100%	-0.05	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_15162		Yes	No	RxLRfam43	GSR	100%	-0.39	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_15105		Yes	No	RxLRfam1	InBtw	100%	-0.14	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_15086		Yes	No	RxLRfam88	GSR	100%	0.14	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_15038		Yes	No	RxLRfam1	GSR	100%	-0.22	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_15032		Yes	No	RxLRfam1	InBtw	100%	-0.16	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_14986		Yes	No	RxLRfam6	GSR	100%	0.09	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_14960		Yes	No	RxLRfam21	GSR	99%	-0.74	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_14955		Yes	No	RxLRfam21	InBtw	100%	-0.76	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_14932		Yes	No	RxLRfam21	GSR	100%	-0.76	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_14738		Yes	No	RxLRfam3	GSR	100%	-0.63	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_14737	PexRD8 paralog	Yes	No	RxLRfam3	GSR	100%	-0.44	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_14736	PexRD8 paralog	Yes	No	RxLRfam3	GSR	100%	-0.29	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_14732		Yes	No	RxLRfam3	GSR	100%	0.19	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_14434		Yes	No	RxLRfam13	InBtw	91%	-0.46	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_14374	Pex147-3	Yes	No	RxLRfam58	GSR	100%	-0.10	0	0	0	NA	0.00	0.00	3	NA	2
PITG_14093		Yes	Yes	RxLRfam71	GSR	100%	0.55	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_14086		Yes	No	RxLRfam94	InBtw	100%	0.03	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_14062		Yes	No	NA	InBtw	100%	0.11	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_14046		Yes	No	RxLRfam69	InBtw	100%	-0.36	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_13959		Yes	No	RxLRfam3	GSR	100%	0.93	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_13956		Yes	No	RxLRfam32	GSR	100%	-0.24	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_13936		Yes	No	RxLRfam32	InBtw	100%	0.31	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_13930	PexRD11	Yes	No	RxLRfam32	GSR	100%	0.12	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_13593		Yes	No	RxLRfam18	GSR	100%	0.41	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_13550		Yes	No	RxLRfam4	GSR	100%	0.28	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_13538		Yes	No	RxLRfam50	GSR	100%	0.32	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_13536		Yes	No	RxLRfam37	GSR	100%	0.58	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_13529		Yes	No	RxLRfam50	GSR	100%	3.75	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_13481		Yes	No	RxLRfam3	GSR	100%	0.24	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_13452	PexRD21	Yes	No	RxLRfam108	InBtw	100%	0.10	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_13306	PexRD22	Yes	No	RxLRfam122	InBtw	100%	-0.04	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_13072		Yes	No	RxLRfam44	InBtw	100%	-0.07	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_13018		Yes	No	RxLRfam1	InBtw	100%	0.00	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_12851		Yes	No	RxLRfam91	GSR	100%	0.25	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_12719		Yes	No	RxLRfam36	GSR	100%	0.40	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_12423		Yes	No	RxLRfam121	InBtw	100%	0.19	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_12276		Yes	Yes	RxLRfam70	InBtw	100%	-0.42	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_11839		Yes	No	RxLRfam70	GSR	100%	0.85	0	0	0	NA	0.00	0.00	NA	NA	NA

Appendix 4.1. Features of RXLRs in the sequenced *P. infestans* isolate 06_3928A

Gene ID	Annotation	Secreted	Core ortho	RXLR family	Inter. dist.	Cov	CNV	No. of SNPs	No. of Nonsyn SNPs	No. of Syn SNPs	dN/dS	dN	dS	Induced in potato (dpi) by *P.infestans*		
														06_3928A	T30-4	NL07434
PITG_11429		Yes	No	RxLRfam54	GSR	100%	-0.02	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_11384	PexRD2	Yes	No	RxLRfam6	GSR	100%	0.01	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_11383		Yes	No	RxLRfam6	GSR	100%	-0.35	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_11350		Yes	No	RxLRfam6	GSR	100%	-0.27	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_11344		Yes	No	RxLRfam24	GSR	100%	0.32	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_10818		Yes	No	RxLRfam31	GSR	100%	-0.24	0	0	0	NA	0.00	0.00	NA	NA	2
PITG_10673		Yes	No	RxLRsng165	InBtw	100%	-0.11	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_10640		Yes	No	RxLRfam27	GSR	98%	-0.45	0	0	0	NA	0.00	0.00	3	NA	2
PITG_10639		Yes	No	RxLRfam21	GSR	75%	-0.44	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_10347		Yes	No	RxLRfam1	GSR	100%	0.87	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_10339		Yes	No	RxLRfam1	GSR	100%	-0.56	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_10248		Yes	No	RxLRfam15	GSR	100%	-0.01	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_10244		Yes	Yes	RxLRfam25	InBtw	100%	0.09	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_10227		Yes	No	RxLRfam13	GSR	100%	0.46	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_10116		Yes	No	RxLRfam1	GSR	100%	-0.01	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_09935		Yes	No	RxLRfam18	GSR	100%	-0.09	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_09915		Yes	No	RxLRfam18	GSR	100%	-0.07	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_09861		Yes	Yes	RxLRfam53	GSR	100%	-0.31	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_09838		Yes	No	RxLRfam92	GSR	100%	-0.19	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_09837		Yes	No	RxLRfam92	GSR	100%	-0.16	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_09836		Yes	No	RxLRfam92	InBtw	100%	-0.19	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_09773		Yes	No	RxLRfam6	GSR	100%	-0.07	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_09771		Yes	No	RxLRfam91	GSR	100%	-0.16	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_09758		Yes	No	RxLRfam119	GSR	100%	0.14	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_09754		Yes	No	RxLRfam119	GSR	100%	-0.33	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_09741		Yes	No	RxLRfam6	GSR	100%	-0.63	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_09689		Yes	No	RxLRfam56	GSR	100%	-0.05	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_09685		Yes	No	RxLRfam56	GSR	100%	-0.14	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_09647		Yes	No	RxLRfam2	GSR	100%	-0.04	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_09632	PexRD45	Yes	No	RxLRfam5	GSR	81%	-0.86	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_09622		Yes	No	RxLRfam2	InBtw	96%	-0.32	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_09616		Yes	No	NA	GSR	100%	-0.30	0	0	0	NA	0.00	0.00	NA	NA	2
PITG_09510		Yes	No	RxLRfam20	GSR	100%	-0.14	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_09503		Yes	No	RxLRfam20	GSR	100%	-0.05	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_09499		Yes	No	RxLRfam20	GSR	100%	0.38	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_09498		Yes	No	RxLRfam20	GSR	100%	0.55	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_09496		Yes	No	RxLRfam20	GSR	100%	-0.17	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_09213		Yes	No	RxLRfam27	InBtw	100%	0.25	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_09054		Yes	No	RxLRfam39	GSR	99%	-0.29	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_08949	*Avr2* paralog	Yes	No	RxLRfam7	GSR	100%	0.19	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_08624		Yes	No	RxLRfam89	InBtw	100%	0.04	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_08317		Yes	No	RxLRsng250	InBtw	100%	-0.50	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_08150		Yes	No	RxLRfam19	GSR	100%	0.04	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_08133		Yes	No	RxLRsng158	InBtw	100%	0.01	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_08074		Yes	No	RxLRfam1	GSR	100%	-0.22	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_07954		Yes	No	RxLRfam2	InBtw	100%	-0.24	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_07741		Yes	No	RxLRsng238	GSR	100%	-0.26	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_07634		Yes	No	RxLRfam1	GSR	100%	-0.56	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_07597		Yes	No	RxLRfam26	GSR	100%	-0.15	0	0	0	NA	0.00	0.00	NA	NA	NA

Appendix 4.1. Features of RXLRs in the sequenced *P. infestans* isolate 06_3928A

Gene ID	Annotation	Secreted	Core ortho	RXLR family	Inter. dist.	Cov	CNV	No. of SNPs	No. of Nonsyn SNPs	No. of Syn SNPs	dN/dS	dN	dS	Induced in potato (dpi) by *P.infestans*		
														06_3928A	T30-4	NL07434
PITG_07569		Yes	No	RxLRfam30	GSR	100%	0.07	0	0	0	NA	0.00	0.00	NA	NA	2
PITG_07566		Yes	No	RxLRfam30	GSR	100%	-0.33	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_07558		Yes	No	RxLRfam2	GSR	100%	-0.22	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_07556		Yes	No	RxLRfam2	GSR	100%	-0.23	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_07500		Yes	No	RxLRfam7	GSR	100%	-0.59	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_07499		Yes	No	RxLRfam7	GSR	100%	-0.47	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_07482		Yes	No	RxLRfam7	GSR	100%	0.41	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_07451		Yes	No	RxLRfam116	InBtw	100%	0.23	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_07414		Yes	No	RxLRfam53	GSR	100%	-0.33	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_06552		Yes	No	RxLRfam88	GSR	100%	0.40	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_06485		Yes	No	RxLRsng184	InBtw	100%	-0.35	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_06432		Yes	No	RxLRfam2	GSR	100%	-0.03	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_06419		Yes	No	RxLRfam8	GSR	100%	-0.04	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_06413		Yes	No	RxLRfam8	GSR	100%	-0.28	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_06375		Yes	No	RxLRfam1	GDR	100%	-0.20	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_06305		Yes	No	RxLRfam3	InBtw	100%	-0.27	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_06290		Yes	No	RxLRfam3	GSR	100%	-0.21	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_06092		Yes	No	RxLRsng197	GSR	100%	0.54	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_06083		Yes	No	RxLRsng167	InBtw	100%	-0.19	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_06030		Yes	No	RxLRfam1	GSR	100%	0.05	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_05983		Yes	No	RxLRfam86	GSR	100%	-0.13	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_05981		Yes	No	RxLRsng217	GSR	100%	-0.05	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_05980		Yes	No	RxLRfam86	GSR	100%	-0.07	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_05978		Yes	No	RxLRfam86	GSR	100%	0.05	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_05146		Yes	No	RxLRfam12	GSR	100%	0.48	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_05133		Yes	No	RxLRfam1	GSR	100%	0.29	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_05118		Yes	No	RxLRfam7	GSR	100%	0.12	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_05076		Yes	No	RxLRfam1	InBtw	100%	-0.29	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_05074		Yes	No	RxLRfam1	GSR	99%	-0.42	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_05072		Yes	No	RxLRfam1	InBtw	99%	-0.40	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_05068		Yes	No	RxLRfam115	GSR	100%	-0.38	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_04388	PexRD25	Yes	No	RxLRfam1	GSR	100%	-0.03	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_04373		Yes	No	RxLRfam68	InBtw	100%	-0.32	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_04353		Yes	No	RxLRfam1	GSR	100%	0.09	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_04351		Yes	No	RxLRfam50	GSR	100%	0.39	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_04350		Yes	No	RxLRfam1	GSR	100%	0.12	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_04331		Yes	No	RxLRfam113	GSR	100%	-0.04	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_04329		Yes	No	RxLRfam47	InBtw	100%	0.26	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_04326		Yes	No	RxLRfam47	GSR	100%	-0.03	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_04300		Yes	No	RxLRfam81	GSR	100%	-0.03	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_04282		Yes	No	RxLRfam85	InBtw	100%	-0.28	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_04279		Yes	No	RxLRfam25	InBtw	100%	-0.46	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_04194		Yes	No	RxLRfam5	GSR	100%	0.09	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_04178		Yes	No	RxLRfam10	GSR	100%	0.27	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_04169		Yes	No	RxLRfam10	InBtw	100%	-0.12	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_04164		Yes	No	RxLRfam10	GSR	100%	0.09	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_04153		Yes	No	RxLRfam17	GSR	100%	-0.22	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_04148		Yes	No	RxLRfam83	InBtw	100%	-0.13	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_04145	PexRD29	Yes	No	RxLRfam17	GSR	100%	-0.29	0	0	0	NA	0.00	0.00	NA	NA	NA

Appendix 4.1. Features of RXLRs in the sequenced *P. infestans* isolate 06_3928A

Gene ID	Annotation	Secreted	Core ortho	RXLR family	Inter. dist.	Cov	CNV	No. of SNPs	No. of Nonsyn SNPs	No. of Syn SNPs	dN/dS	dN	dS	Induced in potato (dpi) by *P.infestans*		
														06_3928A	T30-4	NL07434
PITG_04139		Yes	No	RxLRfam83	InBtw	100%	-0.21	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_04099		Yes	No	RxLRfam82	GSR	100%	1.15	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_04081		Yes	No	RxLRfam5	GSR	100%	-0.48	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_04074		Yes	No	RxLRsng195	InBtw	100%	0.07	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_04050		Yes	No	RxLRfam81	GSR	100%	0.16	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_03155		Yes	Yes	RxLRsng229	GDR	100%	0.19	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_02900		Yes	No	RxLRfam46	GSR	100%	-0.11	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_02897		Yes	No	RxLRfam111	InBtw	100%	0.09	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_02843	PexRD30	Yes	No	RxLRfam65	InBtw	100%	0.29	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_02830		Yes	No	RxLRfam65	InBtw	100%	-0.25	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_01875		Yes	No	RxLRfam109	InBtw	100%	-0.33	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_00821		Yes	No	RxLRfam108	InBtw	100%	0.15	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_00707		Yes	No	RxLRfam107	GDR	100%	-0.09	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_00579		Yes	Yes	RxLRfam14	GSR	100%	-0.15	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_00366	PexRD43	Yes	No	RxLRfam80	GSR	100%	-0.27	0	0	0	NA	0.00	0.00	NA	NA	NA
PITG_15972		Yes	No	RxLRfam7	GSR	100%	-0.85	0	0	0	NA	0.00	0.00	NA	2 and 3	NA
PITG_11484		Yes	No	RxLRfam120	GSR	100%	-0.22	0	0	0	NA	0.00	0.00	NA	2 and 3	NA
PITG_08278		Yes	No	RxLRfam7	GSR	100%	0.06	0	0	0	NA	0.00	0.00	NA	2 and 3	NA
PITG_18221		Yes	No	RxLRfam124	GSR	0%	NA	0	0	0	NA	NA	NA	NA	2	NA
PITG_23231		Yes	No	RxLRfam54	Not	0%	NA	0	0	0	NA	NA	NA	NA	NA	NA
PITG_21681		Yes	No	RxLRfam14	Not	0%	NA	0	0	0	NA	NA	NA	NA	NA	NA
PITG_20857		Yes	No	RxLRfam5	GSR	0%	NA	0	0	0	NA	NA	NA	NA	NA	NA
PITG_17217		Yes	No	RxLRfam45	InBtw	0%	NA	0	0	0	NA	NA	NA	NA	NA	NA
PITG_15718		Yes	No	RxLRfam14	GSR	0%	NA	0	0	0	NA	NA	NA	NA	NA	NA
PITG_15712		Yes	No	RxLRsng162	InBtw	0%	NA	0	0	0	NA	NA	NA	NA	NA	NA
PITG_12010		Yes	No	RxLRfam47	GSR	0%	NA	0	0	0	NA	NA	NA	NA	NA	NA
PITG_18215		Yes	No	RxLRfam124	GSR	0%	NA	0	0	0	NA	NA	NA	NA	2 and 3	NA
PITG_16663	*Avr1*	Yes	No	RxLRfam2	GSR	0%	NA	0	0	0	NA	NA	NA	NA	2 and 3	NA
PITG_10341		Yes	No	RxLRfam1	GSR	100%	-0.21	7	6	1	2.02	0.01	0.00	NA	NA	NA
PITG_23210		Yes	No	RxLRsng185	Not	100%	0.20	6	5	1	1.79	0.02	0.01	NA	NA	NA
PITG_15225		Yes	No	RxLRfam28	GSR	100%	0.28	5	4	1	1.68	0.01	0.01	NA	NA	NA
PITG_01905		Yes	No	RxLRfam110	GSR	100%	-0.11	5	4	1	1.54	0.02	0.01	NA	NA	NA
PITG_05096		Yes	No	RxLRfam1	InBtw	88%	0.54	10	8	2	1.47	0.01	0.01	NA	2	NA
PITG_05095		Yes	No	RxLRfam1	GSR	99%	-0.07	9	7	2	1.29	0.01	0.01	NA	NA	NA
PITG_04063		Yes	No	RxLRfam1	InBtw	100%	-0.25	4	3	1	1.22	0.00	0.00	NA	NA	NA
PITG_01904		Yes	No	RxLRfam15	GSR	82%	0.44	13	10	3	1.20	0.05	0.04	NA	NA	NA
PITG_15226		Yes	No	RxLRfam28	GSR	100%	0.07	4	3	1	1.17	0.01	0.01	NA	NA	NA
PITG_14203		Yes	No	RxLRfam33	GSR	100%	0.55	12	9	3	1.13	0.03	0.02	NA	NA	NA
PITG_11952		Yes	No	RxLRfam23	GSR	100%	0.47	9	7	2	1.04	0.01	0.01	NA	NA	NA
PITG_00619		Yes	No	RxLRfam14	GDR	100%	-0.32	4	3	1	1.03	0.01	0.01	NA	NA	NA
PITG_15277		Yes	No	RxLRfam1	InBtw	100%	-0.36	10	7	3	0.88	0.02	0.02	NA	NA	NA
PITG_15726		Yes	No	NA	GSR	100%	4.19	6	4	2	0.87	0.01	0.01	NA	NA	NA
PITG_13503		Yes	No	RxLRfam8	GSR	95%	-0.35	4	3	1	0.86	0.01	0.01	NA	NA	NA
PITG_12722		Yes	No	RxLRfam4	GDR	100%	-0.14	3	2	1	0.81	0.01	0.01	NA	NA	NA
PITG_16845		Yes	No	RxLRfam1	GSR	100%	-0.53	3	2	1	0.75	0.00	0.00	NA	NA	NA
PITG_23024		Yes	No	RxLRfam1	GSR	100%	0.14	8	5	3	0.69	0.01	0.01	NA	NA	NA
PITG_22972		Yes	No	RxLRfam7	GSR	100%	-0.52	10	6	4	0.69	0.04	0.06	NA	NA	NA
PITG_10808		Yes	Yes	RxLRfam31	GSR	100%	-0.15	3	2	1	0.65	0.00	0.01	NA	NA	NA
PITG_11953		Yes	No	RxLRfam56	GSR	100%	0.42	20	11	9	0.56	0.03	0.05	NA	NA	NA

Appendix 4.1. Features of RXLRs in the sequenced *P. infestans* isolate 06_3928A

Gene ID	Annotation	Secreted	Core ortho	RXLR family	Inter. dist.	Cov	CNV	No. of SNPs	No. of Nonsyn SNPs	No. of Syn SNPs	dN/dS	dN	dS	Induced in potato (dpi) by *P.infestans*		
														06_3928A	T30-4	NL07434
PITG_07736		Yes	No	RxLRfam37	InBtw	100%	0.03	6	4	2	0.55	0.01	0.03	NA	NA	NA
PITG_04355		Yes	No	RxLRfam114	GSR	100%	-0.33	5	3	2	0.53	0.01	0.02	NA	NA	NA
PITG_18156		Yes	No	RxLRfam39	GSR	100%	-0.03	7	4	3	0.52	0.01	0.02	NA	NA	NA
PITG_15037		Yes	No	RxLRfam1	GSR	100%	-0.13	2	1	1	0.45	0.00	0.01	NA	NA	NA
PITG_05771		Yes	No	RxLRfam16	GDR	100%	-0.02	2	1	1	0.45	0.00	0.00	NA	NA	NA
PITG_22824		Yes	No	NA	GSR	100%	-0.18	2	1	1	0.43	0.00	0.00	NA	NA	NA
PITG_13537		Yes	No	RxLRfam37	GSR	70%	-0.35	2	1	1	0.43	0.00	0.01	NA	NA	NA
PITG_23048		Yes	No	RxLRfam98	GSR	100%	0.49	2	1	1	0.42	0.00	0.01	NA	NA	NA
PITG_13125		Yes	No	RxLRfam16	GSR	100%	0.37	2	1	1	0.42	0.00	0.01	NA	NA	NA
PITG_07766		Yes	No	RxLRfam37	InBtw	100%	-0.13	2	1	1	0.40	0.00	0.01	NA	NA	NA
PITG_22828		Yes	No	RxLRfam26	InBtw	100%	-0.36	2	1	1	0.38	0.00	0.01	NA	NA	NA
PITG_04354		Yes	No	RxLRfam114	GSR	100%	-0.15	15	8	7	0.37	0.02	0.07	NA	NA	NA
PITG_23092		Yes	No	RxLRsng204	InBtw	100%	-0.51	4	2	2	0.36	0.01	0.01	NA	NA	NA
PITG_13507		Yes	No	RxLRfam8	GSR	92%	-0.62	5	2	3	0.25	0.00	0.01	NA	NA	NA
PITG_13535		Yes	No	RxLRfam37	GSR	100%	0.09	3	1	2	0.20	0.00	0.02	NA	NA	NA
PITG_23047		Yes	No	RxLRfam98	GSR	100%	0.01	4	1	3	0.14	0.00	0.02	NA	NA	NA
PITG_13509		Yes	No	RxLRfam8	GSR	100%	-0.10	4	1	3	0.13	0.00	0.02	NA	NA	NA
PITG_23011		Yes	No	RxLRfam69	Not	100%	-0.22	1	0	1	0.00	0.00	0.01	NA	NA	NA
PITG_22825		Yes	No	RxLRsng208	GSR	100%	-0.16	1	0	1	0.00	0.00	0.01	NA	NA	NA
PITG_22722		Yes	No	RxLRfam1	InBtw	100%	-0.42	1	0	1	0.00	0.00	0.00	NA	NA	NA
PITG_21933		Yes	No	RxLRfam93	Not	67%	0.13	1	0	1	0.00	0.00	0.01	NA	NA	NA
PITG_14684		Yes	No	NA	GSR	100%	-0.09	1	0	1	0.00	0.00	0.01	NA	NA	NA
PITG_14054		Yes	No	RxLRfam2	GSR	100%	0.12	1	0	1	0.00	0.00	0.01	3	NA	NA
PITG_12046		Yes	No	RxLRfam83	InBtw	100%	0.26	1	0	1	0.00	0.00	0.01	NA	NA	NA
PITG_09111		Yes	No	RxLRfam1	GSR	100%	0.03	1	0	1	0.00	0.00	0.01	NA	NA	NA
PITG_01907		Yes	No	RxLRfam110	GSR	100%	0.00	1	0	1	0.00	0.00	0.00	NA	NA	NA

Core ortho.,core orthologs; Int. dis., Intergenic distance; Cov, breadth of coverage; Syn SNPs, Synonymous SNPs; Nonsyn SNPs, Nonsynonymous SNPs.

Appendix 4.2. List of genes with duplications in the sequenced *P. infestans* 06_3928A genome

Gene ID	Annotation	Secreted	Core ortholog	Effector type	RXLR family	Intergenic distance	CNV
PITG_11913	heat shock cognate 70 kDa protein	No	Yes	NA	NA	InBtw	7.50
PITG_01323	conserved hypothetical protein	No	No	NA	NA	GSR	6.18
PITG_21039	conserved hypothetical protein	Yes	No	NA	NA	InBtw	5.72
PITG_23234	M96 mating-specific protein, pseudogene	No	No	NA	NA	Not	5.52
PITG_21038	conserved hypothetical protein	Yes	No	NA	NA	Not	5.45
PITG_14787	secreted RxLR effector peptide, putative	Yes	No	RXLR	RxLRfam6	GSR	4.89
PITG_21978	predicted protein	No	No	NA	NA	Not	4.50
PITG_15726	RXLR effector family protein, putative	Yes	No	RXLR	NA	GSR	4.19
PITG_04993	predicted protein	No	No	NA	NA	GSR	4.03
PITG_04088	predicted protein	No	No	NA	NA	GSR	4.02
PITG_16549	hypothetical protein	No	No	NA	NA	GSR	3.91
PITG_07090	hypothetical protein	No	No	NA	NA	GSR	3.90
PITG_02168	conserved hypothetical protein	Yes	No	NA	NA	GSR	3.84
PITG_13529	secreted RxLR effector peptide, putative	Yes	No	RXLR	RxLRfam50	GSR	3.75
PITG_15353	predicted protein	No	No	NA	NA	GSR	3.60
PITG_10406	predicted protein	No	No	NA	NA	InBtw	3.58
PITG_20653	predicted protein	No	No	NA	NA	Not	3.51
PITG_14783	secreted RxLR effector peptide, putative	Yes	No	RXLR	RxLRfam6	GSR	3.46
PITG_04994	predicted protein	No	No	NA	NA	InBtw	3.40
PITG_07565	Folate-Biopterin Transporter (FBT) Family	No	Yes	NA	NA	InBtw	3.32
PITG_21862	predicted protein	No	No	NA	NA	Not	3.02
PITG_09200	hypothetical protein	No	No	NA	NA	GSR	2.97
PITG_21893	predicted protein	No	No	NA	NA	Not	2.82
PITG_08050	hypothetical protein	No	No	NA	NA	InBtw	2.82
PITG_22632	polysaccharide lyase, putative	No	No	NA	NA	Not	2.59
PITG_08526	conserved hypothetical protein	No	No	NA	NA	GSR	2.55
PITG_22425	predicted protein	No	No	NA	NA	Not	2.52
PITG_22237	predicted protein	No	No	NA	NA	Not	2.51
PITG_13690	predicted protein	No	No	NA	NA	InBtw	2.47
PITG_09255	polysaccharide lyase, putative	No	No	NA	NA	GSR	2.46
PITG_02927	hypothetical protein	No	No	NA	NA	InBtw	2.46
PITG_21335	hypothetical protein	No	No	NA	NA	InBtw	2.45
PITG_03470	conserved hypothetical protein	No	No	NA	NA	GSR	2.45
PITG_19218	predicted protein	No	No	NA	NA	Not	2.44
PITG_11879	predicted protein	No	No	NA	NA	GSR	2.37
PITG_11943	hypothetical protein	No	Yes	NA	NA	InBtw	2.36
PITG_13098	elicitin-like protein	Yes	No	elicitins	NA	GSR	2.34
PITG_22683	secreted RxLR effector peptide, putative	Yes	No	RXLR	RxLRsng209	InBtw	2.34
PITG_22786	M96 mating-specific protein, putative, 5' partial	No	No	NA	NA	GDR	2.33
PITG_09197	Major Intrinsic Protein (MIP) Family	No	Yes	NA	NA	GSR	2.33
PITG_13107	predicted protein	No	No	NA	NA	InBtw	2.32
PITG_09128	predicted protein	No	No	NA	NA	GSR	2.31
PITG_08854	predicted protein	No	No	NA	NA	GSR	2.28
PITG_10409	conserved hypothetical protein	No	No	NA	NA	GDR	2.27
PITG_14829	predicted protein	No	No	NA	NA	InBtw	2.24
PITG_09195	hypothetical protein	No	No	NA	NA	GSR	2.24
PITG_21772	predicted protein	No	No	NA	NA	Not	2.24
PITG_11942	conserved hypothetical protein	Yes	No	NA	NA	InBtw	2.21
PITG_09110	hypothetical protein	No	No	NA	NA	GSR	2.20
PITG_22904	helitron helicase-like protein	No	No	NA	NA	GSR	2.18
PITG_19556	predicted protein	No	No	NA	NA	GSR	2.16
PITG_20582	hypothetical protein	No	No	NA	NA	GSR	2.14

Appendix 4.2. List of genes with duplications in the sequenced *P. infestans* 06_3928A genome

Gene ID	Annotation	Secreted	Core ortholog	Effector type	RXLR family	Intergenic distance	CNV
PITG_12004	predicted protein	No	No	NA	NA	InBtw	2.13
PITG_22138	predicted protein	No	No	NA	NA	Not	2.12
PITG_19438	predicted protein	No	No	NA	NA	GDR	2.12
PITG_14192	predicted protein	No	No	NA	NA	GSR	2.12
PITG_04985	hypothetical protein similar to xylitol dehydrogenase	No	No	NA	NA	GSR	2.11
PITG_15728	secreted RxLR effector peptide, putative	Yes	No	RXLR	RxLRfam23	GSR	2.03
PITG_09196	Major Intrinsic Protein (MIP) Family	No	No	NA	NA	GSR	2.03
PITG_16285	secreted RxLR effector peptide, putative	No	No	RXLR	RxLRfam47	GSR	2.03
PITG_22252	hypothetical protein	No	No	NA	NA	Not	1.99
PITG_11944	conserved hypothetical protein	Yes	Yes	NA	NA	GSR	1.98
PITG_11511	predicted protein	No	No	NA	NA	InBtw	1.98
PITG_08647	polysaccharide lyase, putative	No	No	NA	NA	InBtw	1.97
PITG_21835	hypothetical protein	No	No	NA	NA	Not	1.95
PITG_04397	conserved hypothetical protein	No	No	NA	NA	InBtw	1.92
PITG_21938	conserved hypothetical protein	No	No	NA	NA	Not	1.91
PITG_18806	predicted protein	No	No	NA	NA	InBtw	1.91
PITG_22383	glycoside hydrolase, putative	No	No	NA	NA	Not	1.90
PITG_18851	predicted protein	No	Yes	NA	NA	Not	1.90
PITG_23194	cys-rich secreted peptide, putative	Yes	No	cys-rich	NA	Not	1.89
PITG_05708	predicted protein	No	No	NA	NA	InBtw	1.89
PITG_21486	predicted protein	No	No	NA	NA	Not	1.87
PITG_12738	predicted protein	No	No	NA	NA	GSR	1.87
PITG_08524	predicted protein	No	No	NA	NA	GSR	1.87
PITG_19800	secreted RxLR effector peptide, putative	Yes	No	RXLR	RxLRfam50	Not	1.86
PITG_13100	conserved hypothetical protein	Yes	No	NA	NA	GSR	1.85
PITG_12552	hypothetical protein	No	No	NA	NA	InBtw	1.80
PITG_19510	Crinkler (CRN) family protein, pseudogene	No	No	CRINKLER	NA	Not	1.79
PITG_13688	predicted protein	No	No	NA	NA	GSR	1.78
PITG_18737	predicted protein	No	No	NA	NA	InBtw	1.76
PITG_16348	predicted protein	No	No	NA	NA	GSR	1.76
PITG_12199	predicted protein	No	No	NA	NA	GSR	1.70
PITG_11938	hypothetical protein	No	No	NA	NA	InBtw	1.69
PITG_19266	predicted protein	No	No	NA	NA	InBtw	1.68
PITG_08866	glycoside hydrolase, putative	No	No	NA	NA	GDR	1.68
PITG_16280	cysteine protease family C44, putative	No	No	NA	NA	GSR	1.66
PITG_09318	hypothetical protein	No	No	NA	NA	GSR	1.66
PITG_17208	polysaccharide lyase, putative	No	No	NA	NA	GSR	1.66
PITG_20487	conserved hypothetical protein	No	No	NA	NA	InBtw	1.66
PITG_04204	conserved hypothetical protein	No	No	NA	NA	InBtw	1.65
PITG_19415	hypothetical protein	No	No	NA	NA	GSR	1.64
PITG_14275	predicted protein	No	No	NA	NA	InBtw	1.63
PITG_16283	secreted RxLR effector peptide, putative	Yes	No	RXLR	RxLRfam1	GSR	1.61
PITG_23028	elicitin-like protein, pseudogene	No	No	SCR	NA	InBtw	1.60
PITG_11871	hypothetical protein	Yes	No	NA	NA	GSR	1.60
PITG_13696	hypothetical protein, contains CRN-like motif, pseudogene	No	No	NA	NA	InBtw	1.60
PITG_13101	conserved hypothetical protein	No	No	NA	NA	GSR	1.59
PITG_13708	hypothetical protein, contains CRN-like motif, pseudogene	No	No	NA	NA	GSR	1.59
PITG_09190	predicted protein	No	No	NA	NA	GSR	1.59
PITG_04087	predicted protein	No	No	NA	NA	GSR	1.59
PITG_12205	predicted protein	No	No	NA	NA	GSR	1.59
PITG_13102	conserved hypothetical protein	No	No	NA	NA	GSR	1.58
PITG_18807	predicted protein	No	No	NA	NA	GDR	1.55

Appendix 4.2. List of genes with duplications in the sequenced *P. infestans* 06_3928A genome

Gene ID	Annotation	Secreted	Core ortholog	Effector type	RXLR family	Intergenic distance	CNV
PITG_17519	predicted protein	No	No	NA	NA	InBtw	1.55
PITG_21485	predicted protein	No	No	NA	NA	GDR	1.54
PITG_22296	predicted protein	No	No	NA	NA	Not	1.54
PITG_04096	predicted protein	No	No	NA	NA	GSR	1.54
PITG_21773	predicted protein	No	No	NA	NA	Not	1.53
PITG_14130	hypothetical protein	No	No	NA	NA	InBtw	1.53
PITG_01534	predicted protein	No	No	NA	NA	InBtw	1.52
PITG_20581	glycine-rich protein. similar to fibroin	No	No	NA	NA	Not	1.52
PITG_02612	predicted protein	No	No	NA	NA	GSR	1.50
PITG_04462	predicted protein	No	No	NA	NA	GSR	1.50
PITG_19278	hypothetical protein	No	No	NA	NA	GSR	1.50
PITG_22524	predicted protein	No	No	NA	NA	Not	1.50
PITG_09105	predicted protein	No	No	NA	NA	GSR	1.49
PITG_21970	predicted protein	No	No	NA	NA	Not	1.49
PITG_11627	predicted protein	No	No	NA	NA	InBtw	1.49
PITG_13817	predicted protein	No	No	NA	NA	InBtw	1.48
PITG_10549	conserved hypothetical protein	No	No	NA	NA	InBtw	1.48
PITG_22070	hypothetical protein	No	No	NA	NA	Not	1.48
PITG_15731	ATP-binding Cassette (ABC) Superfamily	No	No	NA	NA	GSR	1.47
PITG_09141	hypothetical protein	No	No	NA	NA	GDR	1.47
PITG_22227	polysaccharide lyase, putative	No	No	NA	NA	Not	1.46
PITG_13589	conserved hypothetical protein	No	No	NA	NA	InBtw	1.46
PITG_08517	predicted protein	No	No	NA	NA	InBtw	1.46
PITG_12553	predicted protein	No	No	NA	NA	InBtw	1.46
PITG_09193	Major Intrinsic Protein (MIP) Family	No	No	NA	NA	GSR	1.45
PITG_09102	predicted protein	No	No	NA	NA	GSR	1.45
PITG_15335	thioredoxin/dynein outer arm protein	No	No	NA	NA	GSR	1.45
PITG_09297	hypothetical protein	No	Yes	NA	NA	GDR	1.44
PITG_09256	polysaccharide lyase, putative	No	No	NA	NA	GSR	1.44
PITG_13117	ATP-binding Cassette (ABC) Superfamily	No	No	NA	NA	InBtw	1.43
PITG_18897	hypothetical protein similar to novel protein	No	No	NA	NA	GSR	1.43
PITG_22671	protein kinase, putative, pseudogene	No	No	NA	NA	GSR	1.42
PITG_21926	predicted protein	No	No	NA	NA	Not	1.41
PITG_12008	predicted protein	No	No	NA	NA	GSR	1.41
PITG_20889	conserved hypothetical protein	No	No	NA	NA	Not	1.40
PITG_16260	predicted protein	No	No	NA	NA	InBtw	1.40
PITG_09121	conserved hypothetical protein	No	No	NA	NA	InBtw	1.40
PITG_09083	predicted protein	No	No	NA	NA	GSR	1.39
PITG_12003	predicted protein	No	No	NA	NA	InBtw	1.38
PITG_11860	predicted protein	No	No	NA	NA	GSR	1.37
PITG_11946	conserved hypothetical protein	Yes	Yes	NA	NA	GSR	1.37
PITG_20008	predicted protein	No	No	NA	NA	GSR	1.37
PITG_14295	predicted protein	No	No	NA	NA	InBtw	1.36
PITG_23156	small cysteine rich protein SCR58	Yes	No	SCR	NA	GSR	1.36
PITG_09199	hypothetical protein	No	No	NA	NA	GSR	1.36
PITG_21132	conserved hypothetical protein	No	No	NA	NA	Not	1.36
PITG_16551	predicted protein	No	No	NA	NA	GSR	1.36
PITG_09165	conserved hypothetical protein	No	No	NA	NA	GSR	1.36
PITG_13106	predicted protein	No	No	NA	NA	GSR	1.36
PITG_17654	predicted protein	No	No	NA	NA	GSR	1.35
PITG_21981	conserved hypothetical protein	No	No	NA	NA	Not	1.35
PITG_09156	predicted protein	No	Yes	NA	NA	GSR	1.35

Appendix 4.2. List of genes with duplications in the sequenced *P. infestans* 06_3928A genome

Gene ID	Annotation	Secreted	Core ortholog	Effector type	RXLR family	Intergenic distance	CNV
PITG_10326	predicted protein	No	No	NA	NA	GSR	1.34
PITG_18231	glycoside hydrolase, putative	No	No	NA	NA	GSR	1.33
PITG_23017	predicted protein	No	No	NA	NA	InBtw	1.33
PITG_18580	predicted protein	No	No	NA	NA	GSR	1.33
PITG_19945	predicted protein	No	No	NA	NA	Not	1.32
PITG_22295	predicted protein	No	No	NA	NA	Not	1.32
PITG_05159	conserved hypothetical protein	No	No	NA	NA	GSR	1.30
PITG_22319	conserved hypothetical protein	No	No	NA	NA	Not	1.30
PITG_22724	secreted RxLR effector peptide, putative	Yes	No	RXLR	RxLRfam67	GSR	1.29
PITG_04988	hypothetical protein	No	No	NA	NA	GSR	1.29
PITG_17210	polysaccharide lyase, putative	No	No	NA	NA	GSR	1.29
PITG_21164	predicted protein	No	No	NA	NA	Not	1.29
PITG_14790	predicted protein	No	No	NA	NA	GSR	1.29
PITG_17381	predicted protein	No	No	NA	NA	GSR	1.29
PITG_16566	predicted protein	No	No	NA	NA	InBtw	1.29
PITG_16282	secreted RxLR effector peptide, putative	Yes	No	RXLR	RxLRfam18	GSR	1.29
PITG_14119	conserved hypothetical protein	Yes	No	NA	NA	GSR	1.29
PITG_18761	conserved hypothetical protein	No	No	NA	NA	GSR	1.28
PITG_10235	predicted protein	No	No	NA	NA	InBtw	1.28
PITG_05038	chromodomain protein, putative	No	No	NA	NA	InBtw	1.27
PITG_07519	predicted protein	No	No	NA	NA	GSR	1.27
PITG_22893	elicitin-like kprotein, pseudogene	No	No	NA	NA	GSR	1.27
PITG_13816	predicted protein	No	No	NA	NA	InBtw	1.27
PITG_20009	predicted protein	No	No	NA	NA	GSR	1.27
PITG_16259	predicted protein	No	No	NA	NA	InBtw	1.27
PITG_14118	conserved hypothetical protein	No	No	NA	NA	GSR	1.27
PITG_13568	conserved hypothetical protein	No	No	NA	NA	GDR	1.26
PITG_14015	predicted protein	No	No	NA	NA	InBtw	1.26
PITG_20334	ATP-binding Cassette (ABC) Superfamily	No	No	NA	NA	Not	1.25
PITG_19936	NPP1-like protein	Yes	No	NLP	NA	GSR	1.25
PITG_11510	predicted protein	No	No	NA	NA	InBtw	1.25
PITG_09203	Major Intrinsic Protein (MIP) Family	No	No	NA	NA	InBtw	1.25
PITG_19760	predicted protein	No	No	NA	NA	GSR	1.25
PITG_10548	predicted protein	No	No	NA	NA	GDR	1.25
PITG_19563	hypothetical protein	No	No	NA	NA	GSR	1.25
PITG_18482	polysaccharide lyase, putative	Yes	No	enzyme, lyase	NA	GSR	1.24
PITG_20773	conserved hypothetical protein	No	No	NA	NA	GDR	1.23
PITG_09656	predicted protein	No	No	NA	NA	GSR	1.22
PITG_21107	secreted RxLR effector peptide, putative	Yes	No	RXLR	RxLRfam3	Not	1.22
PITG_01392	predicted protein	No	No	NA	NA	GSR	1.22
PITG_17752	predicted protein	No	No	NA	NA	GSR	1.22
PITG_15680	predicted protein	No	No	NA	NA	InBtw	1.20
PITG_21133	predicted protein	No	No	NA	NA	Not	1.20
PITG_21349	aspartyl-tRNA synthetase	No	No	NA	NA	Not	1.20
PITG_06027	predicted protein	No	No	NA	NA	GSR	1.19
PITG_11961	ATP-binding Cassette (ABC) Superfamily	No	No	NA	NA	InBtw	1.19
PITG_21152	secreted RxLR effector peptide, putative	No	No	RXLR	RxLRfam1	Not	1.19
PITG_22437	predicted protein	No	No	NA	NA	Not	1.19
PITG_20859	conserved hypothetical protein	No	No	NA	NA	GSR	1.19
PITG_19622	hypothetical protein	No	No	NA	NA	InBtw	1.19
PITG_19486	conserved hypothetical protein	No	No	NA	NA	GSR	1.18
PITG_09236	predicted protein	No	No	NA	NA	GSR	1.17

Appendix 4.2. List of genes with duplications in the sequenced *P. infestans* 06_3928A genome

Gene ID	Annotation	Secreted	Core ortholog	Effector type	RXLR family	Intergenic distance	CNV
PITG_09317	predicted protein	No	No	NA	NA	GSR	1.17
PITG_17531	expressed protein, contains CRN-like motif	No	No	NA	NA	GSR	1.17
PITG_01460	predicted protein	No	No	NA	NA	GSR	1.16
PITG_01459	predicted protein	No	No	NA	NA	InBtw	1.16
PITG_20772	protein kinase	No	No	NA	NA	GDR	1.16
PITG_17433	hypothetical protein	No	No	NA	NA	GSR	1.16
PITG_17163	predicted protein	No	No	NA	NA	InBtw	1.16
PITG_22933	secreted RxLR effector peptide, putative	Yes	No	RXLR	RxLRfam98	GSR	1.16
PITG_19937	predicted protein	No	No	NA	NA	GSR	1.15
PITG_19274	predicted protein	No	No	NA	NA	GSR	1.15
PITG_04099	secreted RxLR effector peptide, putative	Yes	No	RXLR	RxLRfam82	GSR	1.15
PITG_09521	hypothetical protein	No	No	NA	NA	InBtw	1.15
PITG_10179	predicted protein	No	No	NA	NA	GSR	1.14
PITG_04316	conserved hypothetical protein	No	No	NA	NA	GSR	1.14
PITG_03153	predicted protein	No	No	NA	NA	InBtw	1.14
PITG_19848	predicted protein	No	No	NA	NA	InBtw	1.14
PITG_13698	conserved hypothetical protein, contains CRN-like motif	No	No	NA	NA	GSR	1.14
PITG_14254	phosphoenolpyruvate carboxykinase	No	No	NA	NA	GSR	1.14
PITG_10547	predicted protein	No	No	NA	NA	GDR	1.14
PITG_18478	polysaccharide lyase, putative	No	No	NA	NA	GSR	1.14
PITG_09061	predicted protein	No	No	NA	NA	Not	1.13
PITG_19341	predicted protein	No	No	NA	NA	GSR	1.13
PITG_13704	expressed protein, contains CRN-like motif	No	No	NA	NA	GSR	1.13
PITG_12037	cysteine protease family C44, putative	No	No	NA	NA	InBtw	1.12
PITG_09175	protease inhibitor EpiC2A	Yes	No	Enzyme Inhibitor	NA	GSR	1.12
PITG_18350	hypothetical protein	No	No	NA	NA	InBtw	1.12
PITG_02563	predicted protein	No	No	NA	NA	GSR	1.12
PITG_02377	predicted protein	No	No	NA	NA	GSR	1.12
PITG_11846	hypothetical protein	No	No	NA	NA	GSR	1.12
PITG_11873	predicted protein	No	No	NA	NA	GSR	1.12
PITG_06490	hypothetical protein	No	No	NA	NA	InBtw	1.12
PITG_13700	hypothetical protein, contains CRN-like motif, pseudogene	No	No	NA	NA	GSR	1.12
PITG_11988	Mitochondrial Carrier (MC) Family	No	No	NA	NA	InBtw	1.12
PITG_16747	hypothetical protein	No	No	NA	NA	InBtw	1.12
PITG_09287	hypothetical protein	No	Yes	NA	NA	GDR	1.12
PITG_11901	phosphoenolpyruvate carboxykinase	No	No	NA	NA	GSR	1.11
PITG_21163	predicted protein	No	No	NA	NA	Not	1.11
PITG_23170	secreted peptide candidate, ORF supported by proteomics	No	No	NA	NA	GSR	1.11
PITG_09394	pyruvate kinase	No	No	NA	NA	InBtw	1.10
PITG_16303	predicted protein	No	No	NA	NA	InBtw	1.10
PITG_17526	expressed protein, contains CRN-like motif	No	No	NA	NA	GSR	1.10
PITG_21701	hypothetical protein	No	No	NA	NA	Not	1.10
PITG_01571	predicted protein	No	No	NA	NA	GSR	1.09
PITG_11583	conserved hypothetical protein	Yes	No	NA	NA	InBtw	1.09
PITG_07554	predicted protein	No	No	NA	NA	GSR	1.09
PITG_14175	hypothetical protein	No	Yes	NA	NA	GDR	1.09
PITG_11892	predicted protein	No	No	NA	NA	GSR	1.09
PITG_04390	predicted protein	No	No	NA	NA	GSR	1.09
PITG_09323	predicted protein	No	No	NA	NA	InBtw	1.09
PITG_13454	predicted protein	No	No	NA	NA	GSR	1.09
PITG_17540	conserved hypothetical protein, contains CRN-like motif	No	No	NA	NA	GSR	1.09
PITG_21106	predicted protein	No	Yes	NA	NA	GSR	1.09

Appendix 4.2. List of genes with duplications in the sequenced *P. infestans* 06_3928A genome

Gene ID	Annotation	Secreted	Core ortholog	Effector type	RXLR family	Intergenic distance	CNV
PITG_17538	predicted protein	No	No	NA	NA	GSR	1.09
PITG_09238	conserved hypothetical protein	No	No	NA	NA	GSR	1.09
PITG_19555	predicted protein	No	No	NA	NA	GSR	1.08
PITG_16815	predicted protein	No	No	NA	NA	Not	1.08
PITG_13482	predicted protein	No	No	NA	NA	GSR	1.08
PITG_12856	conserved hypothetical protein	No	No	NA	NA	GSR	1.07
PITG_17542	hypothetical protein, contains CRN-like motif, pseudogene	No	No	NA	NA	InBtw	1.07
PITG_02519	M96 mating-specific protein, putative	Yes	No	NA	NA	GDR	1.07
PITG_11934	predicted protein	No	No	NA	NA	InBtw	1.07
PITG_15593	predicted protein	No	No	NA	NA	GSR	1.07
PITG_15421	predicted protein	No	No	NA	NA	GSR	1.07
PITG_18347	protein kinase	No	No	NA	NA	GSR	1.07
PITG_02644	conserved hypothetical protein	No	No	NA	NA	InBtw	1.06
PITG_14487	predicted protein	No	No	NA	NA	GSR	1.06
PITG_11881	predicted protein	No	No	NA	NA	GSR	1.06
PITG_20140	hypothetical protein	No	No	NA	NA	GSR	1.06
PITG_18351	hypothetical protein	No	No	NA	NA	InBtw	1.06
PITG_01987	hypothetical protein	No	No	NA	NA	InBtw	1.06
PITG_17667	predicted protein	No	No	NA	NA	GSR	1.06
PITG_11959	hypothetical protein	No	Yes	NA	NA	InBtw	1.06
PITG_02434	conserved hypothetical protein	Yes	No	NA	NA	InBtw	1.06
PITG_19366	polysaccharide lyase, putative	Yes	No	enzyme, lyase	NA	GSR	1.05
PITG_09198	predicted protein	No	No	NA	NA	GSR	1.05
PITG_02188	conserved hypothetical protein	No	No	NA	NA	GSR	1.05
PITG_17219	predicted protein	No	No	NA	NA	Not	1.05
PITG_00246	hypothetical protein	No	No	NA	NA	GDR	1.05
PITG_09319	predicted protein	No	No	NA	NA	GSR	1.04
PITG_12209	predicted protein	No	No	NA	NA	GSR	1.04
PITG_09194	Major Intrinsic Protein (MIP) Family	No	No	NA	NA	GSR	1.04
PITG_10232	secreted RxLR effector peptide, putative	Yes	No	RXLR	RxLRfam69	GSR	1.04
PITG_06086	predicted protein	No	No	NA	NA	InBtw	1.04
PITG_21293	conserved hypothetical protein	No	No	NA	NA	GSR	1.04
PITG_08547	predicted protein	No	No	NA	NA	InBtw	1.04
PITG_13676	conserved hypothetical protein	No	Yes	NA	NA	GDR	1.04
PITG_15346	hypothetical protein	No	No	NA	NA	GSR	1.04
PITG_22282	predicted protein	No	No	NA	NA	Not	1.04
PITG_08639	conserved hypothetical protein	No	No	NA	NA	InBtw	1.04
PITG_17647	predicted protein	No	No	NA	NA	GSR	1.04
PITG_11552	predicted protein	No	No	NA	NA	GSR	1.04
PITG_11963	ATP-binding Cassette (ABC) Superfamily	No	No	NA	NA	InBtw	1.04
PITG_14071	predicted protein	No	No	NA	NA	GSR	1.03
PITG_08391	predicted protein	No	No	NA	NA	GSR	1.03
PITG_02179	hypothetical protein	No	No	NA	NA	GSR	1.03
PITG_15334	predicted protein	No	No	NA	NA	GSR	1.03
PITG_19831	secreted RxLR effector peptide, putative	Yes	No	RXLR	RxLRfam40	GSR	1.02
PITG_13691	predicted protein	No	No	NA	NA	GDR	1.02
PITG_13937	predicted protein	No	No	NA	NA	InBtw	1.02
PITG_14826	predicted protein	No	No	NA	NA	GDR	1.02
PITG_16277	hypothetical protein	No	No	NA	NA	GSR	1.02
PITG_09162	predicted protein	No	No	NA	NA	GSR	1.02
PITG_09228	conserved hypothetical protein	No	No	NA	NA	GSR	1.01
PITG_00904	predicted protein	No	No	NA	NA	GSR	1.01

Appendix 4.2. List of genes with duplications in the sequenced *P. infestans* 06_3928A genome

Gene ID	Annotation	Secreted	Core ortholog	Effector type	RXLR family	Intergenic distance	CNV
PITG_09132	predicted protein	No	No	NA	NA	InBtw	1.01
PITG_05803	cysteine protease family C48, putative	No	No	NA	NA	GSR	1.01
PITG_18660	predicted protein	No	No	NA	NA	Not	1.01
PITG_15538	hypothetical protein	No	No	NA	NA	GSR	1.01
PITG_14825	predicted protein	No	No	NA	NA	GDR	1.01
PITG_18383	hypothetical protein	No	No	NA	NA	GSR	1.01
PITG_05063	conserved hypothetical protein	No	No	NA	NA	InBtw	1.00
PITG_17522	conserved hypothetical protein, contains CRN-like motif	No	No	NA	NA	GSR	1.00

Appendix 4.3. List of genes deleted in the sequenced *P. infestans* 06_3928A genome

Gene ID	Annotation	Secreted	Core ortholog	Effector type	RXLR family	Intergenic distance	Breadth of coverage (%)
PITG_12010		YES	NO	RXLR	RxLRfam47	GSR	0
PITG_15712		YES	NO	RXLR	RxLRsng162	InBtw	0
PITG_15718		YES	NO	RXLR	RxLRfam14	GSR	0
PITG_16663	*Avr1*	YES	NO	RXLR	RxLRfam2	GSR	0
PITG_17217		YES	NO	RXLR	RxLRfam45	InBtw	0
PITG_18215		YES	NO	RXLR	RxLRfam124	GSR	0
PITG_18221		YES	NO	RXLR	RxLRfam124	GSR	0
PITG_20857		YES	NO	RXLR	RxLRfam5	GSR	0
PITG_21681		YES	NO	RXLR	RxLRfam14	Not	0
PITG_22741		YES	NO	Elicitin	NA	GSR	0
PITG_23059		YES	NO	SCR	NA	GSR	0
PITG_23138		YES	NO	NA	NA	GSR	0
PITG_23228		YES	NO	NLP	NA	Not	0
PITG_23231		YES	NO	RXLR	RxLRfam54	Not	0
PITG_01012		NO	NO	NA	NA	GSR	0
PITG_05417		NO	NO	NA	NA	InBtw	0
PITG_05711		NO	NO	NA	NA	GSR	0
PITG_07583		NO	NO	NA	NA	GDR	0
PITG_08855		NO	NO	NA	NA	InBtw	0
PITG_08856		NO	NO	NA	NA	GDR	0
PITG_11053		NO	NO	NA	NA	GSR	0
PITG_12447		NO	YES	NA	NA	InBtw	0
PITG_12735		NO	NO	NA	NA	GSR	0
PITG_13504		NO	NO	NA	NA	InBtw	0
PITG_13532		NO	NO	NA	NA	GSR	0
PITG_15708		NO	NO	NA	NA	GSR	0
PITG_15714		NO	NO	NA	NA	GSR	0
PITG_16702		NO	NO	NA	NA	InBtw	0
PITG_17317		NO	NO	NA	NA	GSR	0
PITG_17320		NO	NO	NA	NA	InBtw	0
PITG_17574		NO	NO	NA	NA	InBtw	0
PITG_17816		NO	NO	CRINKLER	NA	GSR	0
PITG_17820		NO	NO	NA	NA	GSR	0
PITG_18218		NO	NO	NA	NA	GSR	0
PITG_18675		NO	NO	RXLR	RxLRfam5	GSR	0
PITG_18697		NO	NO	NA	NA	InBtw	0
PITG_19566		NO	NO	NA	NA	InBtw	0
PITG_20080		NO	NO	NA	NA	Not	0
PITG_20137		NO	NO	NA	NA	Not	0
PITG_20421		NO	NO	NA	NA	GSR	0
PITG_20858		NO	NO	NA	NA	InBtw	0
PITG_21438		NO	NO	NA	NA	Not	0
PITG_22272		NO	NO	NA	NA	Not	0
PITG_22379		NO	NO	NA	NA	Not	0
PITG_22420		NO	NO	NA	NA	Not	0
PITG_22422		NO	NO	NA	NA	Not	0
PITG_23107		NO	NO	NLP	NA	GSR	0

Appendix 4.4. List of candidate assembled RXLRs from unmapped reads of *P. infestans* 06_3928A genome

Pex ID	Protein length (aa)	HMM score	Signal peptide length (aa)	Full length	RXLR starts at aa position	RXLR-EER motif	Similarity in *P. infestans* T30-4	Amino acid sequence
Pex644	188	0.993	22	Yes	43	RFLR-EER	PITG_22798, RxLRsng233	MRRCYILIAIAVVLSGIASVVADSSQDKLMAV EGDQTTGTVNRFLRRDDELSAENTEERIVA GDIPLSARMINNIYKVEKRIVDPKLADELLEK PGLKTLKTHLDAALPYSERAKVFERWHAD GVDPSSITKALKVHPAIAKKYNTVSTMYDLY VKSAAIKRLTELKRKSDNDLADAVRLKRQRI NEZ
Pex50259	154	0.997	21	No	40	RSLR-EER	NA	MHLRNALVWVVTTLLIGSVASDHPTVFQHF NGKVNALSSRSLRLHEERGIPVSTIANIKGM LTSKRVSDKTLDSWRKAGKTADKVFVWLSL GRGKGELFDNPNFAKWVKYVDDLSASHPE RKSSISTLTSYYDDEPLSKMIIAAQKNPDTR ALA
Pex30588	137	0.999	21	No	51	RSLR-XXX	NA	MRRSSILYVAAVALCISFCDAASAATNSEFS PIMPFGTLQSAYSTALTSTRSLRGSKRDDD NKDMDFVQENRAGIQLTHIDDLLKQLALNE KMVLQNLNKFDDDLMRKLRQNPSWARTIL RWKDRDLHPTQVAAILN
Pex46622	126	0.999	20	Yes	41	RLLR-EER	(PITG_09739, PITG_09773) RxLRfam6	MRISQAVVVVTVAFLASSEALSTRMDDKVS KVATHDGPSQRLLRIHHTAIEDEDDSEERGL KEKDFKRLAVYADELGINVEKATKNTAYLRE VADEYAKYKSLLNQLIKKRKSKGSPMITYEH HGZ
Pex15083 *	117	0.977	20	Yes	49	RLLR-EER	(PITG_22870, PITG_08943) *Avr2*, RxLRfam7	MRLAYIFAVTMAGALPYCNALHAAPGAKAL NKIKTFPDFAAPSRMDGNRLLRRVDNEESE TEEERGFNLKDTLKKLNPIKAAGKAKDKAK EVTEKITDADWKKLVNYLQSKGNKRSZ
Pex14182	111	0.998	21	Yes	43	RFLR-EER	NA	MRGVETILTAVLCILCGTTDAAMTSDETIAAS VATKNGVLAKRFLRAQGPPDEERGRLKDV FEKVKRLARYNKWIFSDKSPDWVDKKYPQ FSQGYEKFWENRLVGGGKYAZ

* Pex15083 was identified in this study as a candidate assembled RXLR effector. Pex15083 amino acid sequence corresponds to the Avirulence protein AVR2 variant in *P. infestans* 06_3928A isolate (Gilroy et al., 2011).

Appendix 4.5. List of 4934 *Phytophthora infestans* genes that are induced during infection on potato in the strains 06_3928A, T30-4 and NL07434 (See attached CD)

Appendix 4.6. List of genes showing an extended induction period of 2 and 3 dpi on potato in *P. infestans* 06_3928A isolate

Gene ID	Annotation	Core ortholog	Effecto type	RXLR family	Inter-genic distance
PITG_23077	small cysteine-rich protein SCR91	No	Small Cys Rich	NA	InBtw
PITG_23123	small cysteine rich protein SCR50	No	Small Cys Rich	NA	GSR
PITG_11450	conserved hypothetical protein	Yes	Small Cys Rich	NA	InBtw
PITG_07529	conserved hypothetical protein	No	Small Cys Rich	NA	GDR
PITG_23156	small cysteine rich protein SCR58	No	Small Cys Rich	NA	GSR
PITG_09216	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam55	GSR
PITG_18683	avrblb2 family secreted RxLR effector peptide, putative	No	RXLR	RxLRfam5	GSR
PITG_22089	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam18	Not
PITG_05911	secreted RxLR effector peptide (Avh9.1), putative	No	RXLR	RxLRfam18	InBtw
PITG_05912	secreted RxLR effector peptide (Avh9.1), putative	No	RXLR	RxLRfam18	GSR
PITG_20300	avrblb2 family secreted RxLR effector peptide, putative	No	RXLR	RxLRfam5	GSR
PITG_20303	avrblb2 family secreted RxLR effector peptide, putative	No	RXLR	RxLRfam5	Not
PITG_04090	avrblb2 family secreted RxLR effector peptide, putative	No	RXLR	RxLRfam5	GSR
PITG_09732	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam1	GSR
PITG_04085	avrblb2 family secreted RxLR effector peptide, putative	No	RXLR	RxLRfam5	InBtw
PITG_20301	avrblb2 family secreted RxLR effector peptide, putative	No	RXLR	RxLRfam5	GSR
PITG_10654	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam46	GSR
PITG_15278	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam1	InBtw
PITG_02860	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam80	GSR
PITG_22547	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam97	Not
PITG_16705	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam1	GSR
PITG_06087	secreted RxLR effector peptide, putative	Yes	RXLR	RxLRfam87	InBtw
PITG_21388	avrblb1 secreted RxLR effector peptide, ipi01	No	RXLR	RxLRfam54	Not
PITG_14371	secreted RxLR effector peptide, avr3a	No	RXLR	RxLRfam58	GSR
PITG_16294	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam97	GSR
PITG_00582	secreted RxLR effector peptide, putative	No	RXLR	RxLRsng212	GSR
PITG_21740	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam1	Not
PITG_07550	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam117	GSR
PITG_11947	secreted RxLR effector peptide, putative	Yes	RXLR	RxLRsng164	GSR
PITG_15110	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam1	InBtw
PITG_06478	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam16	GSR
PITG_15039	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam1	GSR
PITG_04314	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam49	GSR
PITG_12737	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam43	GSR
PITG_04266	secreted RxLR effector peptide, putative	No	RXLR	RxLRsng248	InBtw
PITG_04089	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam5	GSR
PITG_22648	RXLR effector family protein, putative	No	RXLR	NA	Not
PITG_22922	secreted RxLR effector peptide, putative	Yes	RXLR	RxLRfam2	InBtw
PITG_23226	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam100	Not
PITG_17316	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam1	InBtw
PITG_21362	secreted RxLR effector peptide, putative, 3' partial	No	RXLR	RxLRfam57	GSR
PITG_17309	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam1	InBtw
PITG_14368	avr3a family secreted RxLR effector peptide, putative	No	RXLR	RxLRfam58	GSR
PITG_06099	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam36	GSR
PITG_23015	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam100	GSR
PITG_15930	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam2	InBtw
PITG_10232	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam69	GSR
PITG_15679	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam23	GSR
PITG_10540	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam57	InBtw
PITG_16427	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam9	InBtw

Appendix 4.6. List of genes showing an extended induction period of 2 and 3 dpi on potato in *P. infestans* 06_3928A isolate

Gene ID	Annotation	Core ortholog	Effecto type	RXLR family	Inter-genic distance
PITG_18670	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam5	InBtw
PITG_22804	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam27	GSR
PITG_22870	avr2 secreted RxLR effector peptide, putative	No	RXLR	RxLRfam7	GSR
PITG_13093	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam38	InBtw
PITG_14443	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam69	InBtw
PITG_17063	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam45	Not
PITG_22604	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam5	Not
PITG_15753	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam38	GSR
PITG_14787	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam6	GSR
PITG_05846	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam23	GSR
PITG_20336	secreted RxLR effector peptide, 3' partial	No	RXLR	RxLRfam9	Not
PITG_14783	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam6	GSR
PITG_01934	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam6	GSR
PITG_22926	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam52	GSR
PITG_05910	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam52	InBtw
PITG_23131	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam128	GSR
PITG_16233	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam9	InBtw
PITG_08174	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam19	InBtw
PITG_06094	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam36	GSR
PITG_04049	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam67	InBtw
PITG_07555	secreted RxLR effector peptide, putative	No	RXLR	RxLRsng247	GSR
PITG_22757	secreted RxLR effector peptide, putative	No	RXLR	RxLRsng203	GSR
PITG_05750	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam29	InBtw
PITG_23035	secreted RxLR effector peptide, putative	Yes	RXLR	RxLRfam1	InBtw
PITG_12731	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam1	GSR
PITG_21190	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam2	Not
PITG_23230	secreted RxLR effector peptide, putative	No	RXLR	RxLRfam9	Not
PITG_00774	secreted RxLR effector peptide, putative	No	RXLR	RxLRsng199	GSR
PITG_23076	NPP1-like protein	No	NLP	NA	GSR
PITG_09716	NPP1-like protein	Yes	NLP	NA	GSR
PITG_04248	NPP1-like protein	No	NLP	NA	InBtw
PITG_19938	NPP1-like protein	No	NLP	NA	InBtw
PITG_04208	NPP1-like protein	No	NLP	NA	InBtw
PITG_22668	NPP1-like protein	No	NLP	NA	InBtw
PITG_22734	NPP1-like protein, 3' partial	No	NLP	NA	GSR
PITG_12139	conserved hypothetical protein	No	NA	NA	InBtw
PITG_11060	conserved hypothetical protein	Yes	NA	NA	InBtw
PITG_14583	conserved hypothetical protein	Yes	NA	NA	GDR
PITG_11916	conserved hypothetical protein	No	NA	NA	GSR
PITG_04202	conserved hypothetical protein	No	NA	NA	GSR
PITG_07143	catalase-peroxidase, putative	No	NA	NA	GSR
PITG_02909	carbohydrate-binding protein, putative	Yes	NA	NA	GSR
PITG_22562	croquemort-like mating protein, putative	Yes	NA	NA	Not
PITG_11891	conserved hypothetical protein	No	NA	NA	InBtw
PITG_18224	conserved hypothetical protein	No	NA	NA	GSR
PITG_07303	carbohydrate-binding protein, putative	Yes	NA	NA	GDR
PITG_18119	conserved hypothetical protein	No	NA	NA	InBtw
PITG_04213	conserved hypothetical protein	No	NA	NA	GSR
PITG_07586	conserved hypothetical protein	No	NA	NA	GSR
PITG_16959	transglutaminase elicitor M81D	Yes	NA	NA	GSR

Appendix 4.6. List of genes showing an extended induction period of 2 and 3 dpi on potato in *P. infestans* 06_3928A isolate

Gene ID	Annotation	Core ortholog	Effecto type	RXLR family	Inter-genic distance
PITG_04949	conserved hypothetical protein	Yes	NA	NA	InBtw
PITG_18923	putative GPI-anchored serine-rich hypothetical protein	Yes	NA	NA	InBtw
PITG_11755	putative GPI-anchored serine-threonine rich protein	Yes	NA	NA	InBtw
PITG_13785	conserved hypothetical protein	Yes	NA	NA	GSR
PITG_02930	berberine-like protein	No	NA	NA	GSR
PITG_06212	conserved hypothetical protein	No	NA	NA	GSR
PITG_03694	disulfide-isomerase, putative	Yes	NA	NA	GDR
PITG_11890	putative GPI-anchored serine-threonine rich protein	No	NA	NA	GDR
PITG_07833	similar to sea slug pheromone	No	NA	NA	InBtw
PITG_11883	conserved hypothetical protein	No	NA	NA	GDR
PITG_19245	conserved hypothetical protein	No	NA	NA	GSR
PITG_15771	hsp70-like protein	Yes	NA	NA	GDR
PITG_05156	secretory protein OPEL-like	Yes	NA	NA	GSR
PITG_10410	SCP-like extracellular protein	Yes	NA	NA	InBtw
PITG_11689	putative GPI-anchored serine-rich hypothetical protein	No	NA	NA	InBtw
PITG_20346	hypothetical protein	No	NA	NA	InBtw
PITG_22758	arabinofuranosidase	Yes	NA	NA	GSR
PITG_12666	conserved hypothetical protein	Yes	NA	NA	GSR
PITG_11459	conserved hypothetical protein	No	NA	NA	InBtw
PITG_18934	calreticulin precursor	No	NA	NA	InBtw
PITG_04385	putative GPI-anchored acidic serine-threonine rich	No	NA	NA	GSR
PITG_01985	iron/zinc purple acid phosphatase-like protein	No	NA	NA	InBtw
PITG_15170	conserved hypothetical protein	No	NA	NA	Not
PITG_06325	conserved hypothetical protein	Yes	NA	NA	InBtw
PITG_22899	secreted peptide candidate, ORF supported by proteomics	No	NA	NA	GSR
PITG_07249	conserved hypothetical protein	Yes	NA	NA	GDR
PITG_06170	conserved hypothetical protein	No	NA	NA	Not
PITG_11340	conserved hypothetical protein	No	NA	NA	InBtw
PITG_13157	hypothetical protein (PITT_13157)	No	NA	NA	GSR
PITG_10972	thioredoxin-like protein	Yes	NA	NA	GDR
PITG_00035	60S ribosomal export protein NMD3, putative	Yes	NA	NA	GDR
PITG_21363	putative GPI-anchored serine-threonine rich protein	No	NA	NA	Not
PITG_11936	conserved hypothetical protein	Yes	NA	NA	InBtw
PITG_05660	putative GPI-anchored serine rich tenascin-like glycoprotein	Yes	NA	NA	InBtw
PITG_11603	peptidyl-prolyl cis-trans isomerase CYP19-4 precursor	No	NA	NA	InBtw
PITG_07032	conserved hypothetical protein	Yes	NA	NA	GDR
PITG_11685	conserved hypothetical protein	Yes	NA	NA	GSR
PITG_02964	carbohydrate-binding protein, putative	Yes	NA	NA	GSR
PITG_06134	conserved hypothetical protein	No	NA	NA	InBtw
PITG_05400	conserved hypothetical protein	No	NA	NA	GSR
PITG_11271	conserved hypothetical protein	No	NA	NA	GDR
PITG_18118	conserved hypothetical protein	No	NA	NA	InBtw
PITG_14939	serine/threonine protein kinase	Yes	NA	NA	GSR
PITG_06994	Phospholipase?D,Pi-sPLD-like-7	Yes	NA	NA	InBtw
PITG_01702	conserved hypothetical protein	No	NA	NA	GDR
PITG_01058	conserved hypothetical protein	Yes	NA	NA	InBtw
PITG_20795	ribosomal protein	Yes	NA	NA	GDR
PITG_07836	conserved hypothetical protein	Yes	NA	NA	GSR
PITG_02305	conserved hypothetical protein	Yes	NA	NA	GDR
PITG_14518	conserved hypothetical protein	No	NA	NA	GSR

Appendix 4.6. List of genes showing an extended induction period of 2 and 3 dpi on potato in *P. infestans* 06_3928A isolate

Gene ID	Annotation	Core ortholog	Effecto type	RXLR family	Inter-genic distance
PITG_06984	stromal cell-derived factor 2 precursor	No	NA	NA	InBtw
PITG_10544	putative GPI-anchored acidic protein	Yes	NA	NA	GSR
PITG_04386	HAM34-like putative membrane protein	No	NA	NA	InBtw
PITG_05579	catalase-peroxidase, putative	No	NA	NA	InBtw
PITG_02910	conserved hypothetical protein	No	NA	NA	InBtw
PITG_07843	protein disulfide-isomerase, putative	Yes	NA	NA	GSR
PITG_14720	aldose 1-epimerase, putative	No	NA	NA	GSR
PITG_11898	conserved hypothetical protein	Yes	Enzyme Inhibitor	NA	InBtw
PITG_13636	trypsin protease GIP-like	No	Enzyme Inhibitor	NA	GSR
PITG_13680	chymotrypsin, serine protease family S01A, putative	No	Enzyme Inhibitor	NA	InBtw
PITG_06175	conserved hypothetical protein	Yes	Enzyme Inhibitor	NA	GDR
PITG_07452	protease inhibitor Epi12	No	Enzyme Inhibitor	NA	InBtw
PITG_13671	glucanase inhibitor protein 3	No	Enzyme Inhibitor	NA	InBtw
PITG_05440	protease inhibitor Epi6	No	Enzyme Inhibitor	NA	GSR
PITG_09173	protease inhibitor EpiC2B	No	Enzyme Inhibitor	NA	GSR
PITG_05437	Epi6-like protease inhibitor	No	Enzyme Inhibitor	NA	GSR
PITG_00058	protease inhibitor EpiC4	Yes	Enzyme Inhibitor	NA	GSR
PITG_09175	protease inhibitor EpiC2A	No	Enzyme Inhibitor	NA	GSR
PITG_01369	protease inhibitor Epi2	No	Enzyme Inhibitor	NA	GDR
PITG_22936	Epi2-like protease inhibitor	No	Enzyme Inhibitor	NA	GSR
PITG_13638	glucanase inhibitor protein 1	No	Enzyme Inhibitor	NA	GSR
PITG_18117	conserved hypothetical protein	No	Enzyme hydrolase	NA	Not
PITG_04125	glycosyl transferase, putative	Yes	Enzyme hydrolase	NA	GDR
PITG_18396	conserved hypothetical protein	No	Enzyme hydrolase	NA	InBtw
PITG_10637	conserved hypothetical protein	No	Enzyme hydrolase	NA	InBtw
PITG_08944	endoglucanase, putative	Yes	Enzyme hydrolase	NA	GSR
PITG_15239	serine protease family S33, putative	No	Enzyme hydrolase	NA	InBtw
PITG_04158	glycoside hydrolase, putative	No	Enzyme hydrolase	NA	GSR
PITG_01029	pectinesterase, putative	No	Enzyme hydrolase	NA	GSR
PITG_06788	exoglucanase 1 precursor	Yes	Enzyme hydrolase	NA	GSR
PITG_16991	cell 12A endoglucanase	Yes	Enzyme hydrolase	NA	GSR
PITG_02545	pectinesterase, putative	No	Enzyme hydrolase	NA	GSR
PITG_14237	glycoside hydrolase, putative	No	Enzyme hydrolase	NA	GSR
PITG_17507	glucosylceramidase, putative	No	Enzyme hydrolase	NA	InBtw
PITG_02700	serine protease family S01A, putative	No	Enzyme hydrolase	NA	GSR
PITG_08912	pectinesterase, putative	No	Enzyme hydrolase	NA	GSR
PITG_04135	glycoside hydrolase, putative	Yes	Enzyme hydrolase	NA	GSR
PITG_04141	glycoside hydrolase, putative	No	Enzyme hydrolase	NA	GSR
PITG_17501	glucosylceramidase, putative	No	Enzyme hydrolase	NA	GDR
PITG_04123	glycoside hydrolase, putative	Yes	Enzyme hydrolase	NA	GDR
PITG_07720	calcineurin-like phosphoesterase, putative	No	Enzyme hydrolase	NA	GDR
PITG_16958	transglutaminase elicitor-like protein	Yes	Elicitins	NA	GSR
PITG_16956	M81 transglutaminase-like protein	Yes	Elicitins	NA	InBtw
PITG_05339	elicitor-like transglutaminase M81-like protein	No	Elicitins	NA	InBtw

APPENDIX 5: Gene expression analysis of *Puccinia monoica* pseudoflowers

Appendix 5.1. List of 948 significantly regulated genes in *Puccinia monoica* induced pseudoflowers ('Pf') compared to uninfected *Boechera stricta* stem and leaves ('SL') (See attached CD)

Appendix 5.2. List of 859 significantly regulated genes in uninfected *Boechera stricta* flowers ('F') compared to uninfected *B. stricta* stem and leaves ('SL') (See attached CD)

Appendix 5.3. Gene ontology biological processes (GOBP) enriched in *Puccinia monoica*-induced pseudoflowers ('Pf') compared to uninfected *Boechera stricta* stem and leaves ('SL')

GOBP*	Corrected P-value	Description	No. of genes	Average gene expression Log2 ('Pf' / 'SL')
9834	7.26E-09	secondary cell wall biogenesis	11	-2.0296
19748	7.26E-09	secondary metabolic process	36	-0.9717
48513	1.38E-08	organ development	30	0.1185
48731	1.38E-08	system development	30	0.1185
50896	1.38E-08	response to stimulus	137	-0.5375
48507	4.94E-08	meristem development	9	0.0097
45962	5.35E-08	positive regulation of development, heterochronic	2	1.2089
6575	1.17E-07	cellular amino acid derivative metabolic process	27	-1.1481
32501	1.31E-07	multicellular organismal process	67	-0.2365
10410	1.60E-07	hemicellulose metabolic process	8	-1.6982
45491	1.60E-07	xylan metabolic process	8	-1.6982
48367	1.71E-07	shoot development	19	0.3987
9908	2.01E-07	flower development	13	-0.0028
22621	2.01E-07	shoot system development	19	0.3987
50793	2.01E-07	regulation of developmental process	15	0.2866
9698	2.05E-07	phenylpropanoid metabolic process	20	-1.4542
10383	2.09E-07	cell wall polysaccharide metabolic process	9	-1.6948
70882	2.09E-07	cellular cell wall organization or biogenesis	17	-1.7791
10382	2.10E-07	cellular cell wall macromolecule metabolic process	8	-1.7661
10413	2.41E-07	glucuronoxylan metabolic process	7	-1.7802
10417	2.41E-07	glucuronoxylan biosynthetic process	7	-1.7802
45492	2.41E-07	xylan biosynthetic process	7	-1.7802
42546	2.74E-07	cell wall biogenesis	16	-1.9486
9809	5.42E-07	lignin biosynthetic process	9	-1.3629
9808	5.80E-07	lignin metabolic process	11	-1.4886
7275	1.06E-06	multicellular organismal development	63	-0.2501
10073	1.06E-06	meristem maintenance	6	-0.0005
32502	1.06E-06	developmental process	69	-0.1377
44038	1.06E-06	cell wall macromolecule biosynthetic process	7	-1.7802
70589	1.06E-06	cellular component macromolecule biosynthetic process	7	-1.7802
70592	1.06E-06	cell wall polysaccharide biosynthetic process	7	-1.7802
9620	1.85E-06	response to fungus	17	-0.6760
9699	1.87E-06	phenylpropanoid biosynthetic process	16	-1.3574
42398	1.92E-06	cellular amino acid derivative biosynthetic process	21	-1.2742
6725	2.06E-06	cellular aromatic compound metabolic process	27	-1.0786
34637	2.18E-06	cellular carbohydrate biosynthetic process	21	-1.6444
40034	2.18E-06	regulation of development, heterochronic	3	0.4309
5975	4.22E-06	carbohydrate metabolic process	50	-0.9893
16051	4.73E-06	carbohydrate biosynthetic process	21	-1.6444
71554	4.99E-06	cell wall organization or biogenesis	26	-1.4073
10016	7.42E-06	shoot morphogenesis	12	0.3550
48437	7.42E-06	floral organ development	9	-0.3187
9611	1.13E-05	response to wounding	16	-0.7025
19438	1.18E-05	aromatic compound biosynthetic process	19	-1.2662
48608	1.54E-05	reproductive structure development	29	-0.1501
9791	1.59E-05	post-embryonic development	34	-0.2504
44036	1.60E-05	cell wall macromolecule metabolic process	10	-1.3853
6519	2.11E-05	cellular amino acid and derivative metabolic process	36	-0.9924
9638	2.32E-05	phototropism	1	1.1825
10051	2.42E-05	xylem and phloem pattern formation	6	-0.9428
6950	3.22E-05	response to stress	85	-0.7716

Appendix 5.3. Gene ontology biological processes (GOBP) enriched in *Puccinia monoica*-induced pseudoflowers ('Pf') compared to uninfected *Boechera stricta* stem and leaves ('SL')

GOBP*	Corrected P-value	Description	No. of genes	Average gene expression Log2 ('Pf' / 'SL')
48827	3.22E-05	phyllome development	12	0.5688
3	3.37E-05	reproduction	36	-0.3832
22414	3.52E-05	reproductive process	35	-0.3338
9888	4.27E-05	tissue development	12	-0.1183
3002	4.54E-05	regionalization	9	-0.6236
48438	5.62E-05	floral whorl development	7	-0.3629
9832	7.39E-05	plant-type cell wall biogenesis	11	-2.0296
48856	7.39E-05	anatomical structure development	46	0.0091
3006	7.96E-05	reproductive developmental process	31	-0.2597
9719	9.63E-05	response to endogenous stimulus	39	0.0692
7389	1.01E-04	pattern specification process	10	-0.6792
80060	1.04E-04	integument development	2	-1.0944
30154	1.22E-04	cell differentiation	12	-0.0119
44262	1.24E-04	cellular carbohydrate metabolic process	31	-1.3257
51707	1.26E-04	response to other organism	31	-0.3976
50832	1.29E-04	defense response to fungus	11	-0.3262
9607	1.31E-04	response to biotic stimulus	32	-0.3468
42221	1.37E-04	response to chemical stimulus	74	-0.4413
6952	1.50E-04	defense response	35	-0.6848
9725	1.59E-04	response to hormone stimulus	35	0.3293
33692	1.65E-04	cellular polysaccharide biosynthetic process	13	-1.8785
9628	2.66E-04	response to abiotic stimulus	46	-0.5457
271	3.22E-04	polysaccharide biosynthetic process	13	-1.8785
9606	3.61E-04	tropism	2	1.3256
10074	3.61E-04	maintenance of meristem identity	3	-0.3718
9887	4.07E-04	organ morphogenesis	9	0.4843
71669	4.36E-04	plant-type cell wall organization or biogenesis	15	-1.6102
9637	4.86E-04	response to blue light	4	0.6724
48825	4.96E-04	cotyledon development	2	1.2813
10077	5.03E-04	maintenance of inflorescence meristem identity	1	-1.1793
10158	5.30E-04	abaxial cell fate specification	1	2.2328
48467	5.34E-04	gynoecium development	5	-0.2243
44281	5.40E-04	small molecule metabolic process	65	-0.7654
48569	6.11E-04	post-embryonic organ development	9	-0.3187
44264	7.28E-04	cellular polysaccharide metabolic process	15	-1.7243
9416	7.69E-04	response to light stimulus	18	-0.3375
48869	7.74E-04	cellular developmental process	17	0.3447
1708	7.78E-04	cell fate specification	3	-0.0553
5976	8.07E-04	polysaccharide metabolic process	16	-1.6868
10033	9.24E-04	response to organic substance	44	-0.0247
10075	9.89E-04	regulation of meristem growth	3	0.3708
9314	1.37E-03	response to radiation	18	-0.3375
19827	1.59E-03	stem cell maintenance	3	-0.3718
48864	1.59E-03	stem cell development	3	-0.3718
9855	1.62E-03	determination of bilateral symmetry	2	-1.0944
51704	1.62E-03	multi-organism process	34	-0.5033
48509	1.90E-03	regulation of meristem development	4	0.5437
48863	1.90E-03	stem cell differentiation	3	-0.3718
9850	2.04E-03	auxin metabolic process	4	0.0588
6796	2.04E-03	phosphate metabolic process	50	-0.7790
6793	2.08E-03	phosphorus metabolic process	50	-0.7790

Appendix 5.3. Gene ontology biological processes (GOBP) enriched in *Puccinia monoica*-induced pseudoflowers ('Pf') compared to uninfected *Boechera stricta* stem and leaves ('SL')

GOBP*	Corrected P-value	Description	No. of genes	Average gene expression Log2 ('Pf' / 'SL')
9684	2.29E-03	indoleacetic acid biosynthetic process	2	1.3252
9739	2.39E-03	response to gibberellin stimulus	6	-0.0639
44283	2.55E-03	small molecule biosynthetic process	35	-0.9599
10817	2.56E-03	regulation of hormone levels	8	0.6020
9653	2.69E-03	anatomical structure morphogenesis	19	0.4360
45596	2.69E-03	negative regulation of cell differentiation	3	-0.3718
9639	2.77E-03	response to red or far red light	10	-0.3259
48481	3.08E-03	ovule development	3	-0.3951
9683	3.23E-03	indoleacetic acid metabolic process	2	1.3252
9718	3.23E-03	anthocyanin biosynthetic process	4	-0.7403
9799	3.37E-03	specification of symmetry	2	-1.0944
48366	3.92E-03	leaf development	11	0.5210
9311	4.15E-03	oligosaccharide metabolic process	4	-0.9529
10476	4.45E-03	gibberellin mediated signaling pathway	1	0.9870
45165	4.45E-03	cell fate commitment	3	-0.0553
9631	4.79E-03	cold acclimation	2	0.5874
9965	4.79E-03	leaf morphogenesis	7	0.4722
65007	4.83E-03	biological regulation	100	-0.2228
48440	5.05E-03	carpel development	4	-0.6475
6468	5.95E-03	protein amino acid phosphorylation	42	-0.8386
9958	5.95E-03	positive gravitropism	2	1.3256
10076	5.95E-03	maintenance of floral meristem identity	2	-1.2920
42445	5.95E-03	hormone metabolic process	6	0.4433
9694	7.04E-03	jasmonic acid metabolic process	5	-0.7853
9415	7.51E-03	response to water	13	-0.5962
9312	7.96E-03	oligosaccharide biosynthetic process	3	-0.8381
6569	9.05E-03	tryptophan catabolic process	1	1.1816
46218	9.05E-03	indolalkylamine catabolic process	1	1.1816
48638	9.05E-03	regulation of developmental growth	3	0.3708
80006	9.05E-03	internode patterning	1	-1.1793
10050	9.41E-03	vegetative phase change	1	-1.1250
46283	9.93E-03	anthocyanin metabolic process	4	-0.7403
48878	1.25E-02	chemical homeostasis	7	1.0121
51239	1.39E-02	regulation of multicellular organismal process	9	0.4580
10047	1.41E-02	fruit dehiscence	3	-1.3383
10120	1.41E-02	camalexin biosynthetic process	2	-0.0615
52317	1.41E-02	camalexin metabolic process	2	-0.0615
51093	1.44E-02	negative regulation of developmental process	5	-0.2226
6955	1.48E-02	immune response	16	-1.0854
9737	1.48E-02	response to abscisic acid stimulus	16	0.0376
42446	1.48E-02	hormone biosynthetic process	4	1.2688
48522	1.56E-02	positive regulation of cellular process	7	-1.0347
71495	1.62E-02	cellular response to endogenous stimulus	11	-0.6177
10218	1.63E-02	response to far red light	4	-0.2322
45449	1.63E-02	regulation of transcription	48	-0.3222
16137	1.74E-02	glycoside metabolic process	7	-1.2637
31407	1.78E-02	oxylipin metabolic process	5	-0.7853
45087	1.78E-02	innate immune response	15	-1.0813
9700	1.84E-02	indole phytoalexin biosynthetic process	2	-0.0615
42431	1.84E-02	indole metabolic process	2	-0.0615
46217	1.84E-02	indole phytoalexin metabolic process	2	-0.0615

Appendix 5.3. Gene ontology biological processes (GOBP) enriched in *Puccinia monoica*-induced pseudoflowers ('Pf') compared to uninfected *Boechera stricta* stem and leaves ('SL')

GOBP*	Corrected P-value	Description	No. of genes	Average gene expression Log2 ('Pf' / 'SL')
48518	1.84E-02	positive regulation of biological process	8	-1.0651
52314	1.84E-02	phytoalexin metabolic process	2	-0.0615
52315	1.84E-02	phytoalexin biosynthetic process	2	-0.0615
50801	1.91E-02	ion homeostasis	5	0.0089
71310	1.91E-02	cellular response to organic substance	13	-0.7174
45595	1.92E-02	regulation of cell differentiation	3	-0.3718
30244	1.94E-02	cellulose biosynthetic process	5	-2.0681
2376	2.05E-02	immune system process	16	-1.0854
9889	2.06E-02	regulation of biosynthetic process	50	-0.3467
31326	2.06E-02	regulation of cellular biosynthetic process	50	-0.3467
19439	2.08E-02	aromatic compound catabolic process	3	-0.9757
5987	2.09E-02	sucrose catabolic process	1	-1.2972
10131	2.09E-02	sucrose catabolic process, using invertase or sucrose synthase	1	-1.2972
16098	2.09E-02	monoterpenoid metabolic process	1	-2.2188
16099	2.09E-02	monoterpenoid biosynthetic process	1	-2.2188
16310	2.09E-02	phosphorylation	42	-0.8386
23033	2.13E-02	signaling pathway	22	-0.6813
6833	2.29E-02	water transport	3	-1.1897
10223	2.29E-02	secondary shoot formation	2	-0.0584
10346	2.29E-02	shoot formation	2	-0.0584
16138	2.29E-02	glycoside biosynthetic process	6	-1.2581
42044	2.29E-02	fluid transport	3	-1.1897
5985	2.34E-02	sucrose metabolic process	3	-1.7267
9695	2.34E-02	jasmonic acid biosynthetic process	4	-1.3176
43687	2.34E-02	post-translational protein modification	48	-0.7452
9753	2.37E-02	response to jasmonic acid stimulus	10	-0.6280
10556	2.37E-02	regulation of macromolecule biosynthetic process	48	-0.3222
32870	2.37E-02	cellular response to hormone stimulus	9	-0.7312
50789	2.37E-02	regulation of biological process	83	-0.4678
9414	2.48E-02	response to water deprivation	11	-0.6493
30243	2.59E-02	cellulose metabolic process	5	-2.0681
9851	2.71E-02	auxin biosynthetic process	2	1.3252
70887	2.83E-02	cellular response to chemical stimulus	14	-0.7519
10588	2.88E-02	cotyledon vascular tissue pattern formation	1	1.4688
34754	2.88E-02	cellular hormone metabolic process	3	1.2422
48653	2.88E-02	anther development	2	-0.7094
65008	2.90E-02	regulation of biological quality	20	0.7351
42219	3.13E-02	cellular amino acid derivative catabolic process	3	-0.9757
19219	3.18E-02	regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	48	-0.3222
7167	3.39E-02	enzyme linked receptor protein signaling pathway	7	-0.6419
7169	3.39E-02	transmembrane receptor protein tyrosine kinase signaling pathway	7	-0.6419
10154	3.39E-02	fruit development	16	-0.1639
9891	3.42E-02	positive regulation of biosynthetic process	5	-1.2995
31328	3.42E-02	positive regulation of cellular biosynthetic process	5	-1.2995
9074	3.48E-02	aromatic amino acid family catabolic process	1	1.1816
42402	3.48E-02	cellular biogenic amine catabolic process	1	1.1816
9733	3.52E-02	response to auxin stimulus	13	0.2787
23052	3.59E-02	signaling	37	-0.8130
42435	3.90E-02	indole derivative biosynthetic process	3	0.4486
51171	3.91E-02	regulation of nitrogen compound metabolic process	48	-0.3222
31408	4.05E-02	oxylipin biosynthetic process	4	-1.3176

Appendix 5.3. Gene ontology biological processes (GOBP) enriched in *Puccinia monoica*-induced pseudoflowers ('Pf') compared to uninfected *Boechera stricta* stem and leaves ('SL')

GOBP*	Corrected P-value	Description	No. of genes	Average gene expression Log2 ('Pf' / 'SL')
48443	4.23E-02	stamen development	3	-0.9412
48466	4.23E-02	androecium development	3	-0.9412
80090	4.23E-02	regulation of primary metabolic process	50	-0.3467
9755	4.25E-02	hormone-mediated signaling pathway	9	-0.7312
9814	4.32E-02	defense response, incompatible interaction	6	-0.9447
16054	4.39E-02	organic acid catabolic process	6	-0.5641
32787	4.39E-02	monocarboxylic acid metabolic process	16	-0.4368
46395	4.39E-02	carboxylic acid catabolic process	6	-0.5641
9723	4.55E-02	response to ethylene stimulus	6	0.0319
50794	4.66E-02	regulation of cellular process	75	-0.4865
71555	4.66E-02	cell wall organization	11	-1.0742
40008	4.85E-02	regulation of growth	3	0.3708

*GOBP indicates Gene Ontology Biological Process.

Appendix 5.4. Gene ontology biological processes (GOBP) enriched in uninfected *Boechera stricta* flowers ('F') compared to uninfected *B. stricta* stem and leaves ('SL')

GOBP*	Corrected P-value	Description	No. of genes	Average gene expression Log2 ('F' / 'SL')
71554	6.63E-14	cell wall organization or biogenesis	40	-0.31237112
42546	4.00E-09	cell wall biogenesis	19	-2.077819204
9834	4.42E-09	secondary cell wall biogenesis	11	-2.851846798
32502	6.45E-09	developmental process	84	1.233875941
70882	9.89E-09	cellular cell wall organization or biogenesis	19	-2.077819204
32501	2.15E-08	multicellular organismal process	81	1.229187468
71555	2.91E-08	cell wall organization	24	0.748460814
9555	8.33E-08	pollen development	19	2.274689056
9832	8.33E-08	plant-type cell wall biogenesis	15	-1.983048995
42545	1.33E-07	cell wall modification	20	1.138246455
7275	1.73E-07	multicellular organismal development	77	1.291197399
9908	1.97E-07	flower development	18	2.62728838
10208	4.22E-07	pollen wall assembly	9	3.09091768
10927	4.22E-07	cellular component assembly involved in morphogenesis	9	3.09091768
3	5.77E-07	reproduction	51	1.836253181
48229	5.77E-07	gametophyte development	21	2.237435181
48437	6.11E-07	floral organ development	15	2.645401541
71669	6.36E-07	plant-type cell wall organization or biogenesis	20	-1.262365865
45229	6.36E-07	external encapsulating structure organization	10	2.418894349
10584	7.65E-07	pollen exine formation	8	3.171035085
48438	9.45E-07	floral whorl development	14	2.73396688
44281	1.62E-06	small molecule metabolic process	77	0.027827018
5975	1.96E-06	carbohydrate metabolic process	56	0.0372454
6575	2.27E-06	cellular amino acid derivative metabolic process	25	-0.8848358
9698	3.32E-06	phenylpropanoid metabolic process	19	-1.350741902
10382	5.00E-06	cellular cell wall macromolecule metabolic process	7	-2.653937843
271	5.02E-06	polysaccharide biosynthetic process	16	-1.985830041
6725	6.99E-06	cellular aromatic compound metabolic process	27	-0.581727701
10413	8.10E-06	glucuronoxylan metabolic process	6	-2.583126707
10417	8.10E-06	glucuronoxylan biosynthetic process	6	-2.583126707
45492	8.10E-06	xylan biosynthetic process	6	-2.583126707
10073	8.10E-06	meristem maintenance	5	2.603052073
10022	8.10E-06	meristem determinacy	1	4.03786358
33692	1.01E-05	cellular polysaccharide biosynthetic process	15	-2.249287221
10582	2.49E-05	floral meristem determinacy	1	4.03786358
19953	2.62E-05	sexual reproduction	11	2.457462473
48569	2.71E-05	post-embryonic organ development	16	2.398147706
44038	2.71E-05	cell wall macromolecule biosynthetic process	6	-2.583126707
70589	2.71E-05	cellular component macromolecule biosynthetic process	6	-2.583126707
70592	2.71E-05	cell wall polysaccharide biosynthetic process	6	-2.583126707
9638	2.71E-05	phototropism	1	-2.006163517
22414	3.11E-05	reproductive process	45	1.83644921
9808	3.31E-05	lignin metabolic process	10	-2.487958981
10383	3.76E-05	cell wall polysaccharide metabolic process	7	-2.653937843
19748	4.15E-05	secondary metabolic process	27	-0.574436615
6519	4.23E-05	cellular amino acid and derivative metabolic process	36	-0.682264229
10410	4.23E-05	hemicellulose metabolic process	6	-2.583126707
45491	4.23E-05	xylan metabolic process	6	-2.583126707
44262	4.81E-05	cellular carbohydrate metabolic process	34	-0.548190196
48440	4.81E-05	carpel development	8	2.912946126
48856	5.43E-05	anatomical structure development	61	1.614891646

Appendix 5.4. Gene ontology biological processes (GOBP) enriched in uninfected *Boechera stricta* flowers ('F') compared to uninfected *B. stricta* stem and leaves ('SL')

GOBP*	Corrected P-value	Description	No. of genes	Average gene expression Log2 ('F' / 'SL')
48507	5.73E-05	meristem development	7	2.073489687
44264	6.29E-05	cellular polysaccharide metabolic process	17	-1.942273663
3006	6.31E-05	reproductive developmental process	40	1.908237429
42398	7.66E-05	cellular amino acid derivative biosynthetic process	19	-0.650556779
44283	7.76E-05	small molecule biosynthetic process	42	0.233966447
5976	7.76E-05	polysaccharide metabolic process	18	-1.725145812
34637	8.63E-05	cellular carbohydrate biosynthetic process	19	-1.246893248
48481	9.05E-05	ovule development	6	2.57836137
48467	1.10E-04	gynoecium development	8	2.912946126
19438	1.15E-04	aromatic compound biosynthetic process	18	-0.213979549
44282	1.33E-04	small molecule catabolic process	17	-0.288767808
9699	1.40E-04	phenylpropanoid biosynthetic process	14	-0.97353524
9063	1.57E-04	cellular amino acid catabolic process	7	0.660126199
48608	1.91E-04	reproductive structure development	34	1.954558356
32787	2.04E-04	monocarboxylic acid metabolic process	24	0.865017325
50793	2.16E-04	regulation of developmental process	16	1.266492714
48513	2.24E-04	organ development	31	1.526128533
48731	2.29E-04	system development	31	1.526128533
9310	2.30E-04	amine catabolic process	7	0.660126199
65007	2.92E-04	biological regulation	121	0.532404299
19752	2.92E-04	carboxylic acid metabolic process	40	0.343823287
43436	2.92E-04	oxoacid metabolic process	40	0.343823287
6082	2.96E-04	organic acid metabolic process	40	0.343823287
16051	3.12E-04	carbohydrate biosynthetic process	21	-1.095018389
42180	4.01E-04	cellular ketone metabolic process	40	0.343823287
48646	4.87E-04	anatomical structure formation involved in morphogenesis	12	3.005879722
10254	5.30E-04	nectary development	1	4.03786358
10050	5.44E-04	vegetative phase change	1	-1.644007088
10158	5.44E-04	abaxial cell fate specification	1	3.206764462
6629	5.95E-04	lipid metabolic process	37	1.046003357
30244	5.95E-04	cellulose biosynthetic process	7	-2.730628922
10876	6.23E-04	lipid localization	15	1.094663758
44085	6.25E-04	cellular component biogenesis	32	-0.372681487
6631	6.78E-04	fatty acid metabolic process	17	1.142922173
9791	9.38E-04	post-embryonic development	36	1.602322497
9889	1.06E-03	regulation of biosynthetic process	63	0.646906003
31326	1.06E-03	regulation of cellular biosynthetic process	63	0.646906003
30243	1.06E-03	cellulose metabolic process	7	-2.730628922
16054	1.16E-03	organic acid catabolic process	9	0.564560412
46395	1.16E-03	carboxylic acid catabolic process	9	0.564560412
80090	1.37E-03	regulation of primary metabolic process	65	0.683159368
48518	1.38E-03	positive regulation of biological process	12	-0.635667359
48869	1.41E-03	cellular developmental process	25	2.153268681
9944	1.49E-03	polarity specification of adaxial/abaxial axis	3	2.844215397
6624	1.68E-03	vacuolar protein processing	1	-1.816975853
55114	1.86E-03	oxidation reduction	14	-0.952457885
65001	1.99E-03	specification of axis polarity	3	2.844215397
51239	2.45E-03	regulation of multicellular organismal process	11	0.930415936
48580	2.45E-03	regulation of post-embryonic development	8	0.401781098
31323	2.47E-03	regulation of cellular metabolic process	66	0.654239837
3002	2.64E-03	regionalization	8	1.538459604

Appendix 5.4. Gene ontology biological processes (GOBP) enriched in uninfected *Boechera stricta* flowers ('F') compared to uninfected *B. stricta* stem and leaves ('SL')

GOBP*	Corrected P-value	Description	No. of genes	Average gene expression Log2 ('F' / 'SL')
9809	2.81E-03	lignin biosynthetic process	6	-2.580936986
10556	3.28E-03	regulation of macromolecule biosynthetic process	60	0.753167936
19439	3.28E-03	aromatic compound catabolic process	4	-1.436415169
9943	3.28E-03	adaxial/abaxial axis specification	3	2.844215397
48443	3.34E-03	stamen development	7	2.681072074
48466	3.34E-03	androecium development	7	2.681072074
6820	3.49E-03	anion transport	9	0.580116527
44036	3.55E-03	cell wall macromolecule metabolic process	7	-2.653937843
80086	3.55E-03	stamen filament development	3	4.173611637
45449	4.10E-03	regulation of transcription	59	0.788148068
51171	4.13E-03	regulation of nitrogen compound metabolic process	62	0.784319856
44255	4.13E-03	cellular lipid metabolic process	26	0.638164707
6869	5.85E-03	lipid transport	13	0.548337787
48522	5.85E-03	positive regulation of cellular process	9	-0.34564395
9909	6.10E-03	regulation of flower development	5	0.120382863
42219	6.10E-03	cellular amino acid derivative catabolic process	4	-1.436415169
1708	6.10E-03	cell fate specification	2	4.309761859
19219	7.16E-03	regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	60	0.80635646
9955	7.46E-03	adaxial/abaxial pattern formation	3	2.844215397
44092	8.73E-03	negative regulation of molecular function	7	1.526050371
19222	8.80E-03	regulation of metabolic process	69	0.561361771
10565	8.95E-03	regulation of cellular ketone metabolic process	4	-0.666342224
6548	9.89E-03	histidine catabolic process	2	-1.465572919
9077	9.89E-03	histidine family amino acid catabolic process	2	-1.465572919
6569	9.89E-03	tryptophan catabolic process	1	2.274286584
46218	9.89E-03	indolalkylamine catabolic process	1	2.274286584
9250	1.02E-02	glucan biosynthetic process	8	-2.083053001
51094	1.04E-02	positive regulation of developmental process	4	-1.556202444
9637	1.04E-02	response to blue light	2	-1.626986169
7389	1.16E-02	pattern specification process	8	1.538459604
10468	1.23E-02	regulation of gene expression	62	0.681732173
48878	1.23E-02	chemical homeostasis	7	0.787952532
9798	1.44E-02	axis specification	3	2.844215397
10047	1.50E-02	fruit dehiscence	3	-1.514195909
9653	1.55E-02	anatomical structure morphogenesis	26	2.011403031
19216	1.65E-02	regulation of lipid metabolic process	3	-1.478332661
46700	1.66E-02	heterocycle catabolic process	4	0.185998538
6633	1.98E-02	fatty acid biosynthetic process	10	1.231893668
9719	2.00E-02	response to endogenous stimulus	36	0.603859361
10252	2.01E-02	auxin homeostasis	4	2.553540659
8284	2.09E-02	positive regulation of cell proliferation	2	0.1651172
45595	2.13E-02	regulation of cell differentiation	5	2.248610211
48582	2.13E-02	positive regulation of post-embryonic development	3	-1.505737586
60255	2.21E-02	regulation of macromolecule metabolic process	62	0.681732173
43086	2.22E-02	negative regulation of catalytic activity	7	1.526050371
50896	2.24E-02	response to stimulus	117	-0.392046318
6073	2.28E-02	cellular glucan metabolic process	10	-1.594376797
9804	2.43E-02	coumarin metabolic process	2	0.012152238
9805	2.43E-02	coumarin biosynthetic process	2	0.012152238
45165	2.43E-02	cell fate commitment	2	4.309761859
80110	2.43E-02	sporopollenin biosynthetic process	2	2.298256075

Appendix 5.4. Gene ontology biological processes (GOBP) enriched in uninfected *Boechera stricta* flowers ('F') compared to uninfected *B. stricta* stem and leaves ('SL')

GOBP*	Corrected P-value	Description	No. of genes	Average gene expression Log2 ('F' / 'SL')
44042	2.70E-02	glucan metabolic process	10	-1.594376797
31325	3.02E-02	positive regulation of cellular metabolic process	5	-0.645024244
6355	3.04E-02	regulation of transcription, DNA-dependent	32	1.037454326
65008	3.12E-02	regulation of biological quality	24	0.873680713
51252	3.20E-02	regulation of RNA metabolic process	32	1.037454326
9851	3.25E-02	auxin biosynthetic process	3	0.956714647
50794	3.32E-02	regulation of cellular process	86	0.431509347
9725	3.35E-02	response to hormone stimulus	32	0.678349626
9606	3.35E-02	tropism	1	-2.006163517
50789	3.40E-02	regulation of biological process	94	0.396032801
9911	3.40E-02	positive regulation of flower development	2	-1.434432689
43193	3.40E-02	positive regulation of gene-specific transcription	2	-0.103372574
51179	3.50E-02	localization	65	0.900175219
9893	3.53E-02	positive regulation of metabolic process	5	-0.645024244
65009	3.63E-02	regulation of molecular function	10	0.951955531
6098	3.64E-02	pentose-phosphate shunt	4	0.084777167
32989	3.83E-02	cellular component morphogenesis	17	2.328797347
9888	3.83E-02	tissue development	10	1.328334292
9733	3.86E-02	response to auxin stimulus	13	0.440793476
9891	3.88E-02	positive regulation of biosynthetic process	5	-0.645024244
31328	3.88E-02	positive regulation of cellular biosynthetic process	5	-0.645024244
6740	3.94E-02	NADPH regeneration	4	0.084777167
10193	3.94E-02	response to ozone	4	-0.888270706
46271	3.94E-02	phenylpropanoid catabolic process	3	-2.673315753
46274	3.94E-02	lignin catabolic process	3	-2.673315753
48638	3.94E-02	regulation of developmental growth	3	2.340108838
15711	3.94E-02	organic anion transport	2	-1.42828641
15800	3.94E-02	acidic amino acid transport	2	-1.42828641
46713	3.94E-02	boron transport	2	0.165458646
46890	3.94E-02	regulation of lipid biosynthetic process	2	-1.507895638
9074	3.94E-02	aromatic amino acid family catabolic process	1	2.274286584
42402	3.94E-02	cellular biogenic amine catabolic process	1	2.274286584
16053	3.95E-02	organic acid biosynthetic process	18	0.788756739
46394	3.95E-02	carboxylic acid biosynthetic process	18	0.788756739
6066	3.95E-02	alcohol metabolic process	14	0.190752049
42592	3.95E-02	homeostatic process	10	0.348348099
10033	4.09E-02	response to organic substance	42	0.505640104
6857	4.14E-02	oligopeptide transport	7	-0.734035415
15833	4.14E-02	peptide transport	7	-0.734035415
9850	4.16E-02	auxin metabolic process	3	0.956714647
50790	4.91E-02	regulation of catalytic activity	10	0.951955531

*GOBP indicates Gene Ontology Biological Process.

REFERENCES

Abe, M., Kobayashi, Y., Yamamoto, S., Daimon, Y., Yamaguchi, A., Ikeda, Y., Ichinoki, H., Notaguchi, M., Goto, K., and Araki, T. (2005). FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. Science *309*, 1052-1056.

Abrahams, S., Lee, E., Walker, A.R., Tanner, G.J., Larkin, P.J., and Ashton, A.R. (2003). The Arabidopsis TDS4 gene encodes leucoanthocyanidin dioxygenase (LDOX) and is essential for proanthocyanidin synthesis and vacuole development. Plant J *35*, 624-636.

Agriculture, U.D.o. (1960). Index of Plant Diseases in the United States. (US Department of Agriculture, Washington DC, 1960).

Agrios, G.N. (2005). Plant Pathology. Elsevier, San Diego, CA.

Aida, M., Ishida, T., Fukaki, H., Fujisawa, H., and Tasaka, M. (1997). Genes involved in organ separation in Arabidopsis: an analysis of the cup-shaped cotyledon mutant. Plant Cell *9*, 841-857.

Aked, J., and Hall, J.L. (1993). The Uptake of Glucose, Fructose and Sucrose into Pea Powdery Mildew (Erysiphe pisi DC) from the Apoplast of Pea Leaves. New Phytologist *123*, 277-282.

Allen, R.L., Bittner-Eddy, P.D., Grenville-Briggs, L.J., Meitz, J.C., Rehmany, A.P., Rose, L.E., and Beynon, J.L. (2004). Host-parasite coevolutionary conflict between Arabidopsis and downy mildew. Science *306*, 1957-1960.

Alvarez, J.P., Pekker, I., Goldshmidt, A., Blum, E., Amsellem, Z., and Eshed, Y. (2006). Endogenous and synthetic microRNAs stimulate simultaneous, efficient, and localized regulation of multiple targets in diverse species. Plant Cell *18*, 1134-1151.

Alvarez-Fernandez, M., Barrett, A.J., Gerhartz, B., Dando, P.M., Ni, J., and Abrahamson, M. (1999). Inhibition of mammalian legumain by some cystatins is due to a novel second reactive site. J Biol Chem *274*, 19195-19203.

Anastasiou, E., Kenz, S., Gerstung, M., MacLean, D., Timmer, J., Fleck, C., and Lenhard, M. (2007). Control of plant organ size by KLUH/CYP78A5-dependent intercellular signaling. Dev Cell *13*, 843-856.

Aranishi, F., and Mano, N. (2000). Antibacterial cathepsins in different types of ambicoloured Japanese flounder skin. . Fish Shellfish Immunol *10*, 87-89.

Armstrong, M.R., Whisson, S.C., Pritchard, L., Bos, J.I., Venter, E., Avrova, A.O., Rehmany, A.P., Bohme, U., Brooks, K., Cherevach, I.*, et al.* (2005). An ancestral oomycete locus contains late blight avirulence gene Avr3a, encoding a protein that is recognized in the host cytoplasm. Proc Natl Acad Sci U S A *102*, 7766-7771.

Bailey, K., Cevik, V., Holton, N., Byrne-Richardson, J., Sohn, K.H., Coates, M., Woods-Tor, A., Aksoy, H.M., Hughes, L., Baxter, L.*, et al.* (2011). Molecular cloning of ATR5(Emoy2) from Hyaloperonospora arabidopsidis, an avirulence determinant that triggers RPP5-

mediated defense in Arabidopsis. Mol Plant Microbe Interact *24*, 827-838.
Baldauf, S.L. (2003). The deep roots of eukaryotes. Science *300*, 1703-1706.
Baldauf, S.L. (2008). An overview of the phylogeny and diversity of eukaryotes Journal of Systematics and Evolution *46*, 263–273.
Bariola, P.A., MacIntosh, G.C., and Green, P.J. (1999). Regulation of S-like ribonuclease levels in Arabidopsis. Antisense inhibition of RNS1 or RNS2 elevates anthocyanin accumulation. Plant Physiol *119*, 331-342.
Baxter, L., Tripathy, S., Ishaque, N., Boot, N., Cabral, A., Kemen, E., Thines, M., Ah-Fong, A., Anderson, R., Badejoko, W.*, et al.* (2010). Signatures of adaptation to obligate biotrophy in the Hyaloperonospora arabidopsidis genome. Science *330*, 1549-1551.
Bennett, T., Sieberer, T., Willett, B., Booker, J., Luschnig, C., and Leyser, O. (2006). The Arabidopsis MAX pathway controls shoot branching by regulating auxin transport. Curr Biol *16*, 553-563.
Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., Hall, K.P., Evers, D.J., Barnes, C.L., Bignell, H.R.*, et al.* (2008). Accurate whole human genome sequencing using reversible terminator chemistry. Nature *456*, 53-59.
Berardini, T.Z., Mundodi, S., Reiser, L., Huala, E., Garcia-Hernandez, M., Zhang, P., Mueller, L.A., Yoon, J., Doyle, A., Lander, G.*, et al.* (2004). Functional annotation of the Arabidopsis genome using controlled vocabularies. Plant Physiol *135*, 745-755.
Bhattacharjee, S., Hiller, N.L., Liolios, K., Win, J., Kanneganti, T.D., Young, C., Kamoun, S., and Haldar, K. (2006). The malarial host-targeting signal is conserved in the Irish potato famine pathogen. Plos Pathog *2*, e50.
Birch, P.R., Boevink, P.C., Gilroy, E.M., Hein, I., Pritchard, L., and Whisson, S.C. (2008). Oomycete RXLR effectors: delivery, functional redundancy and durable disease resistance. Curr Opin Plant Biol *11*, 373-379.
Birch, P.R., Rehmany, A.P., Pritchard, L., Kamoun, S., and Beynon, J.L. (2006). Trafficking arms: oomycete effectors enter host plant cells. Trends Microbiol *14*, 8-11.
Bird, D., Beisson, F., Brigham, A., Shin, J., Greer, S., Jetter, R., Kunst, L., Wu, X., Yephremov, A., and Samuels, L. (2007). Characterization of Arabidopsis ABCG11/WBC11, an ATP binding cassette (ABC) transporter that is required for cuticular lipid secretion. Plant J *52*, 485-498.
Bischoff, V., Nita, S., Neumetzler, L., Schindelasch, D., Urbain, A., Eshed, R., Persson, S., Delmer, D., and Scheible, W.R. (2010). TRICHOME BIREFRINGENCE and its homolog AT5G01360 encode plant-specific DUF231 proteins required for cellulose biosynthesis in Arabidopsis. Plant Physiol *153*, 590-602.
Blair, J.E., Coffey, M.D., Park, S.Y., Geiser, D.M., and Kang, S. (2008). A multi-locus phylogeny for Phytophthora utilizing markers derived from complete genome sequences. Fungal Genet Biol *45*, 266-277.
Blakeslee, J.J., Bandyopadhyay, A., Lee, O.R., Mravec, J., Titapiwatanakun, B., Sauer, M., Makam, S.N., Cheng, Y., Bouchard, R., Adamec, J.*, et*

al. (2007). Interactions among PIN-FORMED and P-glycoprotein auxin transporters in Arabidopsis. Plant Cell *19*, 131-147.
Blakeslee, J.J., Bandyopadhyay, A., Peer, W.A., Makam, S.N., and Murphy, A.S. (2004). Relocalization of the PIN1 auxin efflux facilitator plays a role in phototropic responses. Plant Physiol *134*, 28-31.
Bohlmann, J., Martin, D., Oldham, N.J., and Gershenzon, J. (2000). Terpenoid secondary metabolism in Arabidopsis thaliana: cDNA cloning, characterization, and functional expression of a myrcene/(E)-beta-ocimene synthase. Arch Biochem Biophys *375*, 261-269.
Bolstad, B.M., Irizarry, R.A., Astrand, M., and Speed, T.P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics *19*, 185-193.
Bos, J.I., Armstrong, M.R., Gilroy, E.M., Boevink, P.C., Hein, I., Taylor, R.M., Zhendong, T., Engelhardt, S., Vetukuri, R.R., Harrower, B.*, et al.* (2010). Phytophthora infestans effector AVR3a is essential for virulence and manipulates plant immunity by stabilizing host E3 ligase CMPG1. Proc Natl Acad Sci U S A *107*, 9909-9914.
Bos, J.I., Kanneganti, T.D., Young, C., Cakir, C., Huitema, E., Win, J., Armstrong, M.R., Birch, P.R., and Kamoun, S. (2006). The C-terminal half of Phytophthora infestans RXLR effector AVR3a is sufficient to trigger R3a-mediated hypersensitivity and suppress INF1-induced cell death in Nicotiana benthamiana. Plant J *48*, 165-176.
Bouwmeester, K., van Poppel, P.M.J.A., and Govers, F. (2009). Chapter 5: Genome biology cracks enigmas of oomycete plant pathogens. Molecular aspects of plant disease resistance. Edited by Jane Parker© 2009 Blackwell Publishing Ltd. . Annual Plant Reviews *34*, 102–133.
Bouzidi, M.F., Parlange, F., Nicolas, P., and Mouzeyar, S. (2007). Expressed Sequence Tags from the oomycete Plasmopara halstedii, an obligate parasite of the sunflower. Bmc Microbiol *7*, 110.
Breitling, R., Armengaud, P., Amtmann, A., and Herzyk, P. (2004). Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. FEBS Lett *573*, 83-92.
Brown, D.M., Goubet, F., Wong, V.W., Goodacre, R., Stephens, E., Dupree, P., and Turner, S.R. (2007). Comparison of five xylan synthesis mutants reveals new insight into the mechanisms of xylan synthesis. Plant J *52*, 1154-1168.
Brown, D.M., Zeef, L.A., Ellis, J., Goodacre, R., and Turner, S.R. (2005). Identification of novel genes in Arabidopsis involved in secondary cell wall formation using expression profiling and reverse genetics. Plant Cell *17*, 2281-2295.
Brown, D.M., Zhang, Z., Stephens, E., Dupree, P., and Turner, S.R. (2009). Characterization of IRX10 and IRX10-like reveals an essential role in glucuronoxylan biosynthesis in Arabidopsis. Plant J *57*, 732-746.
Bustamante, C.D., Fledel-Alon, A., Williamson, S., Nielsen, R., Hubisz, M.T., Glanowski, S., Tanenbaum, D.M., White, T.J., Sninsky, J.J.,

Hernandez, R.D., *et al.* (2005). Natural selection on protein-coding genes in the human genome. Nature *437*, 1153-1157.
Cavalier-Smith, T. (2005). Economy, speed and size matter: evolutionary forces driving nuclear genome miniaturization and expansion. Ann Bot *95*, 147-175.
Champouret, N. (2010). Functional genomics of *Phytophthora infestans* effectors and *Solanum* resistance genes. PhD thesis Wageningen University.
Chaparro-Garcia, A., Wilkinson, R.C., Gimenez-Ibanez, S., Findlay, K., Coffey, M.D., Zipfel, C., Rathjen, J.P., Kamoun, S., and Schornack, S. (2011). The receptor-like kinase SERK3/BAK1 is required for basal resistance against the late blight pathogen phytophthora infestans in Nicotiana benthamiana. Plos One *6*, e16608.
Chapman, A., Lees, A., Cooke, D., and Cooke, L. (2010). The changing Phytophthora infestans population: implications for Late Blight epidemics and control. PPO-Special Report Twelfth EuroBlight workshop Arras (France) *14*, 315 - 316
Chen, F., Tholl, D., D'Auria, J.C., Farooq, A., Pichersky, E., and Gershenzon, J. (2003). Biosynthesis and emission of terpenoid volatiles from Arabidopsis flowers. Plant Cell *15*, 481-494.
Chen, L.Q., Hou, B.H., Lalonde, S., Takanaga, H., Hartung, M.L., Qu, X.Q., Guo, W.J., Kim, J.G., Underwood, W., Chaudhuri, B., *et al.* (2010). Sugar transporters for intercellular exchange and nutrition of pathogens. Nature *468*, 527-532.
Cheng, Y., Qin, G., Dai, X., and Zhao, Y. (2008). NPY genes and AGC kinases define two key steps in auxin-mediated organogenesis in Arabidopsis. Proc Natl Acad Sci U S A *105*, 21017-21022.
Cheung, F., Win, J., Lang, J.M., Hamilton, J., Vuong, H., Leach, J.E., Kamoun, S., Andre Levesque, C., Tisserat, N., and Buell, C.R. (2008). Analysis of the Pythium ultimum transcriptome using Sanger and Pyrosequencing approaches. Bmc Genomics *9*, 542.
Clack, T., Mathews, S., and Sharrock, R.A. (1994). The phytochrome apoprotein family in Arabidopsis is encoded by five genes: the sequences and expression of PHYD and PHYE. Plant Molecular Biology *25*, 413-427.
Clegg, M.T., and Durbin, M.L. (2000). Flower color variation: a model for the experimental study of evolution. Proc Natl Acad Sci U S A *97*, 7016-7023.
Coffeen, W.C., and Wolpert, T.J. (2004). Purification and characterization of serine proteases that exhibit caspase-like activity are associated with programmed cell death in *Avena sativa*. Plant Cell *16*.
Colmenero-Flores, J.M., Moreno, L.P., Smith, C.E., and Covarrubias, A.A. (1999). Pvlea-18, a member of a new late-embryogenesis-abundant protein family that accumulates during water stress and in the growing regions of well-irrigated bean seedlings. Plant Physiol *120*, 93-104.
Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics *21*, 3674-3676.

Conway, L.J., and Poethig, R.S. (1997). Mutations of Arabidopsis thaliana that transform leaves into cotyledons. Proc Natl Acad Sci U S A *94*, 10209-10214.

Corbesier, L., Vincent, C., Jang, S., Fornara, F., Fan, Q., Searle, I., Giakountis, A., Farrona, S., Gissot, L., Turnbull, C.*, et al.* (2007). FT protein movement contributes to long-distance signaling in floral induction of Arabidopsis. Science *316*, 1030-1033.

Cornelis, G.R., Boland, A., Boyd, A.P., Geuijen, C., Iriarte, M., Neyt, C., Sory, M.P., and Stainier, I. (1998). The virulence plasmid of Yersinia, an antihost genome. Microbiol Mol Biol Rev *62*, 1315-1352.

Cubas, P., Lauter, N., Doebley, J., and Coen, E. (1999). The TCP domain: a motif found in proteins regulating plant growth and development. Plant J *18*, 215-222.

de Folter, S., Immink, R.G., Kieffer, M., Parenicova, L., Henz, S.R., Weigel, D., Busscher, M., Kooiker, M., Colombo, L., Kater, M.M.*, et al.* (2005). Comprehensive interaction map of the Arabidopsis MADS Box transcription factors. Plant Cell *17*, 1424-1433.

Dhonukshe, P., Huang, F., Galvan-Ampudia, C.S., Mahonen, A.P., Kleine-Vehn, J., Xu, J., Quint, A., Prasad, K., Friml, J., Scheres, B.*, et al.* (2010). Plasma membrane-bound AGC3 kinases phosphorylate PIN auxin carriers at TPRXS(N/S) motifs to direct apical PIN recycling. Development *137*, 3245-3255.

Ditta, G., Pinyopich, A., Robles, P., Pelaz, S., and Yanofsky, M.F. (2004). The SEP4 gene of Arabidopsis thaliana functions in floral organ and meristem identity. Curr Biol *14*, 1935-1940.

Divon, H.H., and Fluhr, R. (2007). Nutrition acquisition strategies during fungal infection of plants. Fems Microbiol Lett *266*, 65-74.

Dobes, C., Mitchell-Olds, T., and Koch, M.A. (2004). Intraspecific Diversification in North American Boechera stricta (= Arabis drummondii), Boechera x divaricarpa, and Boechera holboellii (Brassicaceae) Inferred from Nuclear and Chloroplast Molecular Markers: An Integrative Approach. American Journal of Botany *91*, 2087-2101.

Dou, D., Kale, S.D., Wang, X., Chen, Y., Wang, Q., Jiang, R.H., Arredondo, F.D., Anderson, R.G., Thakur, P.B., McDowell, J.M.*, et al.* (2008a). Conserved C-terminal motifs required for avirulence and suppression of cell death by Phytophthora sojae effector Avr1b. Plant Cell *20*, 1118-1133.

Dou, D., Kale, S.D., Wang, X., Jiang, R.H., Bruce, N.A., Arredondo, F.D., Zhang, X., and Tyler, B.M. (2008b). RXLR-mediated entry of Phytophthora sojae effector Avr1b into soybean cells does not require pathogen-encoded machinery. Plant Cell *20*, 1930-1947.

Douglas, S.J., Chuck, G., Dengler, R.E., Pelecanda, L., and Riggs, C.D. (2002). KNAT1 and ERECTA regulate inflorescence architecture in Arabidopsis. Plant Cell *14*, 547-558.

Drenth, A., Janssen, E.M., and Govers, F. (1995). Formation and survival of oospores of Phytophthora infestans under natural conditions. Plant Pathology *44*, 1365-3059.

Eimert, K., Wang, S.M., Lue, W.I., and Chen, J. (1995). Monogenic Recessive Mutations Causing Both Late Floral Initiation and Excess Starch Accumulation in Arabidopsis. Plant Cell *7*, 1703-1712.

Elango, N., Kim, S.H., Vigoda, E., and Yi, S.V. (2008). Mutations of different molecular origins exhibit contrasting patterns of regional substitution rate variation. PLoS Comput Biol *4*, e1000015.

Elde, N.C., and Malik, H.S. (2009). The evolutionary conundrum of pathogen mimicry. Nat Rev Microbiol *7*, 787-797.

Elizondo, L.I., Jafar-Nejad, P., Clewing, J.M., and Boerkoel, C.F. (2009). Gene clusters, molecular evolution and disease: a speculation. Curr Genomics *10*, 64-75.

Ellis, J.G., Rafiqi, M., Gan, P., Chakrabarti, A., and Dodds, P.N. (2009). Recent progress in discovery and functional analysis of effector proteins of fungal and oomycete plant pathogens. Curr Opin Plant Biol *12*, 399-405.

Emery, J.F., Floyd, S.K., Alvarez, J., Eshed, Y., Hawker, N.P., Izhaki, A., Baum, S.F., and Bowman, J.L. (2003). Radial patterning of Arabidopsis shoots by class III HD-ZIP and KANADI genes. Curr Biol *13*, 1768-1774.

Eriksson, S., Stransfeld, L., Adamski, N.M., Breuninger, H., and Lenhard, M. (2010). KLUH/CYP78A5-dependent growth signaling coordinates floral organ growth in Arabidopsis. Curr Biol *20*, 527-532.

Erwin, D.C., and Ribeiro, O.K. (1996). *Phytophthora* Diseases Worldwide. American Phytopathological Society APS Press, St Paul, Minnesota, 562 pages.

Evans, T.A., Davidson, C.R., Dominiak, J.D., Mulrooney, R.P., B., C.R., and H., A.S. (2002). Two New Races of Phytophthora phaseoli from Lima Bean in Delaware. Plant Dis *86*, 813-813

Fankhauser, C., Yeh, K.C., Lagarias, J.C., Zhang, H., Elich, T.D., and Chory, J. (1999). PKS1, a substrate phosphorylated by phytochrome that modulates light signaling in Arabidopsis. Science *284*, 1539-1541.

Farr, D.F., Bills, G.F., Chamuris, G.P., and Rossman, A.Y. (1989). Fungi on Plants and Plant Products in the United States. (American Phytopathological Society, St Paul, Minnesota).

Fernandez, O., Bethencourt, L., Quero, A., Sangwan, R.S., and Clement, C. (2010). Trehalose and plant stress responses: friend or foe? Trends Plant Sci *15*, 409-417.

Flier, W.G., Grünwald, N.J., Kroon, L., van der Bosch, T.B.M., Garay Serrano, E., Lozoya Saldana, H., Bonants, P.J.M., Turkensteen, L.J., and (2002). Phytophthora ipomoeae sp. nov., a new homothallic species causing leaf blight on Ipomoea longipedunculata in the Toluca Valley of central Mexico. Mycological Research *106*, 848-856.

Freeman, T.C., Goldovsky, L., Brosch, M., van Dongen, S., Maziere, P., Grocock, R.J., Freilich, S., Thornton, J., and Enright, A.J. (2007). Construction, visualisation, and clustering of transcription networks from microarray expression data. PLoS Comput Biol *3*, 2032-2042.

Fry, W. (2008). Phytophthora infestans: the plant (and R gene) destroyer. Mol Plant Pathol *9*, 385-402.

Fry, W., Grünwald, N.J., Cooke, D.E., McLeod, A., Forbes, G., and Cao, K. (2009). Population genetics and population diversity of

Phytophthora infestans. . In Oomycete Genetics and Genomics: Diversity, Interactions, and Research Tools, pp. 139-164.
Fu, J., Liu, H., Li, Y., Yu, H., Li, X., Xiao, J., and Wang, S. (2011). Manipulating broad-spectrum disease resistance by suppressing pathogen-induced auxin accumulation in rice. Plant Physiol *155*, 589-602.
Gaulin, E., Madoui, M.A., Bottin, A., Jacquet, C., Mathe, C., Couloux, A., Wincker, P., and Dumas, B. (2008). Transcriptome of Aphanomyces euteiches: new oomycete putative pathogenicity factors and metabolic pathways. Plos One *3*, e1723.
Geisler, M., Blakeslee, J.J., Bouchard, R., Lee, O.R., Vincenzetti, V., Bandyopadhyay, A., Titapiwatanakun, B., Peer, W.A., Bailly, A., Richards, E.L.*, et al.* (2005). Cellular efflux of auxin catalyzed by the Arabidopsis MDR/PGP transporter AtPGP1. Plant J *44*, 179-194.
Geisler, M., and Murphy, A.S. (2006). The ABC of auxin transport: the role of p-glycoproteins in plant development. FEBS Lett *580*, 1094-1102.
Gepstein, S. (2004). Leaf senescence--not just a 'wear and tear' phenomenon. Genome Biol *5*, 212.
Ghanta, S., Bhattacharyya, D., and Chattopadhyay, S. (2011). Glutathione signaling acts through NPR1-dependent SA-mediated pathway to mitigate biotic stress. Plant Signal Behav *6*, 607-609.
Giakountis, A., and Coupland, G. (2008). Phloem transport of flowering signals. Curr Opin Plant Biol *11*, 687-694.
Gietz, R.D., Schiestl, R.H., Willems, A.R., and Woods, R.A. (1995). Studies on the transformation of intact yeast cells by the LiAc/SS-DNA/PEG procedure. Yeast *11*, 355-360.
Gijzen, M. (2009). Runaway repeats force expansion of the Phytophthora infestans genome. Genome Biol *10*, 241.
Gijzen, M., and Nurnberger, T. (2006). Nep1-like proteins from plant pathogens: recruitment and diversification of the NPP1 domain across taxa. Phytochemistry *67*, 1800-1807.
Gilroy, E.M., Breen, S., Whisson, S.C., Squires, J., Hein, I., Kaczmarek, M., Turnbull, D., Boevink, P.C., Lokossou, A., Cano, L.M.*, et al.* (2011). Presence/absence, differential expression and sequence polymorphisms between PiAVR2 and PiAVR2-like in Phytophthora infestans determine virulence on R2 plants. New Phytol *191*, 763-776.
Giraud, T., Yockteng, R., Lopez-Villavicencio, M., Refregier, G., and Hood, M.E. (2008). Mating system of the anther smut fungus Microbotryum violaceum: selfing under heterothallism. Eukaryot Cell *7*, 765-775.
Gonneau, M., Pagant, S., Brun, F., and Laloue, M. (2001). Photoaffinity labelling with the cytokinin agonist azido-CPPU of a 34 kDa peptide of the intracellular pathogenesis-related protein family in the moss Physcomitrella patens. Plant Mol Biol *46*, 539-548.
Gregory, W.F., and Maizels, R.M. (2008). Cystatins from filarial parasites: Evolution, adaptation and function in the host–parasite relationship. The International Journal of Biochemistry & Cell Biology *40*, 1389–1398.

Grell, M.N., Jensen, A.B., Olsen, P.B., Eilenberg, J., and Lange, L. (2011). Secretome of fungus-infected aphids documents high pathogen activity and weak host response. Fungal Genet Biol *48*, 343-352.

Grouffaud, S., van West, P., Avrova, A.O., Birch, P.R., and Whisson, S.C. (2008). Plasmodium falciparum and Hyaloperonospora parasitica effector translocation motifs are functional in Phytophthora infestans. Microbiology *154*, 3743-3751.

Grunwald, N.J., and Flier, W.G. (2005). The biology of *Phytophthora infestans* at its center of origin. Annu Rev Phytopathol *43*, 171-190.

Haas, B.J., Kamoun, S., Zody, M.C., Jiang, R.H., Handsaker, R.E., Cano, L.M., Grabherr, M., Kodira, C.D., Raffaele, S., Torto-Alalibo, T.*, et al.* (2009). Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. Nature *461*, 393-398.

Hall, G. (1989). Plasmopara halstedii. CMI Descriptions of Pathogenic Fungi and Bacteria CAB International, Wallingford, UK *No. 979.* .

Halterman, D.A., Chen, Y., Sopee, J., Berduo-Sandoval, J., and Sánchez-Pérez, A. (2010). Competition between Phytophthora infestans Effectors Leads to Increased Aggressiveness on Plants Containing Broad-Spectrum Late Blight Resistance. Plos One *5*, e10536.

Hann, D.R., and Rathjen, J.P. (2007). Early events in the pathogenicity of Pseudomonas syringae on Nicotiana benthamiana. Plant J *49*, 607-618.

Hatai, K., and Hoshiai, G.I. (1994). Pathogenicity of Saprolegnia parasitica Coker. In Salmon Saprolegniasis Edited by: Mueller GJ Portland, Oregon, US Department of Energy, Bonneville Power Administration, 87-98.

Heese, A., Hann, D.R., Gimenez-Ibanez, S., Jones, A.M., He, K., Li, J., Schroeder, J.I., Peck, S.C., and Rathjen, J.P. (2007). The receptor-like kinase SERK3/BAK1 is a central regulator of innate immunity in plants. Proc Natl Acad Sci U S A *104*, 12217-12222.

Hewitt, W.B., and Pearson, R.C. (1988). Downy mildew. In: RC Pearson & AC Goheen (eds) Compendium of Grape Diseases The American Phytopathological Society APS Press, St Paul, MN.

Hjelmeland, K. (1983). Proteinase inhibitors in the muscle and serum of cod (Gadus morhua). Isolation and characterization. . Comp Biochem Physiol B Comp Biochem *76*, 365-372.

Hogenhout, S.A., Van der Hoorn, R.A., Terauchi, R., and Kamoun, S. (2009). Emerging concepts in effector biology of plant-associated organisms. Mol Plant Microbe Interact *22*, 115-122.

Holub, E.B., and Beynon, J.L. (1997). Symbiology of mouse-ear cress (*Arabidopsis thaliana*) and oomycetes. Adv Bot Res *24*.

Huang, H., Tudor, M., Weiss, C.A., Hu, Y., and Ma, H. (1995). The Arabidopsis MADS-box gene AGL3 is widely expressed and encodes a sequence-specific DNA-binding protein. Plant Mol Biol *28*, 549-567.

Hussein, M.M.A., and Hatai, K. (2002). Pathogenicity of Saprolegnia species associated with outbreaks of salmonid saprolegniosis in Japan. Fisheries Science *68*, 1067–1072.

Immink, R.G., Ferrario, S., Busscher-Lange, J., Kooiker, M., Busscher, M., and Angenent, G.C. (2003). Analysis of the petunia MADS-box transcription factor family. Mol Genet Genomics *268*, 598-606.

Jacobs, K.A., Collins-Racie, L.A., Colbert, M., Duckett, M., Golden-Fleet, M., Kelleher, K., Kriz, R., LaVallie, E.R., Merberg, D., Spaulding, V.*, et al.* (1997). A genetic selection for isolating cDNAs encoding secreted proteins. Gene *198*, 289-296.

Jiang, R.H., Tripathy, S., Govers, F., and Tyler, B.M. (2008). RXLR effector reservoir in two Phytophthora species is dominated by a single rapidly evolving superfamily with more than 700 members. Proc Natl Acad Sci U S A *105*, 4874-4879.

Jiang, R.H., Tyler, B.M., and Govers, F. (2006a). Comparative analysis of Phytophthora genes encoding secreted proteins reveals conserved synteny and lineage-specific gene duplications and deletions. Mol Plant Microbe Interact *19*, 1311-1321.

Jiang, R.H., Tyler, B.M., Whisson, S.C., Hardham, A.R., and Govers, F. (2006b). Ancient origin of elicitin gene clusters in Phytophthora genomes. Mol Biol Evol *23*, 338-351.

Joubes, J., Raffaele, S., Bourdenx, B., Garcia, C., Laroche-Traineau, J., Moreau, P., Domergue, F., and Lessire, R. (2008). The VLCFA elongase gene family in Arabidopsis thaliana: phylogenetic analysis, 3D modelling and expression profiling. Plant Mol Biol *67*, 547-566.

Judelson, H.S., Ah-Fong, A.M., Aux, G., Avrova, A.O., Bruce, C., Cakir, C., da Cunha, L., Grenville-Briggs, L., Latijnhouwers, M., Ligterink, W.*, et al.* (2008). Gene expression profiling during asexual development of the late blight pathogen Phytophthora infestans reveals a highly dynamic transcriptome. Mol Plant Microbe Interact *21*, 433-447.

Kalscheuer, R., and Steinbuchel, A. (2003). A novel bifunctional wax ester synthase/acyl-CoA:diacylglycerol acyltransferase mediates wax ester and triacylglycerol biosynthesis in Acinetobacter calcoaceticus ADP1. J Biol Chem *278*, 8075-8082.

Kammenga, J.E., Herman, M.A., Ouborg, N.J., Johnson, L., and Breitling, R. (2007). Microarray challenges in ecology. Trends Ecol Evol *22*, 273-279.

Kamoun, S. (2003). Molecular genetics of pathogenic oomycetes. Eukaryot Cell *2*, 191-199.

Kamoun, S. (2006). A catalogue of the effector secretome of plant pathogenic oomycetes. Annu Rev Phytopathol *44*, 41-60.

Kamoun, S. (2007). Groovy times: filamentous pathogen effectors revealed. Curr Opin Plant Biol *10*, 358-365.

Kamoun, S. (2009). The Secretome of Plant-Associated Fungi and Oomycetes. The Mycota V *Plant Relationships, 2nd Edition*, 173-180.

Kamoun, S., and Smart, C.D. (2005). Late Blight of Potato and Tomato in the Genomics Era. Plant Dis *89*, 692-699.

Kamoun, S., van West, P., de Jong, A.J., de Groot, K.E., Vleeshouwers, V.G., and Govers, F. (1997). A gene encoding a protein elicitor of Phytophthora infestans is down-regulated during infection of potato. Mol Plant Microbe Interact *10*, 13-20.

Kanneganti, T.D., Huitema, E., Cakir, C., and Kamoun, S. (2006). Synergistic interactions of the plant cell death pathways induced by Phytophthora infestans Nepl-like protein PiNPP1.1 and INF1 elicitin. Mol Plant Microbe Interact *19*, 854-863.

Kanrar, S., Bhattacharya, M., Arthur, B., Courtier, J., and Smith, H.M. (2008). Regulatory networks that function to specify flower meristems require the function of homeobox genes PENNYWISE and POUND-FOOLISH in Arabidopsis. Plant J *54*, 924-937.

Kanrar, S., Onguka, O., and Smith, H.M. (2006). Arabidopsis inflorescence architecture requires the activities of KNOX-BELL homeodomain heterodimers. Planta *224*, 1163-1173.

Kanzaki, H., Saitoh, H., Takahashi, Y., Berberich, T., Ito, A., Kamoun, S., and Terauchi, R. (2008). NbLRK1, a lectin-like receptor kinase protein of Nicotiana benthamiana, interacts with Phytophthora infestans INF1 elicitin and mediates INF1-induced cell death. Planta *228*, 977-987.

Kasahara, M., Kagawa, T., Oikawa, K., Suetsugu, N., Miyao, M., and Wada, M. (2002). Chloroplast avoidance movement reduces photodamage in plants. Nature *420*, 829-832.

Kawamura, Y., Hase, S., Takenaka, S., Kanayama, Y., Yoshioka, H., Kamoun, S., and Takahashi, H. (2009). INF1 Elicitin Activates Jasmonic Acid- and Ethylene-mediated Signalling Pathways and Induces Resistance to Bacterial Wilt Disease in Tomato. J Phytopathol *157*, 287–297.

Kemen, E., Gardiner, A., Schultz-Larsen, T., Kemen, A.C., Balmuth, A.L., Robert-Seilaniantz, A., Bailey, K., Holub, E., Studholme, D.J., Maclean, D.*, et al.* (2011). Gene gain and loss during evolution of obligate parasitism in the white rust pathogen of Arabidopsis thaliana. PLoS Biol *9*, e1001094.

Keppler, B.D., and Showalter, A.M. (2010). IRX14 and IRX14-LIKE, two glycosyl transferases involved in glucuronoxylan biosynthesis and drought tolerance in Arabidopsis. Mol Plant *3*, 834-841.

Kitamura, S., Shikazono, N., and Tanaka, A. (2004). TRANSPARENT TESTA 19 is involved in the accumulation of both anthocyanins and proanthocyanidins in Arabidopsis. Plant J *37*, 104-114.

Klein, R.D., Gu, Q., Goddard, A., and Rosenthal, A. (1996). Selection for genes encoding secreted proteins and receptors. Proc Natl Acad Sci U S A *93*, 7108-7113.

Klotz, C. (2004). Studien zur DNA-Vakzinierung von Hühnern mit Eimeria tenella-Antigenen. Dissertation Mathematisch-Naturwissenschaftlichen Fakultät Ider Humboldt-Universität zu Berlin.

Kong, Y., Zhou, G., Yin, Y., Xu, Y., Pattathil, S., and Hahn, M.G. (2011). Molecular analysis of a family of Arabidopsis genes related to galacturonosyltransferases. Plant Physiol *155*, 1791-1805.

Kosiol, C., Vinar, T., da Fonseca, R.R., Hubisz, M.J., Bustamante, C.D., Nielsen, R., and Siepel, A. (2008). Patterns of positive selection in six Mammalian genomes. PLoS Genet *4*, e1000144.

Kouzarides, T. (2002). Histone methylation in transcriptional control. Curr Opin Genet Dev *12*, 198-209.

Koyama, T., Furutani, M., Tasaka, M., and Ohme-Takagi, M. (2007). TCP transcription factors control the morphology of shoot lateral organs via negative regulation of the expression of boundary-specific genes in Arabidopsis. Plant Cell *19*, 473-484.

Koyama, T., Mitsuda, N., Seki, M., Shinozaki, K., and Ohme-Takagi, M. (2010). TCP transcription factors regulate the activities of ASYMMETRIC LEAVES1 and miR164, as well as the auxin response, during differentiation of leaves in Arabidopsis. Plant Cell *22*, 3574-3588.

Kram, B.W., and Carter, C.J. (2009). Arabidopsis thaliana as a model for functional nectary analysis. Sex Plant Reprod *22*, 235-246.

Kram, B.W., Xu, W.W., and Carter, C.J. (2009). Uncovering the Arabidopsis thaliana nectary transcriptome: investigation of differential gene expression in floral nectariferous tissues. BMC Plant Biol *9*, 92.

Krishnaswamy, S., Verma, S., Rahman, M.H., and Kav, N.N. (2011). Functional characterization of four APETALA2-family genes (RAP2.6, RAP2.6L, DREB19 and DREB26) in Arabidopsis. Plant Mol Biol *75*, 107-127.

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol *305*, 567-580.

Kroon, L.P., Bakker, F.T., van den Bosch, G.B., Bonants, P.J., and Flier, W.G. (2004). Phylogenetic analysis of Phytophthora species based on mitochondrial and nuclear DNA sequences. Fungal Genet Biol *41*, 766-782.

Kumaran, M.K., Ye, D., Yang, W.-C., Griffith, M.E., Chaudhury, A.M., and Sundaresan, V. (1999). Molecular cloning of ABNORMAL FLORAL ORGANS: a gene required for flower development in Arabidopsis. Sex Plant Reprod *12*, 118–122.

Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., and Salzberg, S.L. (2004). Versatile and open software for comparing large genomes. Genome Biol *5*, R12.

Lal, S., Gulyani, V., and Khurana, P. (2008). Overexpression of HVA1 gene from barley generates tolerance to salinity and water stress in transgenic mulberry (Morus indica). Transgenic Res *17*, 651-663.

Lariguet, P., Schepens, I., Hodgson, D., Pedmale, U.V., Trevisan, M., Kami, C., de Carbonnel, M., Alonso, J.M., Ecker, J.R., Liscum, E.*, et al.* (2006). PHYTOCHROME KINASE SUBSTRATE 1 is a phototropin 1 binding protein required for phototropism. Proc Natl Acad Sci U S A *103*, 10134-10139.

Lazar, G., and Goodman, H.M. (2006). MAX1, a regulator of the flavonoid pathway, controls vegetative axillary bud outgrowth in Arabidopsis. Proc Natl Acad Sci U S A *103*, 472-476.

Lee, C., Teng, Q., Huang, W., Zhong, R., and Ye, Z.H. (2010). The Arabidopsis family GT43 glycosyltransferases form two functionally nonredundant groups essential for the elongation of glucuronoxylan backbone. Plant Physiol *153*, 526-541.

Lee, C., Zhong, R., Richardson, E.A., Himmelsbach, D.S., McPhail, B.T., and Ye, Z.H. (2007). The PARVUS gene is expressed in cells undergoing secondary wall thickening and is essential for glucuronoxylan biosynthesis. Plant Cell Physiol *48*, 1659-1672.

Lee, S.J., Kelley, B.S., Damasceno, C.M., St John, B., Kim, B.S., Kim, B.D., and Rose, J.K. (2006). A functional screen to characterize the

secretomes of eukaryotic pathogens and their hosts in planta. Mol Plant Microbe Interact *19*, 1368-1377.
Lefebvre, B., Timmers, T., Mbengue, M., Moreau, S., Herve, C., Toth, K., Bittencourt-Silvestre, J., Klaus, D., Deslandes, L., Godiard, L.*, et al.* (2010). A remorin protein interacts with symbiotic receptors and regulates bacterial infection. Proc Natl Acad Sci U S A *107*, 2343-2348.
Levesque, C.A., Brouwer, H., Cano, L., Hamilton, J.P., Holt, C., Huitema, E., Raffaele, S., Robideau, G.P., Thines, M., Win, J.*, et al.* (2010). Genome sequence of the necrotrophic plant pathogen Pythium ultimum reveals original pathogenicity mechanisms and effector repertoire. Genome Biol *11*, R73.
Li, F., Wu, X., Lam, P., Bird, D., Zheng, H., Samuels, L., Jetter, R., and Kunst, L. (2008a). Identification of the wax ester synthase/acyl-coenzyme A: diacylglycerol acyltransferase WSD1 required for stem wax ester biosynthesis in Arabidopsis. Plant Physiol *148*, 97-107.
Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754-1760.
Li, H., Ruan, J., and Durbin, R. (2008b). Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res *18*, 1851-1858.
Liu, J., Elmore, J.M., and Coaker, G. (2009). Investigating the functions of the RIN4 protein complex during plant innate immune responses. Plant Signal Behav *4*, 1107-1110.
Liu, T., Ye, W., Ru, Y., Yang, X., Gu, B., Tao, K., Lu, S., Dong, S., Zheng, X., Shan, W.*, et al.* (2011). Two host cytoplasmic effectors are required for pathogenesis of Phytophthora sojae by suppression of host defenses. Plant Physiol *155*, 490-501.
Liu, Z., Bos, J.I., Armstrong, M., Whisson, S.C., da Cunha, L., Torto-Alalibo, T., Win, J., Avrova, A.O., Wright, F., Birch, P.R.*, et al.* (2005). Patterns of diversifying selection in the phytotoxin-like scr74 gene family of Phytophthora infestans. Mol Biol Evol *22*, 659-672.
Lu, S.M., Lu, W., Qasim, M.A., Anderson, S., Apostol, I., Ardelt, W., Bigler, T., Chiang, Y.W., Cook, J., James, M.N.*, et al.* (2001). Predicting the reactivity of proteins from their sequence alone: Kazal family of protein inhibitors of serine proteinases. Proc Natl Acad Sci U S A *98*, 1410-1415.
Luesse, D.R., DeBlasio, S.L., and Hangarter, R.P. (2006). Plastid movement impaired 2, a new gene involved in normal blue-light-induced chloroplast movements in Arabidopsis. Plant Physiol *141*, 1328-1337.
Luo, B., Xue, X.Y., Hu, W.L., Wang, L.J., and Chen, X.Y. (2007). An ABC transporter gene of Arabidopsis thaliana, AtWBC11, is involved in cuticle development and prevention of organ fusion. Plant Cell Physiol *48*, 1790-1802.
Maere, S., Heymans, K., and Kuiper, M. (2005). BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics *21*, 3448-3449.
Malcomber, S.T., and Kellogg, E.A. (2005). SEPALLATA gene diversification: brave new whorls. Trends Plant Sci *10*, 427-435.

Marks, M.D., Wenger, J.P., Gilding, E., Jilk, R., and Dixon, R.A. (2009). Transcriptome analysis of Arabidopsis wild-type and gl3-sst sim trichomes identifies four additional genes required for trichome development. Mol Plant *2*, 803-822.

Marrs, K.A. (1996). The Functions and Regulation of Glutathione S-Transferases in Plants. Annu Rev Plant Physiol Plant Mol Biol *47*, 127-158.

Martin, F.N., and Loper, J.E. (1999). Soilborne plant diseases caused by Pythium spp.: ecology, epidemiology, and prospects for biological control. Critical Reviews in Plant Sciences *18*, 111-181.

McConnell, J.R., Emery, J., Eshed, Y., Bao, N., Bowman, J., and Barton, M.K. (2001). Role of PHABULOSA and PHAVOLUTA in determining radial patterning in shoots. Nature *411*, 709-713.

Menne, K.M., Hermjakob, H., and Apweiler, R. (2000). A comparison of signal sequence prediction methods using a test set of signal peptides. Bioinformatics *16*, 741-742.

Mitsuda, N., Iwase, A., Yamamoto, H., Yoshida, M., Seki, M., Shinozaki, K., and Ohme-Takagi, M. (2007). NAC transcription factors, NST1 and NST3, are key regulators of the formation of secondary walls in woody tissues of Arabidopsis. Plant Cell *19*, 270-280.

Moller, A.P. (1995). Bumblebee preference for symmetrical flowers. Proc Natl Acad Sci U S A *92*, 2288-2292.

Morgan, W., and Kamoun, S. (2007). RXLR effectors of plant pathogenic oomycetes. Curr Opin Microbiol *10*, 332-338.

Morris, K., ., Thomas, H., and Rogers, L. (1996). Endopeptidases during the developmental and senescence of Lolium temulentum leaves. Phytochemistry *41*, 377–384.

Mortimer, J.C., Miles, G.P., Brown, D.M., Zhang, Z., Segura, M.P., Weimar, T., Yu, X., Seffen, K.A., Stephens, E., Turner, S.R.*, et al.* (2010). Absence of branches from xylan in Arabidopsis gux mutants reveals potential for simplification of lignocellulosic biomass. Proc Natl Acad Sci U S A *107*, 17409-17414.

Mueller, O., Kahmann, R., Aguilar, G., Trejo-Aguilar, B., Wu, A., and de Vries, R.P. (2008). The secretome of the maize pathogen Ustilago maydis. Fungal Genet Biol *45 Suppl 1*, S63-70.

Narita, N.N., Moore, S., Horiguchi, G., Kubo, M., Demura, T., Fukuda, H., Goodrich, J., and Tsukaya, H. (2004). Overexpression of a novel small peptide ROTUNDIFOLIA4 decreases cell proliferation and alters leaf shape in Arabidopsis thaliana. Plant J *38*, 699-713.

Neff, M.M., Fankhauser, C., and Chory, J. (2000). Light: an indicator of time and place. Genes Dev *14*, 257-271.

Ngugi, H.K., and Scherm, H. (2006). Mimicry in plant-parasitic fungi. Fems Microbiol Lett *257*, 171-176.

Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997). A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Int J Neural Syst *8*, 581-599.

Nurnberger, T., and Brunner, F. (2002). Innate immunity in plants and animals: emerging parallels between the recognition of general

elicitors and pathogen-associated molecular patterns. Curr Opin Plant Biol *5*, 318-324.

Ochando, I., Jover-Gil, S., Ripoll, J.J., Candela, H., Vera, A., Ponce, M.R., Martinez-Laborda, A., and Micol, J.L. (2006). Mutations in the microRNA complementarity site of the INCURVATA4 gene perturb meristem function and adaxialize lateral organs in arabidopsis. Plant Physiol *141*, 607-619.

Ogawa, M., Kay, P., Wilson, S., and Swain, S.M. (2009). ARABIDOPSIS DEHISCENCE ZONE POLYGALACTURONASE1 (ADPG1), ADPG2, and QUARTET2 are Polygalacturonases required for cell separation during reproductive development in Arabidopsis. Plant Cell *21*, 216-233.

Oh, S.K., Young, C., Lee, M., Oliva, R., Bozkurt, T.O., Cano, L.M., Win, J., Bos, J.I., Liu, H.Y., van Damme, M.*, et al.* (2009). In planta expression screens of Phytophthora infestans RXLR effectors reveal diverse phenotypes, including activation of the Solanum bulbocastanum disease resistance protein Rpi-blb2. Plant Cell *21*, 2928-2947.

Oikawa, A., Joshi, H.J., Rennie, E.A., Ebert, B., Manisseri, C., Heazlewood, J.L., and Scheller, H.V. (2010). An integrative approach to the identification of Arabidopsis and rice genes involved in xylan and secondary wall development. Plos One *5*, e15481.

Olvera-Carrillo, Y., Campos, F., Reyes, J.L., Garciarrubio, A., and Covarrubias, A.A. (2010). Functional analysis of the group 4 late embryogenesis abundant proteins reveals their relevance in the adaptive response during water deficit in Arabidopsis. Plant Physiol *154*, 373-390.

Pain, A., Bohme, U., Berry, A.E., Mungall, K., Finn, R.D., Jackson, A.P., Mourier, T., Mistry, J., Pasini, E.M., Aslett, M.A.*, et al.* (2008). The genome of the simian and human malaria parasite Plasmodium knowlesi. Nature *455*, 799-803.

Pak, C., and van Doorn, W.G. (2005). Delay of Iris flower senescence by protease inhibitors. New Phytol *165*, 473-480.

Palatnik, J.F., Allen, E., Wu, X., Schommer, C., Schwab, R., Carrington, J.C., and Weigel, D. (2003). Control of leaf morphogenesis by microRNAs. Nature *425*, 257-263.

Panikashvili, D., Savaldi-Goldstein, S., Mandel, T., Yifhar, T., Franke, R.B., Hofer, R., Schreiber, L., Chory, J., and Aharoni, A. (2007). The Arabidopsis DESPERADO/AtWBC11 transporter is required for cutin and wax secretion. Plant Physiol *145*, 1345-1360.

Panikashvili, D., Shi, J.X., Schreiber, L., and Aharoni, A. (2009). The Arabidopsis DCR encoding a soluble BAHD acyltransferase is required for cutin polyester formation and seed hydration properties. Plant Physiol *151*, 1773-1789.

Panikashvili, D., Shi, J.X., Schreiber, L., and Aharoni, A. (2011). The Arabidopsis ABCG13 transporter is required for flower cuticle secretion and patterning of the petal epidermis. New Phytol.

Park, J.E., Park, J.Y., Kim, Y.S., Staswick, P.E., Jeon, J., Yun, J., Kim, S.Y., Kim, J., Lee, Y.H., and Park, C.M. (2007). GH3-mediated auxin homeostasis links growth regulation with stress adaptation response in Arabidopsis. J Biol Chem *282*, 10036-10046.

Parkin, J. (1935). The Structure of the Starch Layer in the Glossy Petal of Ranunculus. II. The British Species Examined. . Annals of Botany *49* 283-289.

Pena, M.J., Zhong, R., Zhou, G.K., Richardson, E.A., O'Neill, M.A., Darvill, A.G., York, W.S., and Ye, Z.H. (2007). Arabidopsis irregular xylem8 and irregular xylem9: implications for the complexity of glucuronoxylan biosynthesis. Plant Cell *19*, 549-563.

Peng, J.C., and Karpen, G.H. (2009). Heterochromatic genome stability requires regulators of histone H3 K9 methylation. PLoS Genet *5*, e1000435.

Peters, A.H., O'Carroll, D., Scherthan, H., Mechtler, K., Sauer, S., Schofer, C., Weipoltshammer, K., Pagani, M., Lachner, M., Kohlmaier, A.*, et al.* (2001). Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. Cell *107*, 323-337.

Pfunder, M., and Roy, B.A. (2000). Pollinator-mediated interactions between a pathogenic fungus, Uromyces pisi (Pucciniaceae), and its host plant,Euphorbia cyparissias (Euphorbiaceae). Am J Bot *87*, 48-55.

Phillips, A.J., Anderson, V.L., Robertson, E.J., Secombes, C.J., and van West, P. (2008). New insights into animal pathogenic oomycetes. Trends Microbiol *16*, 13-19.

Pighin, J.A., Zheng, H., Balakshin, L.J., Goodman, I.P., Western, T.L., Jetter, R., Kunst, L., and Samuels, A.L. (2004). Plant cuticular lipid export requires an ABC transporter. Science *306*, 702-704.

Puhakainen, T., Hess, M.W., Makela, P., Svensson, J., Heino, P., and Palva, E.T. (2004). Overexpression of multiple dehydrin genes enhances tolerance to freezing stress in Arabidopsis. Plant Mol Biol *54*, 743-753.

Puigbo, P., Guzman, E., Romeu, A., and Garcia-Vallve, S. (2007). OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. Nucleic Acids Res *35*, W126-131.

Qutob, D., Huitema, E., Gijzen, M., and Kamoun, S. (2003). Variation in structure and activity among elicitins from Phytophthora sojae. Mol Plant Pathol *4*, 119-124.

Qutob, D., Tedman-Jones, J., Dong, S., Kuflu, K., Pham, H., Wang, Y., Dou, D., Kale, S.D., Arredondo, F.D., Tyler, B.M.*, et al.* (2009). Copy number variation and transcriptional polymorphisms of Phytophthora sojae RXLR effector genes Avr1a and Avr3a. Plos One *4*, e5066.

Raffaele, S., Farrer, R.A., Cano, L.M., Studholme, D.J., MacLean, D., Thines, M., Jiang, R.H., Zody, M.C., Kunjeti, S.G., Donofrio, N.M.*, et al.* (2010a). Genome evolution following host jumps in the Irish potato famine pathogen lineage. Science *330*, 1540-1543.

Raffaele, S., Mongrand, S., Gamas, P., Niebel, A., and Ott, T. (2007). Genome-wide annotation of remorins, a plant-specific protein family: evolutionary and functional perspectives. Plant Physiol *145*, 593-600.

Raffaele, S., Win, J., Cano, L.M., and Kamoun, S. (2010b). Analyses of genome architecture and gene expression reveal novel candidate virulence factors in the secretome of Phytophthora infestans. Bmc Genomics *11*, 637.

Raguso, R.A., Levin, R.A., Foose, S.E., Holmberg, M.W., and McDade, L.A. (2003). Fragrance chemistry, nocturnal rhythms and pollination "syndromes" in Nicotiana. Phytochemistry *63*, 265-284.

Raguso, R.A., and Roy, B.A. (1998). 'Floral' scent production by Puccinia rust fungi that mimic flowers. Mol Ecol *7*, 1127-1136.

Randall, T.A., Dwyer, R.A., Huitema, E., Beyer, K., Cvitanich, C., Kelkar, H., Fong, A.M., Gates, K., Roberts, S., Yatzkan, E.*, et al.* (2005). Large-scale gene discovery in the oomycete Phytophthora infestans reveals likely components of phytopathogenicity shared with true fungi. Mol Plant Microbe Interact *18*, 229-243.

Rasband, W.S. (1997-2011). Image J: Image Processing and Analysis in Java. US National Institutes of Health, Bethesda, Maryland, USA, imagejnihgov/ij/ National Institutes of Health, USA.

Rehmany, A.P., Gordon, A., Rose, L.E., Allen, R.L., Armstrong, M.R., Whisson, S.C., Kamoun, S., Tyler, B.M., Birch, P.R., and Beynon, J.L. (2005). Differential recognition of highly divergent downy mildew avirulence gene alleles by RPP1 resistance genes from two Arabidopsis lines. Plant Cell *17*, 1839-1850.

Roy, B.A. (1993a). Floral mimicry by a plant pathogen. Nature *362*, 56 - 58.

Roy, B.A. (1993b). Patterns of Rust Infection as a Function of Host Genetic Diversity and Host Density in Natural Populations of the Apomictic Crucifer, Arabis holboellii. Evolution *47*, 111-124.

Roy, B.A. (1994). The Effects of Pathogen-Induced Pseudoflowers and Buttercups on Each Other's Insect Visitation. Ecology *75*, 352-358

Roy, B.A. (1996). A Plant Pathogen Influences Pollinator Behavior and May Influence Reproduction of Nonhosts. Ecology *77*, 2445-2457

Roy, B.A. (2001). Patterns of association between crucifers and their flower-mimic pathogens: host jumps are more common than coevolution or cospeciation. Evolution *55*, 41-53.

Roy, B.A., and Raguso, R.A. (1997). Olfactory versus visual cues in a floral mimicry system. Oecologia *109*, 414–426.

Ruhlmann, J.M., Kram, B.W., and Carter, C.J. (2010). CELL WALL INVERTASE 4 is required for nectar production in Arabidopsis. J Exp Bot *61*, 395-404.

Saeed, A.I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T., Thiagarajan, M.*, et al.* (2003). TM4: a free, open-source system for microarray data management and analysis. Biotechniques *34*, 374-378.

Schaller, A. (2004). A cut above the rest: the regulatory function of plant proteases. Planta *220*, 183-197.

Schmid, M., Davison, T.S., Henz, S.R., Pape, U.J., Demar, M., Vingron, M., Scholkopf, B., Weigel, D., and Lohmann, J.U. (2005). A gene expression map of Arabidopsis thaliana development. Nat Genet *37*, 501-506.

Schmitthenner, A.F. (1985). Problems and progress toward control of Phytophthora root rot of soybean. Plant Dis *69*, 362–368.

Schneider, G., and Fechner, U. (2004). Advances in the prediction of protein targeting signals. Proteomics *4*, 1571-1580.

Schornack, S., Huitema, E., Cano, L.M., Bozkurt, T.O., Oliva, R., Van Damme, M., Schwizer, S., Raffaele, S., Chaparro-Garcia, A., Farrer,

R., *et al.* (2009). Ten things to know about oomycete effectors. Mol Plant Pathol *10*, 795-803.
Schornack, S., van Damme, M., Bozkurt, T.O., Cano, L.M., Smoker, M., Thines, M., Gaulin, E., Kamoun, S., and Huitema, E. (2010). Ancient class of translocated oomycete effectors targets the host nucleus. Proc Natl Acad Sci U S A *107*, 17421-17426.
Schranz, M.E., Windsor, A.J., Song, B.H., Lawton-Rauh, A., and Mitchell-Olds, T. (2007). Comparative genetic mapping in Boechera stricta, a close relative of Arabidopsis. Plant Physiol *144*, 286-298.
Scofield, S., Dewitte, W., and Murray, J.A. (2008). A model for Arabidopsis class-1 KNOX gene function. Plant Signal Behav *3*, 257-259.
Seidl, A., Clark, R., and Gevens, A.J. (2010). Characterizing Phytophthora infestans US#22 for fungicide resistance, and pathogenicity and virulence on tomato cultivars. North Cent Div Meet Abstr http://wwwapsnetorg/members/divisions/nc/meetings/Pages/2010MeetingAbstractsaspx.
Seo, P.J., Ryu, J., Kang, S.K., and Park, C.M. (2011). Modulation of sugar metabolism by an INDETERMINATE DOMAIN transcription factor contributes to photoperiodic flowering in Arabidopsis. Plant J *65*, 418-429.
Sherson, S.M., Alford, H.L., Forbes, S.M., Wallace, G., and Smith, S.M. (2003). Roles of cell-wall invertases and monosaccharide transporters in the growth and development of Arabidopsis. J Exp Bot *54*, 525-531.
Shirley, B.W., Kubasek, W.L., Storz, G., Bruggemann, E., Koornneef, M., Ausubel, F.M., and Goodman, H.M. (1995). Analysis of Arabidopsis mutants deficient in flavonoid biosynthesis. Plant J *8*, 659-671.
Sidler, M., Hassa, P., Hasan, S., Ringli, C., and Dudler, R. (1998). Involvement of an ABC transporter in a developmental pathway regulating hypocotyl cell elongation in the light. Plant Cell *10*, 1623-1636.
Slusarenko, A.J., and Schlaich, N.L. (2003). Downy mildew of Arabidopsis thaliana caused by Hyaloperonospora parasitica (formerly Peronospora parasitica). Mol Plant Pathol *4*, 159-170.
Smith, A.P., Nourizadeh, S.D., Peer, W.A., Xu, J., Bandyopadhyay, A., Murphy, A.S., and Goldsbrough, P.B. (2003). Arabidopsis AtGSTF2 is regulated by ethylene and auxin, and encodes a glutathione S-transferase that interacts with flavonoids. Plant J *36*, 433-442.
Smith, H.M., Campbell, B.C., and Hake, S. (2004). Competence to respond to floral inductive signals requires the homeobox genes PENNYWISE and POUND-FOOLISH. Curr Biol *14*, 812-817.
Smith, H.M., and Hake, S. (2003). The interaction of two homeobox genes, BREVIPEDICELLUS and PENNYWISE, regulates internode patterning in the Arabidopsis inflorescence. Plant Cell *15*, 1717-1727.
Song, J., Win, J., Tian, M., Schornack, S., Kaschani, F., Ilyas, M., van der Hoorn, R.A., and Kamoun, S. (2009). Apoplastic effectors secreted by two unrelated eukaryotic plant pathogens target the tomato defense protease Rcr3. Proc Natl Acad Sci U S A *106*, 1654-1659.
Song, J.T., Lu, H., and Greenberg, J.T. (2004). Divergent roles in Arabidopsis thaliana development and defense of two homologous

genes, aberrant growth and death2 and AGD2-LIKE DEFENSE RESPONSE PROTEIN1, encoding novel aminotransferases. Plant Cell *16*, 353-366.

Staples, R.C., Hoch, H.C., Epstein, L., Laccetti, L., and Hassouna, S. (1985). Recognition of host morphology by rust fungi: Responses and mechanisms. Can J Plant Pathol *7*, 314-322.

Stassen, J.H.M., and Van den Ackerveken, G. (2011). How do oomycete effectors interfere with plant life? Curr Opin Plant Biol *14*, 407-414.

Staswick, P.E., Serban, B., Rowe, M., Tiryaki, I., Maldonado, M.T., Maldonado, M.C., and Suza, W. (2005). Characterization of an Arabidopsis enzyme family that conjugates amino acids to indole-3-acetic acid. Plant Cell *17*, 616-627.

Stepanova, A.N., Robertson-Hoyt, J., Yun, J., Benavente, L.M., Xie, D.Y., Dolezal, K., Schlereth, A., Jurgens, G., and Alonso, J.M. (2008). TAA1-mediated auxin biosynthesis is essential for hormone crosstalk and plant development. Cell *133*, 177-191.

Storey, J.D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. Proc Natl Acad Sci U S A *100*, 9440-9445.

Sun, F., P., L., J., X., and H., D. (2010). Mutation in RAP2.6L, a transactivator of the ERF transcription factor

family, enhances Arabidopsis resistance to Pseudomonas syringae. Physiological and Molecular Plant Pathology *74*, 295-302.

Sutton, P.N., Gilbert, M.J., Williams, L.E., and Hall, J.L. (2007). Powdery mildew infection of wheat leaves changes host solute transport and invertase activity. Physiologia Plantarum, *129*, 787–795.

Sutton, P.N., Henry, M.J., and Hall, J.L. (1999). Glucose, and not sucrose, is transported from wheat to wheat powdery mildew. Planta *208*, 426-430.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol Biol Evol.

Tanaka, T., Tanaka, H., Machida, C., Watanabe, M., and Machida, Y. (2004). A new method for rapid visualization of defects in leaf cuticle reveals five intrinsic patterns of surface defects in Arabidopsis. Plant J *37*, 139-146.

Thines, M., Choi, Y.J., Kemen, E., Ploch, S., Holub, E.B., Shin, H.D., and Jones, J.D. (2009). A new species of Albugo parasitic to Arabidopsis thaliana reveals new evolutionary patterns in white blister rusts (Albuginaceae). Persoonia *22*, 123-128.

Thompson, J.D., Gibson, T.J., and Higgins, D.G. (2002). Multiple sequence alignment using ClustalW and ClustalX. Curr Protoc Bioinformatics *Chapter 2*, Unit 2 3.

Tian, M., Benedetti, B., and Kamoun, S. (2005). A Second Kazal-like protease inhibitor from Phytophthora infestans inhibits and interacts with the apoplastic pathogenesis-related protease P69B of tomato. Plant Physiol *138*, 1785-1793.

Tian, M., Huitema, E., Da Cunha, L., Torto-Alalibo, T., and Kamoun, S. (2004). A Kazal-like extracellular serine protease inhibitor from

Phytophthora infestans targets the tomato pathogenesis-related protease P69B. J Biol Chem *279*, 26370-26377.
Tian, M., and Kamoun, S. (2005). A two disulfide bridge Kazal domain from Phytophthora exhibits stable inhibitory activity against serine proteases of the subtilisin family. BMC Biochem *6*, 15.
Tian, M., Win, J., Savory, E., Burkhardt, A., Held, M., Brandizzi, F., and Day, B. (2011). 454 Genome sequencing of Pseudoperonospora cubensis reveals effector proteins with a QXLR translocation motif. Mol Plant Microbe Interact *24*, 543-553.
Tian, M., Win, J., Song, J., van der Hoorn, R., van der Knaap, E., and Kamoun, S. (2007). A Phytophthora infestans cystatin-like protein targets a novel tomato papain-like apoplastic protease. Plant Physiol *143*, 364-377.
Torto, T.A., Li, S., Styer, A., Huitema, E., Testa, A., Gow, N.A., van West, P., and Kamoun, S. (2003). EST mining and functional expression assays identify extracellular effector proteins from the plant pathogen Phytophthora. Genome Res *13*, 1675-1685.
Torto-Alalibo, T., Tian, M., Gajendran, K., Waugh, M.E., van West, P., and Kamoun, S. (2005). Expressed sequence tags from the oomycete fish pathogen Saprolegnia parasitica reveal putative virulence factors. Bmc Microbiol *5*, 46.
Trigueros, M., Navarrete-Gomez, M., Sato, S., Christensen, S.K., Pelaz, S., Weigel, D., Yanofsky, M.F., and Ferrandiz, C. (2009). The NGATHA genes direct style development in the Arabidopsis gynoecium. Plant Cell *21*, 1394-1409.
Tyler, B.M. (2007). Phytophthora sojae: root rot pathogen of soybean and model oomycete. Mol Plant Pathol *8*, 1-8.
Tyler, B.M., Tripathy, S., Zhang, X., Dehal, P., Jiang, R.H., Aerts, A., Arredondo, F.D., Baxter, L., Bensasson, D., Beynon, J.L.*, et al.* (2006). Phytophthora genome sequences uncover evolutionary origins and mechanisms of pathogenesis. Science *313*, 1261-1266.
van de Lagemaat, L.N., Landry, J.R., Mager, D.L., and Medstrand, P. (2003). Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. Trends Genet *19*, 530-536.
Van der Hoorn, R.A., Laurent, F., Roth, R., and De Wit, P.J. (2000). Agroinfiltration is a versatile tool that facilitates comparative analyses of Avr9/Cf-9-induced and Avr4/Cf-4-induced necrosis. Mol Plant Microbe Interact *13*, 439-446.
van der Lee, T., Robold, A., Testa, A., van 't Klooster, J.W., and Govers, F. (2001). Mapping of avirulence genes in Phytophthora infestans with amplified fragment length polymorphism markers selected by bulked segregant analysis. Genetics *157*, 949-956.
van Poppel, P.M., Guo, J., van de Vondervoort, P.J., Jung, M.W., Birch, P.R., Whisson, S.C., and Govers, F. (2008). The Phytophthora infestans avirulence gene Avr4 encodes an RXLR-dEER effector. Mol Plant Microbe Interact *21*, 1460-1470.
van West, P. (2006). Saprolegnia parasitica, an oomycete pathogen with a fishy appetite: new challenges for an old problem. Mycologist *20*, 99 - 104.

van West, P., Shepherd, S.J., Walker, C.A., Li, S., Appiah, A.A., Grenville-Briggs, L.J., Govers, F., and Gow, N.A. (2008). Internuclear gene silencing in Phytophthora infestans is established through chromatin remodelling. Microbiology *154*, 1482-1490.

Venglat, S.P., Dumonceaux, T., Rozwadowski, K., Parnell, L., Babic, V., Keller, W., Martienssen, R., Selvaraj, G., and Datla, R. (2002). The homeobox gene BREVIPEDICELLUS is a key regulator of inflorescence architecture in Arabidopsis. Proc Natl Acad Sci U S A *99*, 4730-4735.

Vidaurre, D.P., Ploense, S., Krogan, N.T., and Berleth, T. (2007). AMP1 and MP antagonistically regulate embryo and meristem development in Arabidopsis. Development *134*, 2561-2567.

Vitolo, M., and Borzani, W. (1983). Measurement of invertase activity of cells of Saccharomyces cerevisiae. Anal Biochem *130*, 469-470.

Vleeshouwers, V., Raffaele, S., Vossen, J., Champouret, N., Oliva, R., Segretin, M.E., Rietman, H., Cano, L.M., Lokossou, A., Kessel, G.*, et al.* (2011). Understanding and Exploiting Late Blight Resistance in the Age of Effectors. Annu Rev Phytopathol.

Vleeshouwers, V.G., Driesprong, J.D., Kamphuis, L.G., Torto-Alalibo, T., Van't Slot, K.A., Govers, F., Visser, R.G., Jacobsen, E., and Kamoun, S. (2006). Agroinfection-based high-throughput screening reveals specific recognition of INF elicitins in Solanum. Mol Plant Pathol *7*, 499-510.

Vleeshouwers, V.G., Rietman, H., Krenek, P., Champouret, N., Young, C., Oh, S.K., Wang, M., Bouwmeester, K., Vosman, B., Visser, R.G.*, et al.* (2008). Effector genomics accelerates discovery and functional profiling of potato disease resistance and phytophthora infestans avirulence genes. Plos One *3*, e2875.

Vleeshouwers, V.G.A.A., and Rietman, H. (2009). In Planta expression systems. In: K Lamour & S Kamoun (eds) S Oomycete Genetics and Genomics: Diversity, Interactions and Research Tools Wiley-Blackwell, Hoboken, NJ.

Vleeshouwers, V.G.A.A., van Dooijeweert, W., Paul Keizer, L.C., Sijpkes, L., Govers, F., and Colon, L.T. (1999). A Laboratory Assay for Phytophthora infestans Resistance in Various Solanum Species Reflects the Field Situation. Eur J Plant Pathol *105*, 241-150.

Volkman, S.K., Sabeti, P.C., DeCaprio, D., Neafsey, D.E., Schaffner, S.F., Milner, D.A., Jr., Daily, J.P., Sarr, O., Ndiaye, D., Ndir, O.*, et al.* (2007). A genome-wide map of diversity in Plasmodium falciparum. Nat Genet *39*, 113-119.

Wagner, U., Edwards, R., Dixon, D.P., and Mauch, F. (2002). Probing the diversity of the Arabidopsis glutathione S-transferase gene family. Plant Mol Biol *49*, 515-532.

Wen, J., Lease, K.A., and Walker, J.C. (2004). DVL, a novel class of small polypeptides: overexpression alters Arabidopsis development. Plant J *37*, 668-677.

Whisson, S.C., Boevink, P.C., Moleleki, L., Avrova, A.O., Morales, J.G., Gilroy, E.M., Armstrong, M.R., Grouffaud, S., van West, P., Chapman, S.*, et al.* (2007). A translocation signal for delivery of oomycete effector proteins into host plant cells. Nature *450*, 115-118.

Whisson, S.C., van der Lee, T., Bryan, G.J., Waugh, R., Govers, F., and Birch, P.R. (2001). Physical mapping across an avirulence locus of Phytophthora infestans using a highly representative, large-insert bacterial artificial chromosome library. Mol Genet Genomics *266*, 289-295.

Whitelam, G.C., Patel, S., and Devlin, P.F. (1998). Phytochromes and photomorphogenesis in Arabidopsis. Philos Trans R Soc Lond B Biol Sci *353*, 1445-1453.

Win, J., Morgan, W., Bos, J., Krasileva, K.V., Cano, L.M., Chaparro-Garcia, A., Ammar, R., Staskawicz, B.J., and Kamoun, S. (2007). Adaptive evolution has targeted the C-terminal domain of the RXLR effectors of plant pathogenic oomycetes. Plant Cell *19*, 2349-2369.

Windsor, A.J., Schranz, M.E., Formanova, N., Gebauer-Jung, S., Bishop, J.G., Schnabelrauch, D., Kroymann, J., and Mitchell-Olds, T. (2006). Partial shotgun sequencing of the Boechera stricta genome reveals extensive microsynteny and promoter conservation with Arabidopsis. Plant Physiol *140*, 1169-1182.

Wu, A.M., Hornblad, E., Voxeur, A., Gerber, L., Rihouey, C., Lerouge, P., and Marchant, A. (2010). Analysis of the Arabidopsis IRX9/IRX9-L and IRX14/IRX14-L pairs of glycosyltransferase genes reveals critical contributions to biosynthesis of the hemicellulose glucuronoxylan. Plant Physiol *153*, 542-554.

Wu, A.M., Rihouey, C., Seveno, M., Hornblad, E., Singh, S.K., Matsunaga, T., Ishii, T., Lerouge, P., and Marchant, A. (2009). The Arabidopsis IRX10 and IRX10-LIKE glycosyltransferases are critical for glucuronoxylan biosynthesis during secondary cell wall formation. Plant J *57*, 718-731.

Xiao, B., Huang, Y., Tang, N., and Xiong, L. (2007). Over-expression of a LEA gene in rice improves drought resistance under the field conditions. Theor Appl Genet *115*, 35-46.

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol *24*, 1586-1591.

Yang, Z., and Nielsen, R. (2000). Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol Biol Evol *17*, 32-43.

Yoon, S., Xuan, Z., Makarov, V., Ye, K., and Sebat, J. (2009). Sensitive and accurate detection of copy number variants using read depth of coverage. Genome Res *19*, 1586-1592.

Zeh, D.W., Zeh, J.A., and Ishida, Y. (2009). Transposable elements and an epigenetic basis for punctuated equilibria. Bioessays *31*, 715-726.

Zhang, Y., and Reinberg, D. (2001). Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. Genes Dev *15*, 2343-2360.

Zhong, R., Pena, M.J., Zhou, G.K., Nairn, C.J., Wood-Jones, A., Richardson, E.A., Morrison, W.H., 3rd, Darvill, A.G., York, W.S., and Ye, Z.H. (2005). Arabidopsis fragile fiber8, which encodes a putative glucuronyltransferase, is essential for normal secondary wall synthesis. Plant Cell *17*, 3390-3408.

Zhu, J., Obrycki, J.J., Ochieng, S.A., Baker, T.C., Pickett, J.A., and Smiley, D. (2005). Attraction of two lacewing species to volatiles produced by host plants and aphid prey. Naturwissenschaften *92*, 277-281.

Zondlo, S.C., and Irish, V.F. (1999). CYP78A5 encodes a cytochrome P450 that marks the shoot apical meristem boundary in Arabidopsis. Plant J *19*, 259-268.