Genomic Evolution in the Oomycetes

Maynooth University
National University of Ireland Maynooth

A thesis submitted to Maynooth University
for the degree of Doctor of Philosophy

Jamie McGowan BSc

June 2020

Supervisor
Dr. David Fitzpatrick,
Department of Biology,
Maynooth University,
Co. Kildare,
Ireland

Head of Department
Prof. Paul Moynagh,
Department of Biology,
Maynooth University,
Co. Kildare,
Ireland

Table of Contents

List of Figures

List of Tables

Declaration of Authorship

This thesis has not previously been submitted in whole or in part to this or any other University for any other degree. This thesis is the sole work of the author, with the exception of the isolation of the three *Phytophthora* species, which were provided by Dr. Richard O'Hanlon (Agri-Food and Biosciences Institute, Belfast, Northern Ireland).

Signed 

Jamie McGowan, BSc

Acknowledgements

Firstly, I would like to thank my supervisor, Dr. David Fitzpatrick. Thanks for all the help, advice and encouragement over the years since my undergraduate research project. I really appreciate all the opportunities you've provided to work on various projects and to travel to conferences. I can't imagine having a better supervisor. I would also like to thank Dr. Rebecca Owens, for the use of her lab and for all her help with the proteomics work. Thanks also to Dr. Richard O'Hanlon for providing the *Phytophthora* isolates.

My research would not have been possible without access to high performance computing infrastructure provided by the Irish Centre for High-End Computing (ICHEC). I would also like to thank the Irish Research Council for funding my PhD research. Mass spectrometry facilities were funded by a Science Foundation Ireland infrastructure award (SFI 12/RI/2346(3)).

Thanks to all the members of the Biology Department, past and present, for making Maynooth such a great place to work, for all the tea breaks, nights out and summer/Christmas parties. Special thanks to my lab partners, Charley and Eoin. Thanks for all your help, good memories and friendship along the way, from all the craic on trips to conferences and nights out, to the days we would all have mini breakdowns when scripts/servers crashed. Thanks for helping to make the PhD an enjoyable experience.

I would also like to thank my family. Thanks to my brothers, Ciarán and Callum, and my parents, who had to put up with me writing this thesis during the Covid-19 crisis, for their continued support and encouragement. Special thanks also to my Nanny, for always checking up on how "school" was going.

Publications

Publications Arising from this Thesis

McGowan, J., Fitzpatrick, D.A., 2017. Genomic, Network, and Phylogenetic Analysis of the Oomycete Effector Arsenal. mSphere. 2, e00408-17.

McGowan, J., Byrne, K.P., Fitzpatrick, D.A., 2019. Comparative Analysis of Oomycete Genome Evolution Using the Oomycete Gene Order Browser (OGOB). Genome Biol. Evol. 11, 189–206.

McGowan, J., Fitzpatrick DA. 2020. Recent advances in oomycete genomics. Advances in Genetics.

McGowan, J., O'Hanlon R., Owens, R.A., Fitzpatrick D.A., 2020. Comparative Genomic and Proteomic Analyses of Three Widespread *Phytophthora* Species: *Phytophthora chlamydospora*, *Phytophthora gonapodyides* and *Phytophthora pseudosyringae*. Microorganisms. 8, 653.

Other Publications

Waldron, R., **McGowan, J**., Gordon, N., McCarthy, C., Mitchell, E.B., Doyle, S., Fitzpatrick, D.A., 2017. Draft Genome Sequence of *Dermatophagoides pteronyssinus* , the European House Dust Mite. Genome Announc. 5, e00789-17.

Bayram, Ö.S., Dettmann, A., Karahoda, B., Moloney, N.M., Ormsby, T., **McGowan, J**., Cea-Sánchez, S., Miralles-Durán, A., Brancini, G.T.P., Luque, E.M., Fitzpatrick, D.A., Cánovas, D., Corrochano, L.M., Doyle, S., Selker, E.U., Seiler, S., Bayram, Ö., 2019. Control of Development, Secondary Metabolism and Light-Dependent Carotenoid Biosynthesis by the Velvet Complex of *Neurospora crassa*. Genetics 212, 691–710.

Waldron, R., **McGowan, J**., Gordon, N., McCarthy, C., Mitchell, E.B., Fitzpatrick, D.A., 2019. Proteome and allergenome of the European house dust mite *Dermatophagoides pteronyssinus*. PLoS One 14, e0216171.

Waldron, R., **McGowan, J**., Gordon, N., Mitchell, E.B., Fitzpatrick, D.A., Doyle, S., 2019. Characterisation of three novel β-1,3 glucanases from the medically important house dust mite *Dermatophagoides pteronyssinus* (airmid). Insect Biochem. Mol. Biol. 115, 103242.

O'Connor, E., **McGowan, J**., McCarthy, C.G.P., Amini, A., Grogan, H., Fitzpatrick, D.A., 2019. Whole Genome Sequence of the Commercially Relevant Mushroom Strain *Agaricus bisporus var. bisporus* ARP23. G3; Genes|Genomes|Genetics 9, 3057–3066.

Presentations

Oral Presentations

2019 Society for Molecular Biology and Evolution. Manchester, UK.

2019 Oomycete Molecular Genetics Network. Oban, UK.

2019 Maynooth University Biology Department Seminar Series. Maynooth University, IE.

2018 Irish Plant Scientists Association Meeting. University College Dublin, IE.

2018 Irish Fungal Society. Maynooth University, IE.

2018 Maynooth University Biology Department Research Day. Maynooth University, IE.

2018 Microbiology Society. Birmingham, UK.

2017 Maynooth University Biology Department Seminar Series. Maynooth University, IE.

2017 Irish Fungal Society. Limerick Institute of Technology, IE.

2017 Maynooth University Biology Department Research Day. Maynooth University, IE.

Poster Presentations

2019 Oomycete Molecular Genetics Network. Oban, UK.

2019 Early Career Microbiologists' Forum. Trinity College Dublin, IE.

2019 Maynooth University Biology Department Research Day. Maynooth University, IE.

2019 Genetics Society of America Fungal Genetics Meeting. Asilomar, California, US.

2019 Virtual Institute for Bioinformatics and Evolution. University College Dublin, IE.

2019 Microbiology Society. Belfast, UK.

2018 Molecular & Computational Biology Symposium. University College Dublin, IE.

2017 Society for Molecular Biology and Evolution. Austin, Texas, US.

Abbreviations

BLAST	Basic Local Alignment Search Tool
BP	Bootstrap support
BPP	Bayesian posterior probabilities
BUSCO	Benchmarking Universal Single Copy Orthologs
CAZyme	Carbohydrate-active enzyme
CBEL	Cellulose binding elicitor lectin
CGOB	Candida Gene Order Browser
CRN	Crinkler
cV8	Clarified V8 juice
EDTA	Ethylenediaminetetraacetic acid
EST	Expressed sequence tag
FDR	False discovery rate
GH	Glycoside hydrolase
GO	Gene Ontology
H_2O_2	Hydrogen peroxide
HMM	Hidden Markov model
HTG	Horizontal gene transfer
ITS	Internal transcribed spacer
Kb	Kilobase
LC-MS/MS	Liquid Chromatography Mass Spectrometry Mass Spectrometry
LGT	Lateral gene transfer
MAMP	Microbe-associated molecular pattern
Mb	Megabase
ML	Maximum likelihood
MYA	Million years ago
NCBI	National Centre for Biotechnology Information
NLP	Necrosis inducing protein
OGOB	Oomycete Gene Order Browser
PCA	Principal Component Analysis
PHI-BASE	Pathogen-Host Interaction Database
rRNA	Ribosomal RNA
RxLR	RxLR effector
SAR	Stramenopila, Alveolata, and Rhizaria
SDS	Sodium dodecyl sulphate
SFI	Suppressor of early Flg22-induced Immune response
TCA	Trichloroacetic acid
TFA	Trifluoroacetic acid
TM	Transmembrane
tRNA	Transfer RNA
YGOB	Yeast Gene Order Browser

Summary

Members of the oomycete class include some of the most devastating pathogens of plants and animals. Oomycetes secrete large arsenals of effector proteins that perform a wide range of functions, including sequestering nutrients from hosts and the environment, degrading host cells to facilitate colonisation, and modulating host immune responses. Recent genome sequencing projects have generated large amounts of oomycete genomic data. Availability of which facilitated bioinformatic characterisation of the secretomes of diverse oomycete species. Using comparative genomic, network, and phylogenetic methods, this thesis reports the identification of lineage and species-specific expansions of effectors. Despite their ubiquity and the threat that oomycetes pose to global food security, there is a lack of dedicated tools to analyse oomycete genomes. To this end, the Oomycete Gene Order Browser (OGOB) was developed. OGOB is a database and novel tool that facilitates comparative genomic and syntenic analyses of oomycete genomes. Analyses using OGOB highlighted the high degree of syntenic conservation within oomycete genera. Furthermore, tandem gene duplication was shown to play a significant role in the expansion and evolution of effector proteins. The data presented in this thesis also describes the first large-scale genomic and proteomic investigations of the widespread phytopathogens *Phytophthora chlamydospora*, *Phytophthora gonapodyides* and *Phytophthora pseudosyringae*. Mass spectrometry analyses identified approximately 300 extracellular proteins per species, many of which are putativly involved in infection or osmotrophy. The expression of approximately 3,000 proteins for each species was validated at the protein level. Comparative genomic analysis of CAZymes suggest that oomycete lifestyles may be linked to their CAZyme repertoires. Overall, the data presented in this thesis expands our knowledge of oomycete genome evolution.

Preface

This thesis is presented in the "Thesis by Publication" format, comprising a published review chapter and three published research articles. Each chapter contains its own bibliography in the format used by the associated journals. Chapter 1 is a review chapter that introduces the oomycetes and discusses the latest findings in oomycete genomics research and mentions some of the results described in later chapters. The first results chapter, Chapter 2, comprises a comparative genomics publication that uses bioinformatics methods to study the evolution of effector proteins across a large dataset of oomycete genomes. Chapter 3 describes the Oomycete Gene Order Browser, a newly developed tool and database for oomycete genomes, that facilitates analyses of oomycete genes and genomes. Chapter 4 reports the genome sequencing of three widespread *Phytophthora* species – *Ph. chlamydospora*, *Ph. gonapodyides* and *Ph. pseudosyringae*. This chapter uses a combination of bioinformatics, comparative genomics and proteomics methods to comprehensively characterise their nuclear genomes, mitochondrial genomes, predicted proteomes, *in silico* secretomes, *in vivo* extracellular proteomes and *in vivo* mycelial proteomes. The thesis concludes with Chapter 5, which is a general discussion of the results described in chapters 2 – 4.

Chapter 1
Introduction

This chapter has been published in Advances in Genetics.

McGowan, J., Fitzpatrick, D.A., 2020. Recent Advances in Oomycete Genomics. Advances in Genetics.

Abstract

The oomycetes are a class of ubiquitous, filamentous microorganisms that include some of the biggest threats to global food security and natural ecosystems. Within the oomycete class are highly diverse species that infect a broad range of animals and plants. Some of the most destructive plant pathogens are oomycetes, such as *Phytophthora infestans*, the agent of potato late blight and the cause of the Irish famine. Recent years have seen a dramatic increase in the number of sequenced oomycete genomes. Here we review the latest developments in oomycete genomics and some of the important insights that have been gained. Coupled with proteomic and transcriptomic analyses, oomycete genome sequences have revealed tremendous insights into oomycete biology, evolution, genome organization, mechanisms of infection, and metabolism. We also present an updated phylogeny of the oomycete class using a phylogenomic approach based on the 65 oomycete genomes that are currently available.

1. The Oomycetes

Oomycetes are filamentous, microbial eukaryotes that morphologically resemble fungi (Beakes et al., 2012), but belong to the species-rich group of stramenopiles (Burki et al., 2020), which contains other diverse organisms such as diatoms and brown algae (**Figure 1**). Oomycete species are highly diverse in terms of their lifestyles, pathogenicity and host ranges. More than 60% of known oomycete species are pathogens of plants (Thines and Kamoun, 2010). Together, they represent one of the biggest threats to global food security and natural ecosystems. Most notorious among oomycete species is *Phytophthora infestans*, a hemibiotrophic pathogen that causes late blight of tomato and potato (Haas et al., 2009), and was the causative agent of the Irish potato famine which resulted in the death of one million people and the displacement of another million. *Phytophthora sojae* is another highly destructive species that causes billions of dollars' worth of soybean crop loss each year (Tyler, 2007). *Phytophthora infestans* and *Ph. sojae* are examples of species with narrow host ranges. In contrast, *Phytophthora ramorum*, the "sudden oak death" pathogen, has a wide host range that can infect more than 100 host species (Rizzo et al., 2005). *Phytophthora cinnamomi* has a very wide host range and is thought to be able to infect more than 3,000 host species (Hardham, 2005). Virtually all dicotyledon plants are susceptible to infection by one or more *Phytophthora* species (Kamoun, 2003). Other economically devastating oomycete species include the obligate biotrophic downy mildews, *Bremia lactucae* and *Plasmopara viticola*, some of the most important pathogens of lettuce and grapevine, respectively (Dussert et al., 2019; Fletcher et al., 2019).

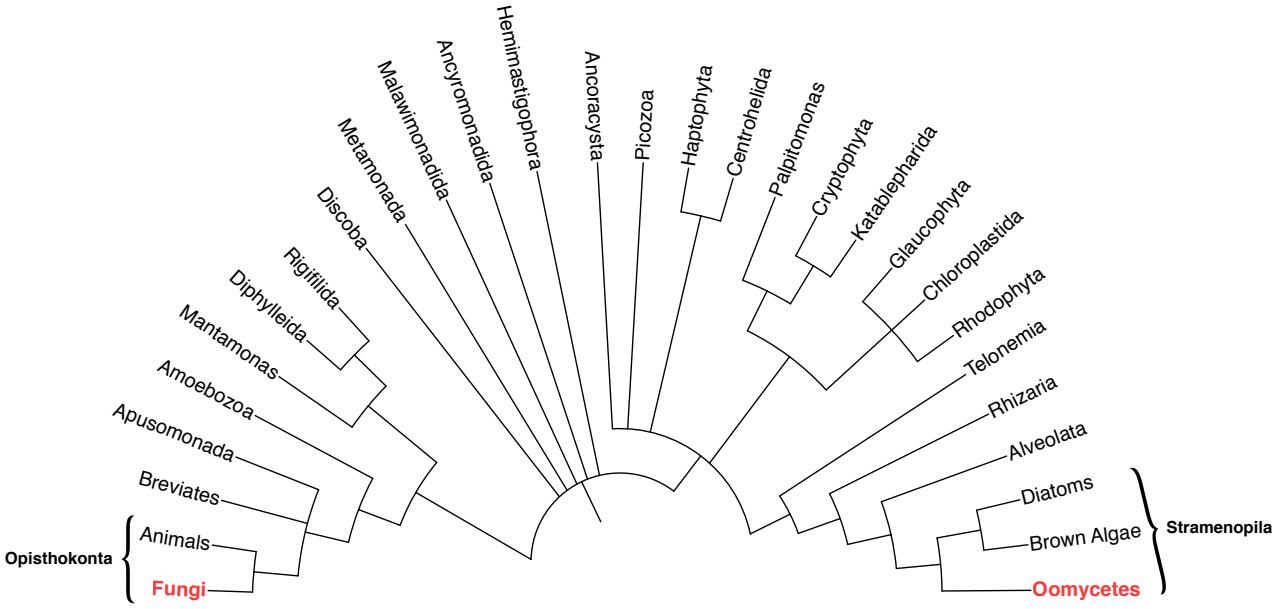


Figure 1. Simplified phylogeny of the eukaryotes showing the distant relationship between oomycetes and fungi. Adapted from Burki et al. (2020).

While the majority of attention is typically placed on plant pathogenic oomycetes, many oomycete species cause infections in other economically and ecologically important organisms. For example, *Saprolegnia* species are pathogens of amphibians, crustaceans, fish, and insects (Jiang et al., 2013). In particular, *Saprolegnia parasitica* is a freshwater fish pathogen that is predominant in farmed salmon populations, resulting in losses of more than 10% (van den Berg et al., 2013). *Pythium oligandrum* and *Pythium periplocum* are pathogens of fungi and oomycetes, acting as powerful biocontrol agents that can combat plant-pathogenic fungi and oomycetes (Benhamou et al., 2012; Kushwaha et al., 2017a). *Pythium guiyangense* is a highly virulent pathogen of mosquitos and is a potential biocontrol agent to efficiently manage mosquito populations (Shen et al., 2019). Some oomycete species, such as *Pythium insidiosum*, can cause fatal infections in humans and other mammals typically resulting in amputations or even death (Gaastra et al., 2010).

The oomycete class is divided into four "crown" orders - the Peronosporales, Pythiales, Albuginales, and Saprolegniales. The Peronosporales order is the best studied order and includes the phytopathogenic *Phytophthora*, *Nothophytophthora* and *Phytopythium* genera as well as the downy mildew *Bremia, Hyaloperonospora, Peronospora, Plasmopara, Pseudoperonospora* and *Sclerospora* genera (Fletcher et al., 2019; McCarthy and Fitzpatrick, 2017). The Pythiales order includes the *Pythium* genus and *Pilasporangium* (Adhikari et al., 2013). Within this order are a mix of animal, fungal, plant and oomycete pathogens. The Albuginales order includes the plant-parasitic *Albugo* genus which causes "white blister rust" on many valuable crop species (Kemen et al., 2011; Links et al., 2011). The Saprolegniales order is the most basal of the four crown orders and includes animal and plant pathogens. Genera belonging to the Saprolegniales order include *Achlya*, *Aphanomyces*, *Saprolegnia* and *Thraustotheca* (Misner et al., 2015).

Oomycetes were previously thought to be fungi due to similar morphological characteristics, filamentous growth, osmotrophic uptake of nutrients and similar ecological roles (Leonard et al., 2018; Richards et al., 2006). Even today, oomycetes are commonly mistakenly referred to as fungi. Several key differences distinguish oomycetes ("pseudo-fungi") from "true-fungi", some of which are summarized in **Table 1.** Phylogenetically, fungi branch with animals within the Opisthokonta (**Figure 1**), whereas the oomycetes are stramenopiles within the SAR (Stramenopila, Alveolata, and Rhizaria) supergroup placing them more closely related to brown algae and diatoms (**Figure 1**) (Burki et al., 2020).

Table 1. Typical features that differ between oomycetes and fungi

Character	Oomycetes	True Fungi
Neighboring taxonomic groups	Diatoms and brown algae	Animals
Hyphal structure	Aseptate and multinucleate	Either single cells or septate hyphae, with one or more nuclei per compartment
Ploidy of vegetative hyphae	Diploid	Typically haploid or dikaryotic
Cell wall composition	Cellulose and β-1,3 and β-1,6 glucans	Chitin and β-1,3 and β-1,6 glucans
Pigmentation	Usually unpigmented	Hyphae and/or spores are commonly pigmented
Secondary metabolites	None described	Common
Motile asexual spores	Biflagellated zoospores with two different types of flagella present – a tinsel anterior flagellum and a smooth posterior whiplash flagellum	Rare but some exceptions such as chytrids, which have monoflagellated zoospores
Sexual spores	Oospores	Various types

Adapted from Judelson & Blanco (2005).

Oomycetes cause infection by secreting large arsenals of effector proteins that breakdown host cell components, modulate host immune responses and trigger host necrosis (Kamoun, 2006; McGowan and Fitzpatrick, 2017). Effector families can be divided into two broad categories based on where they localize – apoplastic effectors that act outside the host cell and cytoplasmic effectors that enter the host cell (Wawra et al., 2012). Apoplastic effectors families include large families of hydrolytic enzymes such as glycoside hydrolases and proteases, as well as toxins including the necrosis-inducing proteins (NLPs) and phytotoxic protein family (PcF) proteins. Cytoplasmic oomycete effectors include the RxLR and Crinkler (CRN) family of effectors, which contain conserved translocation motifs that facilitate their delivery inside host cells (Haas et al., 2009; Tyler et al., 2006).

In recent years, in line with advancements in next-generation sequencing technologies, there has been an increased pace of oomycete genome sequencing. At the time of writing, 65 oomycete species (**Table 2**) have publicly available genome sequences deposited in databases such as the NCBI GenBank (Benson et al., 2012). Such genome sequencing projects have yielded tremendous insights into oomycete biology, evolution, effector arsenals, genome organization, metabolism and mechanisms of infection. Here we review recent developments in oomycete genomics and highlight some of the important findings that have been revealed.

Table 2. List of the 65 oomycete genome assemblies that are available at the time of writing.

Species	**Clade** [a]	**Order**	**Host**	**Genome Size Mb** [b]	**Reference** [c]
Albugo candida		Albuginales	Plants	45.3	(Links et al., 2011)
Albugo laibachii		Albuginales	Plants	35.5	(Kemen et al., 2011)
Lagenidium giganteum		Lagenidiales	Insects	126.3	PRJNA433680
Paralagenidium karlingii		Lagenidiales	Mammals	49.4	PRJNA433826
Bremia lactucae		Peronosporales	Plants	115.9	(Fletcher et al., 2019)
Hyaloperonospora arabidopsidis		Peronosporales	Plants	81.6	(Baxter et al., 2010)
Nothophytophthora valdiviana		Peronosporales	Plants	84.4	(Studholme et al., 2019)
Peronospora belbahrii		Peronosporales	Plants	35.4	(Thines et al., 2019)
Peronospora effusa		Peronosporales	Plants	32.1	(Fletcher et al., 2018)
Peronospora tabacina		Peronosporales	Plants	63.1	(Derevnina et al., 2015)
Phytophthora nicotianae	1	Peronosporales	Plants	80.0	(Liu et al., 2016)
Phytophthora parasitica	1	Peronosporales	Plants	82.4	PRJNA259235
Phytophthora cactorum	1a	Peronosporales	Plants	59.3	(Armitage et al., 2018)

Table 2. Continued.

Species	Clade [a]	Order	Host	Genome Size Mb [b]	Reference [c]
Phytophthora infestans	1c	Peronosporales	Plants	240.0	(Haas et al., 2009)
Phytophthora colocasiae	2a	Peronosporales	Plants	56.6	(Vetukuri, Kushwaha, et al., 2018)
Phytophthora capsici	2b	Peronosporales	Plants	95.2	(Cui et al., 2019; Lamour et al., 2012)
Phytophthora citricola	2c	Peronosporales	Plants	50.3	PRJNA555328
Phytophthora multivora	2c	Peronosporales	Plants	40.1	(Studholme et al., 2015)
Phytophthora plurivora	2c	Peronosporales	Plants	40.4	(Vetukuri, Tripathy, et al., 2018)
Phytophthora pluvialis	3	Peronosporales	Plants	53.6	(Studholme et al., 2015)
Phytophthora litchii	4	Peronosporales	Plants	38.2	(Ye et al., 2016)
Phytophthora megakarya	4	Peronosporales	Plants	101.2	(Ali et al., 2017)
Phytophthora palmivora	4	Peronosporales	Plants	107.4	(Ali et al., 2017)
Phytophthora agathidicida	5	Peronosporales	Plants	37.3	(Studholme et al., 2015)
Phytophthora pinifolia	6b	Peronosporales	Plants	94.6	(Feau et al., 2016)
Phytophthora alni var. alni	7a	Peronosporales	Plants	236.0	(Feau et al., 2016)

Table 2. Continued.

Species	Clade [a]	Order	Host	Genome Size Mb [b]	Reference [c]
Phytophthora cambivora	7a	Peronosporales	Plants	230.6	(Feau et al., 2016)
Phytophthora fragariae	7a	Peronosporales	Plants	73.68	(Gao et al., 2015)
Phytophthora rubi	7a	Peronosporales	Plants	74.7	(Tabima et al., 2017)
Phytophthora pisi	7b	Peronosporales	Plants	58.9	PRJEB6298
Phytophthora sojae	7b	Peronosporales	Plants	95.0	(Tyler et al., 2006)
Phytophthora cinnamomi	7c	Peronosporales	Plants	54.0	(Studholme et al., 2015)
Phytophthora cryptogea	8a	Peronosporales	Plants	63.8	(Feau et al., 2016)
Phytophthora lateralis	8c	Peronosporales	Plants	52.4	(Feau et al., 2016)
Phytophthora ramorum	8c	Peronosporales	Plants	65.0	(Tyler et al., 2006)
Phytophthora kernoviae	10	Peronosporales	Plants	39.4	(Feau et al., 2016)
Phytophthora taxon totara		Peronosporales	Plants	55.6	(Studholme et al., 2015)
Phytopythium vexans		Peronosporales	Plants	33.9	(Adhikari et al., 2013)
Plasmopara halstedii		Peronosporales	Plants	75.3	(Sharma et al., 2015)

Table 2. Continued.

Species	Clade [a]	Order	Host	Genome Size Mb [b]	Reference [c]
Plasmopara muralis		Peronosporales	Plants	59.5	(Dussert et al., 2019)
Plasmopara obducens		Peronosporales	Plants	295.3	PRJNA287843
Plasmopara viticola		Peronosporales	Plants	101.3	(L. Yin et al., 2017)
Pseudoperonospora cubensis		Peronosporales	Plants	64.3	PRJNA80635
Pseudoperonospora humuli		Peronosporales	Plants	40.5	(Rahman et al., 2019)
Sclerospora graminicola		Peronosporales	Plants	253.7	(Nayaka et al., 2017)
Pilasporangium apinafurcum		Pythiales	Plants	37.4	(Uzuhashi, Endoh, Manabe, & Ohkuma, 2017)
Pythium aphanidermatum	A	Pythiales	Plants	35.9	(Adhikari et al., 2013)
Pythium arrhenomanes	B	Pythiales	Plants	44.7	(Adhikari et al., 2013)
Pythium guiyangense		Pythiales	Insects	110.2	(Shen et al., 2019)
Pythium insidiosum	C	Pythiales	Mammals	53.2	(Rujirawat et al., 2015)
Pythium irregulare	F	Pythiales	Plants	42.9	(Adhikari et al., 2013)

Table 2. Continued.

Species	Clade [a]	Order	Host	Genome Size Mb [b]	Reference [c]
Pythium iwayamai	G	Pythiales	Plants	43.3	(Adhikari et al., 2013)
Pythium oligandrum	D	Pythiales	Fungi/Oomycetes	36.8	(Kushwaha et al., 2017a)
Pythium periplocum	D	Pythiales	Fungi/Oomycetes	35.9	(Kushwaha et al., 2017b)
Pythium splendens	I	Pythiales	Plants	53.4	PRJNA548776
Pythium ultimum var. sporangium	I	Pythiales	Plants	37.7	(Adhikari et al., 2013)
Pythium ultimum var. ultimum	I	Pythiales	Plants	42.8	(Lévesque et al., 2010)
Achlya hypogyna		Saprolegniales	Crustaceans	43.4	(Misner et al., 2015)
Aphanomyces astaci		Saprolegniales	Crustaceans	45.3	(Gaulin et al., 2018)
Aphanomyces euteiches		Saprolegniales	Plants	56.9	(Gaulin et al., 2018)
Aphanomyces invadans		Saprolegniales	Fish	71.4	PRJNA188082
Aphanomyces stellatus		Saprolegniales	Fish	62.1	(Gaulin et al., 2018)
Saprolegnia diclina		Saprolegniales	Fish	62.9	PRJNA86859

Table 2. Continued.

Species	Clade [a]	Order	Host	Genome Size Mb [b]	Reference [c]
Saprolegnia parasitica		Saprolegniales	Fish	42.3	(Jiang et al., 2013)
Thraustotheca clavata		Saprolegniales	Free living	39.1	(Misner et al., 2015)

[a] Clade designation according to Blair et al., (2008) for *Phytophthora* and de Cock et al., (2015) for *Pythium*.

[b] Genome size as reported in corresponding publications, otherwise the assembly size is shown

[c] NCBI BioProject accessions are shown for genome assemblies that do not have an associated publication.

2. Oomycete Genomes

In 2006, the first oomycete genome sequences were published for *Ph. ramorum* and *Ph. sojae* (Tyler et al., 2006). Whole-genome shotgun sequencing produced a nine-fold coverage assembly of the 95 Mb *Ph. sojae* genome and a seven-fold coverage assembly of the 65 Mb *Ph. ramorum* genome (**Table 2**). Comparative genomic analysis identified rapid expansion/loss and diversification of secreted protein families associated with pathogenicity, including glycoside hydrolases, proteases and effector families such as elicitins, NLPs and the cytoplasmic RxLR and CRN families (Jiang et al., 2006a, 2006b; Tyler et al., 2006). These analyses showed that the secretome of both species has undergone rapid diversification and is evolving at a faster rate than the overall proteome, as 11% and 17% of the secreted proteins were found to be unique to *Ph. ramorum* and *Ph. sojae*, respectively, compared to 4% and 9% of the overall proteomes of each species respectively. Large multigene families in each species were also reported. The genome sequences show extensive collinearity/synteny of orthologs between the two species, with over 75% of exons aligning in a whole-genome alignment (Tyler et al., 2006).

The *Ph. infestans* genome was published in 2009, revealing a much larger genome assembly size of 229 Mb (**Table 2**) and extremely high repeat content comprising 74% of the genome assembly, compared to 28% and 13% in *Ph. sojae* and *Ph. ramorum* respectively (Haas et al., 2009). The *Ph. infestans* genome also contains large expansions of effector proteins, in particular ~60% more putative RxLR effectors were identified in *Ph. infestans* than *Ph. ramorum* and *Ph. sojae*. Analyses of genome organization showed the existence of genomic regions with high gene density and low repeat content that are syntenically conserved in the three sequenced *Phytophthora* genomes. These blocks are separated by regions with low gene density and expansions of repeat content and mobile genetic elements. These regions are typically populated with rapidly evolving effector

proteins (Haas et al., 2009). This non-random distribution of effector genes in more rapidly evolving, gene-sparse regions of genomes gave rise to the model of "two-speed genomes" (Dong et al., 2015).

The genome of *Pythium ultimum var. ultimum* was published in 2010, the first necrotrophic oomycete and first *Pythium* genome to be sequenced (Lévesque et al., 2010). *Pythium ultimum* is a cosmopolitan plant pathogen with a broad host range, infecting corn, soybean, wheat and other crops (Cheung et al., 2008). The *Py. ultimum* genome is smaller (43 Mb) (**Table 2**) than the *Phytophthora* species that had been sequenced at that time (Lévesque et al., 2010). Genome annotation revealed an expansion of genes involved in proteolysis but a reduction in the number of genes involved in carbohydrate metabolism as well as fewer CRN genes, and no RxLR effectors (Lévesque et al., 2010). These findings led the authors to suggest that *Pythium spp.* has differences in infection mechanisms.

The genome of the obligate biotrophic downy mildew *Hyaloperonospora arabidopsidis* was also published in 2010 (Baxter et al., 2010). Compared to the sequenced *Phytophthora* species, the 78 Mb *Hy. arabidopsidis* (**Table 2**) genome assembly showed a reduction in the number of effector genes (such as RxLRs, Crinklers, elicitins, and NLPs) and secreted degradative enzymes. The *Hyaloperonospora* genome also exhibits losses of genes involved in metabolic pathways, such as genes required for nitrogen and sulfur assimilation (Baxter et al., 2010). Together, these findings suggest a change in infection strategy and signatures of obligate biotrophy evolution. *Hyaloperonospora arabidopsidis*, like most downy mildews, does not form zoospores. Comparative analysis revealed the loss of approximately 80% of *Ph. infestans* flagellar proteins in *Hy. arabidopsidis.* Gene remnants can be detected for the missing flagellar

proteins, suggesting that the loss of flagella from the downy mildews was a relatively recent event (Judelson et al., 2012).

Two *Albugo* (white blister rust) genomes were published in 2011 - *Albugo laibachii* (Kemen et al., 2011) and *Albugo candida* (Links et al., 2011). *Albugo* species are obligate biotrophic plant pathogens that belong to the Albuginales order. They have evolved obligate biotrophy independently from the downy mildew pathogens within the Peronosporales order (Kemen et al., 2011). *Albugo laibachii* is a pathogen of *Arabidopsis thaliana* and *Al. candida* is a pathogen of important *Brassica* species, including canola, oilseed mustard, and cabbage family vegetables. Compared to *Hy. arabidopsidis* and the other sequenced oomycetes, the two *Albugo* genome assemblies are much smaller in size at 35 – 37 Mb (**Table 2**), contain a lower proportion of repetitive sequences and have higher gene density (Kemen et al., 2011; Links et al., 2011). A novel class of cytoplasmic effector proteins with an N-terminal "CHXC" motif was identified in *A. laibachii* (Kemen et al., 2011). Similar to *Hy. arabidopsidis*, both *Albugo* species lack a number of important metabolic enzymes, including genes required for nitrate and sulfate assimilation (Baxter et al., 2010; Kemen et al., 2011; Links et al., 2011). Moreover, both genome assemblies showed a reduction in the number of cell-wall degrading enzymes and effector proteins, and in particular, no NLPs were identified. These findings show convergent evolutionary signatures of obligate biotrophy between the *Albugo* species and *Hy. arabidopsidis* (Baxter et al., 2010; Kemen et al., 2011).

The genome of the fish pathogen *Sa. parasitica* was published in 2013, representing the first animal pathogen oomycete genome (Jiang et al., 2013). Analysis of the 42 Mb assembly (**Table 2**) revealed a large divergence in gene content compared to plant pathogenic oomycetes. There was a lack of RxLRs, CRNs, and NLPs, as well as proteins involved in the breakdown of plant cell walls, such as cutinases and pectin

modifying enzymes. *Saprolegnia parasitica* was shown to possess a large arsenal of 270 proteases, one of the largest of sequenced eukaryotes, that were shown to be expressed at different infection stages (Jiang et al., 2013).

The genome of the pepper and cucurbits pathogen *Phytophthora capsici* was published in 2012 (Lamour et al., 2012). The genome assembly is 64 Mb in length and is typical in terms of genome structure and gene content compared to previously sequenced *Phytophthora* genomes. However, the genome sequence revealed high levels of single nucleotide variation in addition to extensive loss of heterozygosity (LOH). LOH could be attributed to mating-type switches and loss of pathogenicity, contributing to the rapid adaption to fungicides and high levels of diversity within *Ph. capsici* populations (Lamour et al., 2012).

An additional five *Pythium* genomes (*Py. aphanidermatum*, *Py. arrhenomanes*, *Py. irregulare*, *Py. iwayami*, and *Py. ultimum var. sporangiiferum*) and *Phytopythium vexans* were published in 2013 and revealed compact genome sizes ranging from 34 Mb to 45 Mb (**Table 2**) (Adhikari et al., 2013). Similar to the genome sequence of *Pythium ultimum var. ultimum* (Lévesque et al., 2010), these *Pythium*/*Phytopythium* genomes were reported to lack RxLR effectors and possessed fewer numbers of genes involved in carbohydrate metabolism suggesting alternative virulence mechanisms (Adhikari et al., 2013). Synteny analysis of nine oomycete genomes and the distantly related diatom *Thalassiosira pseudonana* (a stramenopile outgroup) revealed a high degree of conservation of gene order between oomycetes that also extended partially with *T. pseudonana* (Adhikari et al., 2013). The animal pathogenic *Py. insidiosum*, isolated from a patient with vascular pythiosis, was later sequenced producing a genome assembly of 53 Mb (**Table 2**) (Rujirawat et al., 2015). This was followed by the genome sequencing of the mycoparasites *Py. periplocum* and *Py. oligandrum* (Kushwaha et al., 2017a,

2017b), which were assembled into 36 Mb and 37 Mb genome assemblies, respectively (**Table 2**). Compared to plant pathogenic *Pythium* species, the mycoparasitic *Pythium* species feature expansions of carbohydrate-active enzymes (CAZymes), including carbohydrate-binding modules, carbohydrate esterases, glycoside hydrolases, glycosyltransferases, polysaccharide lyases, and redox enzymes, that potentially play important roles during mycoparasitism (Kushwaha et al., 2017b). The genome sequence of the insect parasite *Pythium guiyangense* was sequenced using Illumina and PacBio platforms and resulted in a 110 Mb genome assembly (**Table 2**), more than double the size of previously sequenced *Pythium* species (Shen et al., 2019). The large genome size is thought to be the result of interspecies hybridization as most of the *Pythium* core genes are present in two copies, likely originating from two different parental species. Genomic and experimental evidence identified CRN effector proteins that are toxic to insect cells. *Py*. *guiyangense* CRNs are highly divergent as all showed sequence divergence of at least 50% when compared to CRNs from plant pathogenic *Pythium* species (Shen et al., 2019).

Reconstructing the ancestor of the last common ancestor of the Stramenopiles, using whole-genome data from six pathogenic oomycetes and four non-pathogenic Stramenochromes (diatoms and alga), revealed a large ancestral genome containing some ~10,000 genes (Seidl et al., 2012). Analyses of gene family evolution showed that oomycete genome evolution is under constant flux, continuously gaining and losing genes, with gene duplication events outnumbering loss events. The branch leading to the last common ancestor of the Peronosporales and the *Phytophthora* genus, in particular, is characterized by an increased frequency of duplication events, including expansions of many gene families associated with pathogenicity including proteins with signal peptides, carbohydrate metabolizing enzymes, RxLRs and CRNs. Additionally, large numbers of

species-specific gene duplication events were detected, which further contributes to the large gene family sizes of extant oomycetes (Seidl et al., 2012).

Genome sequencing of the two saprolegnian species *Achlya hypogyna* and *Thraustotheca clavata* and comparative genomic analysis of eight other oomycete genomes allowed for the reconstruction of an ancestral secretome (Misner et al., 2015). The reconstructed ancestral oomycete secretome consisted of at least 84 gene families encoding genes putatively involved in "carbohydrate metabolism and transport", "post-translational modification, protein turnover, chaperones", "signal transduction mechanisms" and "amino acid transport and metabolism" (Misner et al., 2015).

Genome sequences for the three *Aphanomyces* species – *Ap. astaci*, *Ap. euteiches,* and *Ap. stellatus* - revealed relatively consistent genome sizes ranging from 45 Mb for *Ap. astaci* to 62 Mb for *Ap. stellatus* (**Table 2**). This was in contrast to larger differences in the number of protein-coding genes which ranged from 16,479 for *Ap. astaci* to 25,573 for *Ap. stellatus* (Gaulin et al., 2018). Comparative genomic analysis revealed that host specialization is correlated with specialized secretomes. The plant pathogen *Ap. euteiches* encodes a large arsenal of diverse cell wall degrading enzymes. In contrast, the crayfish pathogen *Ap. astaci* encodes a large number of secreted proteases and enzymes predicted to target chitin (Gaulin et al., 2018), a major component of crustacean shells.

A highly contiguous, near chromosomal level genome assembly for the downy mildew *Bremia lactucae* was achieved by combining multiple sequencing technologies, resulting in a 91 Mb assembly (**Table 2**), 67.3% of which was identified as being repetitive (Fletcher et al., 2019). Resequencing and flow cytometry analysis of a large number of isolates identified a high prevalence of heterokaryosis (the state of having multiple genetically distinct nuclei within cells) which can lead to rapid changes in populations including differences in virulence levels and sporulation rates (Fletcher et al.,

2019). Sequencing of *Br. lactucae* adds to the growing list of downy mildew genomes available - which includes *Peronospora belbahrii*, *Peronospora effusa*, *Peronospora tabacina*, *Plasmopara halstedii*, *Plasmopara muralis*, *Plasmopara obducens*, *Plasmopara viticola*, *Pseudoperonospora cubensis*, *Pseudoperonospora humuli* and *Sclerospora graminicola* (**Table 2**) (Derevnina et al., 2015; Dussert et al., 2019; Fletcher et al., 2018; Nayaka et al., 2017; Rahman et al., 2019; Sharma et al., 2015; Thines et al., 2019; L. Yin et al., 2017). Comparative genomic and phylogenomic analysis of downy mildews and other Peronosporales lineages revealed *Bremia* and *Plasmopara* species to be more closely related to *Phytophthora* clade 1 (which includes *Ph. cactorum*, *Ph. infestans*, *Ph. nicotianae* and *Ph. parasitica*) than to the other downy mildew lineages, suggesting convergent evolution of biotrophy (Fletcher et al., 2019; McCarthy and Fitzpatrick, 2017; Sharma et al., 2015; L. Yin et al., 2017).

Interestingly, based on hierarchical clustering of metabolic networks, oomycete species can be grouped according to their lifestyle regardless of their phylogenetic relationships (Thines et al., 2019). This includes the clustering of the distantly related *Albugo* species with the downy mildews *Hy. arabidopsidis*, *Pe. effusa*, *Pe. tabacina*, and *Pl. halstedii*. This further suggests the convergent evolution of biotrophic lifestyles. These downy mildew species were predicted to have fewer genes, enzymes, reactions, and metabolites in metabolic networks compared to other oomycete species (Thines et al., 2019). Furthermore, *B. lactucae, Hy. arabidopsidis*, *Pe. belbahrii*, *Pe. effusa*, *Pe. tabacina* were shown to have undergone an extensive reduction in the number of calcium-binding and flagella associated domains. This reduction was not observed in *Plasmopara* species, which are known to produce zoospores, suggesting these missing genes are associated with the production and development zoospores (Fletcher et al., 2019, 2018; Thines et al., 2019).

The genome sequences for a large number of *Phytophthora* forest pathogens have been sequenced in recent years including *Ph. agathidicida*, *Ph. cambivora*, *Ph. cinnamomi*, *Ph. cryptogea*, *Ph. kernoviae*, *Ph. lateralis*, *Ph. multivora*, *Ph. pinifolia*, *Ph. plurivora*, *Ph. pluvialis*, *Ph. taxon totara* and the hybrid *Ph. alni var. alni* (**Table 2**) (Feau et al., 2016; Studholme et al., 2015; Vetukuri et al., 2018b). Additionally, a number of important *Phytophthora* crop pathogens have been sequenced including the cacao black pod rot pathogens *Ph. megakarya* and *Ph. palmivora* (Ali et al., 2017), the strawberry pathogens *Ph. cactorum* and *Ph. fragariae* (Armitage et al., 2018; Gao et al., 2015), the litchi pathogen *Ph. litchii* (Ye et al., 2016), the raspberry pathogen *Ph. rubi* (Tabima et al., 2017), and the taro crop pathogen *Ph. colocasiae* (Vetukuri et al., 2018a). The first *Nothophytophthora* genome, a sister genus of *Phytophthora*, was sequenced in 2019 and tentatively identified as *Nothophytophthora valdiviana* (Studholme et al., 2019). Such genome sequencing projects facilitate investigations into genome evolution and cataloging of effector arsenals of *Phytophthora* species with different host ranges.

3. Oomycete Phylogenomics

Initial phylogenetic analyses of oomycete species were either based on single highly conserved genes, or on a small selection of highly conserved markers such as the internal transcribed spacer (ITS) region, nuclear genes (e.g. β-tubulin and translation elongation factor 1α) and mitochondrial genes (e.g. cox1 and nad1) (Martin et al., 2014). Performing phylogenetic analyses on highly conserved markers may be unable to accurately distinguish taxa at the species level. For example, some closely related *Phytophthora* species share identical ITS, tigA, β-tub, or cox1 genes (Yang and Hong, 2018). These genetic markers are phylogenetically uninformative in these scenarios. Furthermore, conflicts can occur between phylogenies depending on which markers are used. The availability of whole-genome sequences has made it possible to perform whole-genome phylogenetic or phylogenomic analyses of oomycete species. Phylogenomic analyses are based on larger number of genes using genome scale data and allow for other types of phylogenetic markers such as gene duplication events.

A multi-locus phylogeny of 82 *Phytophthora* species divided the genus into 10 well supported clades (Blair et al., 2008). A follow up analysis combining multiple nuclear and mitochondrial loci added further support for the 10 *Phytophthora* clades (Martin et al., 2014). However, the relationships between clades varied depending on which markers were used. The *Pythium* genus was divided into 10 clades (labelled A – J) based on ITS and ribosomal DNA sequences (LéVesque and De Cock, 2004). The *Pythium* genus itself is polyphyletic, separated into two monophyletic subgroupings of clades A – E and clades F – J (McCarthy and Fitzpatrick, 2017). Based on ITS sequences *Phytopythium vexans* (previously *Pythium vexans*) was initially categorized as belonging to *Pythium* clade K (LéVesque and De Cock, 2004) but later reclassified into its own genus *Phytopythium* belonging to the Peronosporales order based on a multi-locus

analysis of nuclear and mitochondrial genes (de Cock et al., 2015), placing this species as an intermediate between *Phytophthora* and *Pythium*.

A recent phylogenomic analysis of the oomycete class used a variety of phylogenomic methods and multiple datasets including a supermatrix analysis of 352 single-copy gene families (7,744 genes) ubiquitously present in 22 Peronosporales, and a supertree analysis of 2,280 ubiquitous single-copy gene families (35,622 genes) as well as 8,355 gene families (209,904 genes) present in at least 4 of 37 oomycete species investigated (McCarthy and Fitzpatrick, 2017). All methods resolved the four crown oomycete orders. Additionally, all methods suggested that *Phytophthora* is a paraphyletic genus, that gave rise to the downy mildews. Furthermore, these results show that the downy mildews are polyphyletic and that obligate biotrophy has evolved at least twice within the Peronosporales (McCarthy and Fitzpatrick, 2017). Further support for the polyphyly of downy mildews was reported in an analysis using six nuclear and mitochondrial loci of 118 taxa within the Peronosporales, including 13 downy mildews (Bourret et al., 2018). An additional phylogenomic analysis based on the concatenation of 49 conserved genes from 13 Peronosporales assemblies (3 *Phytophthora* and 10 downy mildew) provided further evidence for the polyphyly of this group (Fletcher et al., 2018).

Complete genome sequences also allow for the identification of orthologous proteins that can be used in molecular clock analyses to estimate divergence times among species. Molecular clock analyses are particularly useful for taxa that have patchy evidence in the fossil records, such as microbial eukaryotes that are often indistinguishable due to a lack of discriminative morphological characteristics (Matari and Blair, 2014). The earliest fossil evidence for oomycetes comes from the Devonian period ~408 million years ago (MYA) (Krings et al., 2011). A molecular clock analysis of 12 oomycetes and 6 outgroups was performed using 40 orthologs involved in gene

expression regulation, that are highly conserved across 18 genomes (Matari and Blair, 2014). Three molecular clock methods were employed (strict clock, UCLD relaxed clock and random local clock) and calibrated with fossil evidence from oomycetes and diatoms. The results suggest that the oomycetes arose some 430 – 400 MYA, with the two major branches leading to the Peronosporaleans and Saprolegnialeans diverging 225 – 190 MYA (Matari and Blair, 2014). These findings indicate that the oomycetes arose later than their hosts and that the evolution of pathogenicity may have been influenced by environmental changes or facilitated by events such as HGT from pathogenic bacteria and fungi (Matari and Blair, 2014). A follow-up analysis was performed using the random local clock model with the same calibration points and same dataset except for the addition of three *Aphanomyces* genomes (*Ap. astaci*, *Ap. euteiches* and *Ap. stellatus)* (Gaulin et al., 2018). This analysis estimated that *Aphanomyces* diverged from other Saprolegniales species more than 100 MYA, and the three *Aphanomyces* species diverged from each other more than 50 MYA indicating relatively recent host specialization of animal and plant pathogenic *Aphanomyces* species (Gaulin et al., 2018).

Here we present an updated phylogeny of the oomycete class based on the currently available genome sequences (**Table 2**) using a phylogenomic approach. We include the genome of the non-oomycete Stramenopile *Hyphochytrium catenoides* as an outgroup (Leonard et al., 2018). As the majority of genome sequences deposited in NCBI GenBank do not have associated gene models, we used Benchmarking Universal Single-Copy Orthologs (BUSCOs) as phylogenetic markers (Waterhouse et al., 2018). BUSCOs are highly conserved genes that are expected to be found in a genome only as a single copy. BUSCO analysis using the "Alveolata-Stramenopiles" dataset revealed 102 BUSCO families that are present and single-copy in 84% of the 66 genome assemblies, i.e. they are present and single-copy in at least 56 of the 66 species. Each BUSCO family

was individually aligned using MUSCLE (Edgar, 2004). Alignments were subsequently trimmed using trimAl with the parameter "-automated1" which uses a heuristic method to decide which trimming method is most appropriate (Capella-Gutierrez et al., 2009). Trimmed alignments were then concatenated, resulting in a supermatrix of 44,739 amino acid residues. Maximum-likelihood analysis was performed using IQ-TREE (Nguyen et al., 2015) under the JTT+F+R7 model, which was the best fit model according to ModelFinder (Kalyaanamoorthy et al., 2017). Support for each node was calculated based on 100 bootstrap replicates. The phylogeny was visualized and annotated using the Interactive Tree of Life (iTOL) (Letunic and Bork, 2007).

Our phylogeny recovered all genera with generally high bootstrap support (**Figure 2**) and is in broad agreement with previous phylogenetic and phylogenomic studies. All *Phytophthora* species were recovered in their expected clades (**Figure 2**). The relationships within the Peronosporales order, including between individual *Phytophthora* clades, are in complete agreement with a smaller phylogenomic analysis of Peronosporales species based on 18 BUSCO proteins (Fletcher et al., 2019). Our phylogeny groups *Phytophthora* clades 6 (*Ph. pinifolia)* and 7, which is in agreement with a previous phylogeny based on seven nuclear loci and also a multispecies coalescent approach based on two nuclear and four mitochondrial loci (Martin et al., 2014), although we obtained a low bootstrap support value of 49% (**Figure 2**). The relationships between *Phytophthora* clades 2 - 5 in our phylogeny differ from previous nuclear and mitochondrial phylogenies (Blair et al., 2008; Martin et al., 2014). Our phylogeny groups *Phytophthora* clades 2 and 3 (*Ph. pluvialis*) together to the exclusion of clade 4 (**Figure 2**). Although we obtained low bootstrap support (52%) for the grouping of clades 2 and 3, it agrees with previous phylogenomic analyses (Fletcher et al., 2019; McCarthy and Fitzpatrick, 2017). Our results confirm that *Phytophthora* is a paraphyletic genus,

suggesting that clades 1 – 4 are more closely related to each other than to those in clades 5 – 10 (**Figure 2**). *Phytophthora* clades 3, 5, and 6 currently only have one sequenced representative species each. *Phytophthora* clade 9 is currently the only *Phytophthora* clade lacking genomic data. Sequencing of multiple species from each *Phytophthora* clade will likely lead to increased resolution as to the relationships between clades.

Based on the current genome sequences available, it appears that the downy mildews are a polyphyletic group that arose from *Phytophthora* species, with two phylogenetically distinct clades that convergently evolved obligate biotrophy (**Figure 2**). This is in agreement with previous studies (Bourret et al., 2018; Fletcher et al., 2019, 2018; McCarthy and Fitzpatrick, 2017). The first clade consists of *Bremia* and *Plasmopara* which are closely related to *Phytophthora* clade 1 (100% bootstrap support) (**Figure 2**). The second clade contains *Hyaloperonospora*, *Sclerospora*, *Pseudoperonospora*, and *Peronospora* species, which groups with *Ph. taxon totara* (100% bootstrap support) (**Figure 2**). Therefore, obligate biotrophy has evolved at least twice within the Peronosporales order. Obligate biotrophy has evolved independently a third time in the Albuginales order (**Figure 2**).

Nothophytophthora was placed as a sister to the *Phytophthora* genus (**Figure 2**). *Phytopythium vexans* is placed intermediate to the Peronosporales and Pythiales orders (**Figure 2**). As seen in previous studies, the *Pythium* genus is polyphyletic with clades A – D grouping together, and clades F – I grouping together (**Figure 2**). Genomic data is currently not available for *Pythium* clades E, H or J. The insect pathogen *Py. guiyangense* is grouped with the plant pathogenic *Py. irregulare* and *Py. iwayamai* (100% bootstrap support), while the mammal pathogen *Py. insidiosum* is grouped with the plant pathogenic *Py. aphanidermatum* and *Py. arrhenomanes* (77% bootstrap support) (**Figure**

2). This suggests that animal pathogenicity has evolved independently multiple times within the *Pythium* genus.

Lagenidium giganteum is grouped with *Pythium* clades A – D and not *Paralagenidium karlingii* (**Figure 2**). This agrees with a six-gene phylogeny that groups *La. giganteum* with *Pythium* species to the exclusion of *Pa. karlingii*, indicating that *Paralagenidium* is phylogenetically unrelated to the main clades of the oomycetes (Spies et al., 2016). It also agrees with phylogenies based on ITS and cox2 sequences (Vilela et al., 2019). However, our phylogeny suggests that *Pa. karlingii* is even more distantly related as it places it intermediate to the Albuginales and Saprolegniales orders with 100% bootstrap support (**Figure 2**), this is in agreement with the phylogeny shown in (Beakes and Thines, 2016). The previous two phylogenetic analyses (Spies et al., 2016; Vilela et al., 2019) did not include Albuginales sequences. The availability of more genome sequences and whole gene sets will help to further clarify the relationships amongst contentious oomycete clades.

Figure 2. Maximum-likelihood phylogeny of the 65 oomycete species based on the concatenation of 102 conserved BUSCO sequences. The stramenopile *Hyphochytrium catenoides* is included as an outgroup. All nodes have 100% bootstrap support except where indicated. Species are colored according to order.

4. Oomycete Mitochondrial Genomes

As discussed above, mitochondrial genes have been useful for understanding the evolutionary relationships between oomycete species. Several oomycete mitochondrial genomes have been sequenced in full and analyzed, revealing variation in mitochondrial structure and gene content across oomycete lineages. Sequencing of the *Saprolegnia ferax* mitochondrion revealed a densely packed ~ 47 kb circular genome (**Figure 3A**) that encodes 37 protein and rRNA coding genes (18 respiratory chain proteins, 16 ribosomal proteins, the import protein secY, and the large and small ribosomal subunits) and 25 tRNA genes (Grayburn et al., 2004). An 8,618 kb inverted repeat was identified separating the genome into two single-copy regions (**Figure 3A**). The mitochondrion of *Ph. infestans* was determined to be a ~ 38 kb genome (**Figure 3B**) that is A+T rich (76%) and has high gene density with more than 95% of the genome coding (Paquin et al., 1997). The mitochondrial genome sequences of *Ph. ramorum* (**Figure 3C**) and *Ph. sojae* were determined to be ~ 39 kb and ~ 43 kb, respectively (Martin et al., 2007). Similar to *Sa. ferax*, both *Phytophthora* mitochondrial genomes encode 37 protein and rRNA coding genes and 26 (*Ph. ramorum*) or 25 (*Ph. sojae*) tRNA genes that specify 19 amino acids. *Ph. ramorum* possess a 1,150 bp inverted repeat encoding an additional tRNA that isn't present in *Ph. sojae*. Otherwise, the gene order between the two mitochondrial genomes is conserved. Comparison with the *Ph. infestans* mitochondrial genome revealed high conservation of genome collinearity except for two inversions that include 3 and 19 coding regions. Conservation of gene order also extended partially to that of *Sa. ferax*.

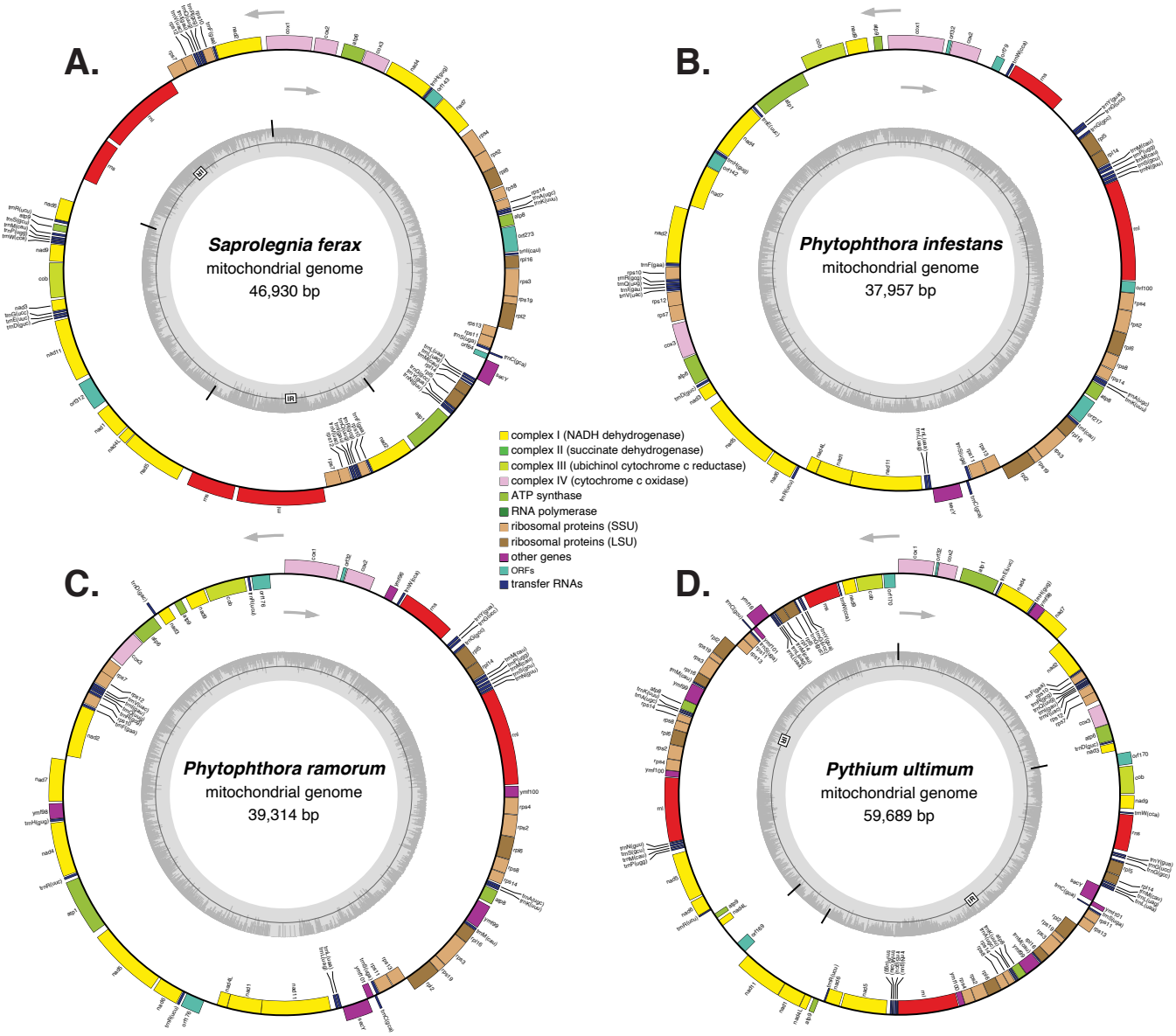


Figure 3. Structures of representative oomycete mitochondrial genomes. (**A**) *Saprolegnia ferax* (NC_005984.1), (**B**) *Phytophthora* infestans (NC_002387.1), (**C**) *Phytophthora ramorum* (NC_009384.1), and (**D**) *Pythium ultimum* (NC_014280.1). Mitochondrial genomes are not drawn to scale. Transcriptional orientation is indicated by arrows. GC content is shown in the inner circles. Inverted repeats are indicated with "IR". The small inverted repeat (1,150 bp) in *Ph. ramorum* is not shown.

The mitochondrial genomes of several clade 1c *Phytophthora* species were sequenced to understand the relationships within the clade, including *Ph. infestans*, *Ph. phaseoli*, *Ph. ipomoeae*, *Ph. mirabilis* and two lineages of *Ph. andina* (Lassiter et al., 2015). Mitochondrial genomes within clade 1c are highly conserved in terms of size, all mitochondrial assemblies are approximately 38 kb and only varied by 179 bp. Gene order is identical among all species and sequence identity of proteins is greater than 99% (Lassiter et al., 2015). The mitochondrial genome of *Py. ultimum* was shown to be larger at ~ 60 kb (**Figure 3D**). The increase in size is due to a large inverted repeat of ~ 22 kb (**Figure 3D**). The genome encodes the same set of protein and RNA encoding genes but many are duplicated in the inverted repeat region (Lévesque et al., 2010). The mitochondrial genome of *Py. insidiosum* was found to be similarly large at ~ 55 kb, containing 2 large inverted repeats (Tangphatsornruang et al., 2016).

Genome sequencing of the *Ph. nicotianae* mitochondria revealed a ~ 37.5 kb sequence that is identical to the previously sequenced *Phytophthora* mitochondrial genomes in terms of gene content (Yuan et al., 2017). Comparative genomic analysis of the *Ph. nicotianae* mitochondrion and 13 other oomycete mitochondrial genomes showed that large inverted repeats are absent in the mitochondria of *Phytophthora* species. This analysis showed that a ~10 kb inversion, containing 8 tRNAs and 11 genes, is common to *Ph. andina, Ph. impomoeae, Ph. infestans, Ph. mirabilis* and *Ph. phaseoli* relative to *Ph. nicotiana*, *Ph. polonica*, *Ph. ramorum* and *Ph. sojae*. Furthermore, while overall gene content is similar between oomycetes, the overall mitochondrial genome size and gene copy number are higher in the Pythiales and Saprolegniales due to the presence of duplications in inverted repeat regions (Yuan et al., 2017). The mitochondrial genome of *Pe. effusa* was determined to be ~41 kb in length and shares the same organization with the *Pe. tabacina* mitochondrial genome, except for a small inverted repeat less than 900

bp that is present in *Pe. effusa* (Derevnina et al., 2015; Fletcher et al., 2018). Sequencing of *Ps. humuli* revealed a 39 kb mitochondrial genome that is AT-rich (77.8%) and gene dense with coding regions representing 90% of the genome. The gene order of the *Ps. humuli* mitochondria is identical to *Ps. cubensis* but the small ribosomal subunit is encoded in the opposite direction compared to *Pe. tabacina* and *Phytophthora* species (Rahman et al., 2019).

5. The Impact of Horizontal Gene Transfer on Oomycete Evolution

Horizontal gene transfer (HGT), or lateral gene transfer (LGT), is the nonvertical transfer of genetic material between different species (Savory et al., 2015). HGT can have significant evolutionary consequences, such as facilitating recipient species to adapt to different ecosystems or exploit new hosts. Before oomycete genomes were first sequenced, it was known that HGT played an important role in the evolution of oomycetes. Phylogenetic analysis of *Ph. infestans* expressed sequence tag (EST) sequences identified an endopolygalacturonase gene (*pipg1*), encoding an enzyme involved in pectin breakdown, that's most closely related homologs belonged to fungi suggesting a possible HGT event from fungi to *Phytophthora* (Torto et al., 2002). Sequencing of the *Ph. ramorum* and *Ph. sojae* genome (Tyler et al., 2006) facilitated more in-depth analyses of HGT which led to the identification of four HGT events from ascomycete fungi to *Phytophthora* with strong phylogenetic evidence (Richards et al., 2006). Interestingly, each of the four genes are likely to be involved in osmotrophy-related functions, implying that HGT may have played a role in the convergent evolution of osmotrophy and filamentous growth between fungi and oomycetes. There is also evidence of HGT events from bacteria to oomycetes, including secreted cutinases (important virulence factors involved in the breakdown of the plant cuticle) that appear to have been transferred from Actinobacteria to oomycetes and later duplicated, with 16 copies being found in the *Ph. sojae* genome (Belbahri et al., 2008). Another analysis that examined metabolic enzymes from eukaryotic genomes showed that 2% of metabolic enzymes in the genomes of *Ph. ramorum* and *Ph. sojae* potentially originated via HGT (Whitaker et al., 2009).

The availability of more oomycete genomes has allowed for more comprehensive, whole-genome scans to identify genes that may have been gained via HGT. Similarly, the

availability of more non-oomycete genomes has led to increased taxon sampling in databases adding further support for the validity of HGT events. Phylogenetic analysis of four oomycetes genomes (*Hy. arabidopsidis*, *Ph. infestans*, *Ph. ramorum* and *Ph. sojae*) using a database of 795 (173 eukaryotic and 622 prokaryotic) genomes identified 34 gene transfers between fungi and oomycetes (Richards et al., 2011). Interestingly, 62% to 76% of genes identified as originating via HGT from fungi in the four analyzed species possess a predicted secretion signal, representing between 2.7% and 7.6% of the total predicted secretomes of these species. Many of the identified genes have functions associated with the breakdown of plant cell walls and the uptake of nitrogen, nucleic acids, phosphate and sugars from the environment (Richards et al., 2011). This adds further support to the hypothesis that HGT events from fungi have played a part in the convergent evolution of osmotrophy and filamentous growth between fungi and oomycetes. Genome analysis of *Sa. parasitica* led to the identification of five gene families (four of which are secreted) that appear to have been gained by *Sa. parasitica* via HGT from bacteria or animals (Jiang et al., 2013). An additional six HGT events were reported from the genome sequences of *Ac. hypogyna* and *Th. clavata*, all of which are predicted to be secreted and involved in pathogenicity or carbohydrate metabolism (Misner et al., 2015).

Reanalysis of 48 HGT gene families based on the 23 oomycetes genomes that were available in 2015 highlighted several important findings (Savory et al., 2015). For example, 33 (69%) of the 48 HGT families are predicted to be secreted and 40 (83%) of the 48 HGT families appear to have a fungal origin. Only seven cases of HGT could be mapped back to the ancestor of the four crown oomycete orders, suggesting that HGT played a limited role in early oomycete evolution. HGT appears to have had a greater impact on plant pathogenic oomycetes, with 33 HGT events identified within the *Phytophthora*, *Hy. arabidopsidis* and *Pythium* clade, compared to only five in the branch

leading to the Saprolegniales order (Savory et al., 2015). Interestingly, many of the HGT derived genes have not only become fixed in the recipient genomes but have been duplicated, sometimes multiple times. For example, *Pythium* and *Phytopythium* species have a mean of 2.1 copies of each HGT derived gene, whereas *Phytophthora* species have a mean of 4.4 (Savory et al., 2015). Acquisition of genes via HGT can potentially change the phenotype of the recipient and also provide genetic material that has the potential to evolve novel or expanded functions. Detailed functional analyses of transporter proteins that were transferred to oomycetes before the divergence of Peronosporales and Saprolegniales revealed an HGT derived paralogue belonging to *Py. aphanidermatum* that has evolved an expanded substrate range enabling it to uptake not only dicarboxylic acid (the ancestral function) but also tricarboxylic acid (Savory et al., 2018).

Analysis of EST sequences from *Py. oligandrum* identified a homolog of a type 2 NLP from bacteria (Horner et al., 2012). Type 2 NLPs were previously thought to be absent in oomycetes and only found in fungi and bacteria. A follow-up analysis of effector proteins in 37 oomycete genomes using network and phylogenetic methods identified type 2 NLPs in three oomycete species - *Py. oligandrum*, *Pilasporangium apinafurcum* and *Pp. vexans* (McGowan and Fitzpatrick, 2017). Phylogenetic analysis suggested that the genes were likely gained via HGT from a Proteobacterial source and later duplicated with 2 copies in *Pp. vexans*, 6 copies in *Pi. apinafurcum* and 17 copies present in *Py. oligandrum*. An additional five instances of HGT from bacteria to oomycetes were reported in another study focusing on 14 plant pathogenic oomycete genomes including a putative secreted protein, a class II fumarase, an oxidoreductase, an alcohol dehydrogenase, and a hydrolase (McCarthy and Fitzpatrick, 2016).

Based on the genome analyses conducted to date, it is clear that HGT has played a significant role in oomycete genome evolution. In particular, HGT has had a major

impact on the secretomes of plant pathogenic oomycetes such as *Phytophthora*. Furthermore, convergent evolution between fungi and oomycetes was likely driven, in part, by HGT. As more genome sequences become available it will be possible to more accurately place the timing of putative HGT events, e.g. in an oomycete ancestor or specific to particular oomycete lineages, and also to rule out possible effects of poor taxon sampling.

6. Genome Mining for Oomycete Effectors

Oomycete species are notorious for secreting large arsenals of effector proteins that perform a wide array of functions during infection. Genome sequencing of oomycete species has facilitated genome mining to identify effector proteins which has led to many studies comparing effector repertoires between species and evolutionary analyses of effector families. The first step in identifying effector proteins is usually predicting which proteins are secreted using tools such as SignalP (Almagro Armenteros et al., 2019) to identify putative signal peptides. Transmembrane domains are then annotated using TMHMM (Krogh et al., 2001). Proteins that contain predicted transmembrane helices downstream of the signal peptide are usually discarded. It is important to note, that unconventionally secreted proteins (i.e. secreted proteins that lack a signal peptide and are signaled for secretion by some other mechanism) will be overlooked using this approach. Additionally, false positives may be reported. Therefore, the gold standard method for identifying secreted proteins is coupling bioinformatics methods with experimental techniques such as mass spectrometry (Meijer et al., 2014) or the yeast secretion trap system (Lee et al., 2006). ApoplastP is a machine learning classifier that can be used to predict if effector proteins localize to the plant apoplast (Sperschneider et al., 2018b). Many effector families contain conserved domains that can be annotated using hidden Markov model (HMM) searches against databases such as InterPro (Finn et al., 2017) or Pfam (Finn et al., 2016), for example, elicitins (PF00964) and NLPs (PF05630). Other effector families may have conserved positionally constrained motifs that can be identified using string searches or regular expression searches. For example, the RxLR family contains an N-terminus “RxLR” motif (where “x” means any amino acid) downstream of the signal peptide cleavage site. Similarly the CRN family contains a conserved N-terminus “LxLFLAK” motif which is usually followed downstream by an

"HVLVxxP" motif (Haas et al., 2009). More robust searches can be carried out to identify effectors with divergent motifs by building HMMs based on the results from regular expression searches. Novel effector proteins have also been identified due to being located in gene-sparse, repeat-rich regions of the genome with close proximity to transposable elements (Raffaele et al., 2010). Additionally, many effector proteins are found in clusters of tandemly duplicated genes (McGowan et al., 2019). Effector-like proteins can also be identified by performing sequence similarity searches against databases of known effectors, for example, using BLAST (Altschul et al., 1997) to identify homologs of experimentally verified effectors in the Pathogen-Host Interaction database (PHI-base) (Urban et al., 2017). The physio-chemical properties of a protein can also be used to identify effectors, for example, many effectors are small (low molecular weight), and cysteine-rich (Sperschneider et al., 2018b, 2015). The cysteine residues might be involved in disulfide bridges, increasing the stability and lifespan of effector proteins in the plant apoplast which contains a large number of degradative proteases (Kamoun, 2006). Genome-wide cataloging of effector proteins has revealed differences in effector repertoires between species with different hosts, some of which are discussed below.

6.1. Apoplastic Effectors

Apoplastic effectors are secreted by the pathogen and exert their pathogenic activity outside of the host cell (Wawra et al., 2012). Oomycete genomes encode large arsenals of secreted degradative enzymes that breakdown host cell components facilitating hyphal penetration of host cells. These degradative enzymes include CAZymes which modify and breakdown carbohydrates. In addition to allowing entry into host cells, the breakdown of host carbohydrates also makes nutrients available for the pathogen to grow.

Phytopathogenic oomycetes encode large arsenals of secreted CAZymes to facilitate the breakdown of plant cell wall components, including cellulases to metabolize cellulose, cutinases to degrade cutin, pectinases to degrade pectin, endoglucanases and β-glucosidases to degrade xyloglucan, and α-glucosidases, α-amylases, α-glucoamylases, and starch-binding modules to metabolize starch (Zerillo et al., 2013). As cellulose is also one of the main components of oomycete cell walls, it is thought that most oomycete cellulases are involved in oomycete cell wall metabolism and only few are involved in the breakdown of plant cell walls. However, they are still important for infection as a *Ph. infestans* cellulose synthase was shown to be involved in the formation of appressoria and required for infection (Grenville-Briggs et al., 2008). Furthermore, proteins with cellulose-binding domains are enriched in the secretome of many *Phytophthora* species suggesting a role in infection (McGowan and Fitzpatrick, 2017). Chitinases and chitin-binding proteins are also expanded in some oomycete genomes including *Ap. astaci*, *Py. oligandrum*, *Sa. diclina* and *Sa. parasitica* (Jiang et al., 2013; McGowan and Fitzpatrick, 2017). As oomycete cell walls contain very little to no chitin, this suggests that oomycete chitin modifying enzymes are involved in the breakdown of exogenous chitin. This is particularly significant for the expansion of chitinases in species such as *Py. oligandrum* and *Ap. astaci,* which parasitize fungi and crayfish respectively. Other types of degradative enzymes include proteases which are also thought to be involved with host cell degradation (Haas et al., 2009). For example, proteases are highly expanded in the animal pathogens *Sa. parasitica* and *Py. guiyangense* and may facilitate cuticle penetration (Jiang et al., 2013; Shen et al., 2019). Interestingly, oomycetes also secrete protease inhibitors to counteract defense proteases produced by the host in response to infection (Tian et al., 2006), highlighting the co-evolutionary arms race between pathogens and their hosts.

Many secreted oomycete proteins or oomycete cell wall/membrane constituents are recognized by the host and trigger a host defense response upon recognition (Oome et al., 2014). Such molecules are referred to as microbe-associated molecular patterns (MAMPs). Well-characterized oomycete MAMPs include elicitins, transglutaminases, and cellulose-binding elicitor lectins (CBELs). Elicitins are known to bind lipids and sequester sterols from host plants, allowing some pathogenic oomycetes to overcome their inability to synthesize their own sterols (Kamoun, 2006). Transglutaminases facilitates the cross-linking of protein-bound glutamine and lysine residues, strengthening structures such as cell walls and conferring resistance to proteolysis (Raaymakers and Van den Ackerveken, 2016). CBEL proteins are known to activate expressions of host defense genes and elicit necrosis and are important for cell wall structure (Raaymakers and Van den Ackerveken, 2016).

Other well studied apoplastic effectors include NLPs and the PcFs. NLPs have a broad taxonomic distribution across the tree of life with homologs found in bacteria, fungi, and oomycetes (Seidl and Van den Ackerveken, 2019). NLPs are known to induce ethylene accumulation leading to rapid tissue necrosis and are associated with the switch to necrotrophy in *Phytophthora* species. Some NLPs have been reported to be noncytotoxic but instead, act as MAMPs (Oome et al., 2014). NLPs are highly expanded in some oomycete species, in particular, some *Phytophthora* species have up to 80 copies (McGowan and Fitzpatrick, 2017). NLPs are completely absent in some species such as *Albugo*, *Aphanomyces* and *Saprolegnia* (Kemen et al., 2011; Links et al., 2011; McGowan and Fitzpatrick, 2017). In contrast to NLPs, PcF proteins are present in lower copy numbers, with 16 identified in *Ph. infestans*, 8 in *Ph. sojae* and only one in *Ph. ramorum* (Haas et al., 2009).

Pcf proteins are small, cysteine-rich proteins that are known to induce plant cell necrosis (Orsomando et al., 2001). Pcf proteins appear to be unique to Peronosporales species based on the currently available genome sequences.

6.2. Cytoplasmic Effectors

Oomycete cytoplasmic effectors differ from apoplastic effectors in that they are secreted and translocated into host cells. Three main classes of oomycete cytoplasmic effectors have been identified – RxLRs, CRNs and CHXC effectors. RxLR effectors are so named because they contain a highly conserved "RxLR" motif in their N-terminal domain. This is often followed downstream by an "EER" motif in many RxLRs (Whisson et al., 2007). RxLR C-terminal domains are typically highly divergent although many contain conserved structural folds caused by the presence of one or more "WY" domains (Win et al., 2012). RxLRs are highly expanded in the Peronosporales order but were not detected at all in *Aphanomyces*, *Pythium* or *Saprolegniales* species (Adhikari et al., 2013; Gaulin et al., 2018; Jiang et al., 2013; Lévesque et al., 2010). However small numbers of RxLR and RxLR-like effectors have been detected in the two *Albugo* genomes (Kemen et al., 2011; Links et al., 2011), making their origin unclear. RxLR effectors are particularly abundant in the genomes of *Phytophthora* species, for example, *Ph. infestans* is predicted to encode 563 RxLRs (Haas et al., 2009), while *Ph. palmivora* and *Ph. megakarya* are predicted to encode 991 and 1,181 RxLRs, respectively (Ali et al., 2017). Many *Phytophthora* RxLRs are expressed in early infection stages and play a role in modulating and suppressing the host immune response (J. Yin et al., 2017). However, the function and molecular mechanisms for most RxLRs are unknown. While the function of most RxLRs is unknown, the subcellular location of a large number of *Ph. infestans* RxLRs was recently determined. *Ph. infestans* RxLRs were shown to localize to diverse

subcellular locations when they enter plant cells, with most localizing to the host cytoplasm, nucleus or plasma membrane (Wang et al., 2019). Furthermore, most of the RxLRs under consideration were shown to enhance colonization. The mechanism of translocation into host cells is unclear but some RxLRs have been shown to enter host cells independent of additional pathogen encoded machinery (Dou et al., 2008). Recent work has shown that the haustorium is involved in the delivery of apoplastic and cytoplasmic effectors both by conventional and nonconventional secretory pathways (Wang et al., 2018, 2017).

The origin of oomycete RxLR effectors is unclear. Interestingly, comparative genomic and syntenic analysis using the Oomycete Gene Order Browser (OGOB) (McGowan et al., 2019) reveals a conserved *Ph. infestans* RxLR effector (SFI4 or PITG_09585) that appears to have been present in the ancestor of the four crown oomycete orders. SFI4 localizes to the host nucleus-cytoplasm and reduces flg22-induced pFRK1-Luc activation in tomato protoplasts, enhancing *Ph. infestans* colonization (Zheng et al., 2014). SFI4 is also highly expressed during infection of potato (J. Yin et al., 2017). In OGOB, SFI4 is in a pillar with orthologs from 17 other species (**Figure 4A**). An ortholog was not identified in *Al. candida* or *Al. laibachii*, suggesting that it may have been lost by these species. As orthologs are found in Peronosporales, Pythiales and Saprolegniales orders, it suggests that this is an ancient RxLR gene locus that was present in an ancestral oomycete species. All protein sequences in this pillar show a high degree of sequence similarity (**Figure 4B**) and are reciprocal best hits to each other in a BLASTp search. Furthermore, all orthologs in this pillar show evidence of microsyntenic conservation (**Figure 4C**), adding further evidence that they are legitimate orthologs. SignalP analysis predicts that 17 out of the 18 proteins in this pillar contain a signal peptide. Multiple sequence alignment reveals the presence of a conserved RxLR motif in

all orthologs in this pillar (**Figure 4B**). All orthologs contain a KDEL or KDEL-like motif at the C-terminus, a motif usually involved in endoplasmic reticulum (ER) retention (i.e. non-secretion), however it is possible that it is masked. A homolog of PITG_09585 was previously detected in *Py. oligandrum* (Horner et al., 2012) but was not considered a bona fide RxLR effector due to the presence of the KDEL C-terminal motif. In addition to *Ph. infestans* SFI4, a number of the orthologs have evidence of transcription during infection. The ortholog in *Hy. arabidopsidis* (HYAP_13904 or RxLR5) was identified during EST sequencing of infected *Arabidopsis* tissue (Cabral et al., 2011). The *Ph. capsici* ortholog (PHYCA_101423) has evidence of expression during infection of *Nicotiana benthamiana* in RNA-Seq experiments (Chen et al., 2013) and during infection of tomato in microarray experiments (Jupe et al., 2013). Additionally, the *Pl. halstedii* ortholog (PHALS_05912) was identified in an RNA-Seq analysis of infected sunflower (Sharma et al., 2015). Interestingly, comparative RNA-seq analysis of *Ph. infestans* and *Py. ultimum* has shown that the two orthologs (PITG_09585 and PYU1_G013121) have similar expression patterns during potato tuber colonization (Ah-Fong et al., 2017b). Whether the conserved orthologs in *Pythium*, *Aphanomyces* and *Saprolegnia* function as effectors warrants further investigation to understand their importance. Together, these data suggest that the RxLR class of effectors is more ancient than previously thought and may have evolved in an ancestral oomycete species prior to their proliferation in Peronosporales species.

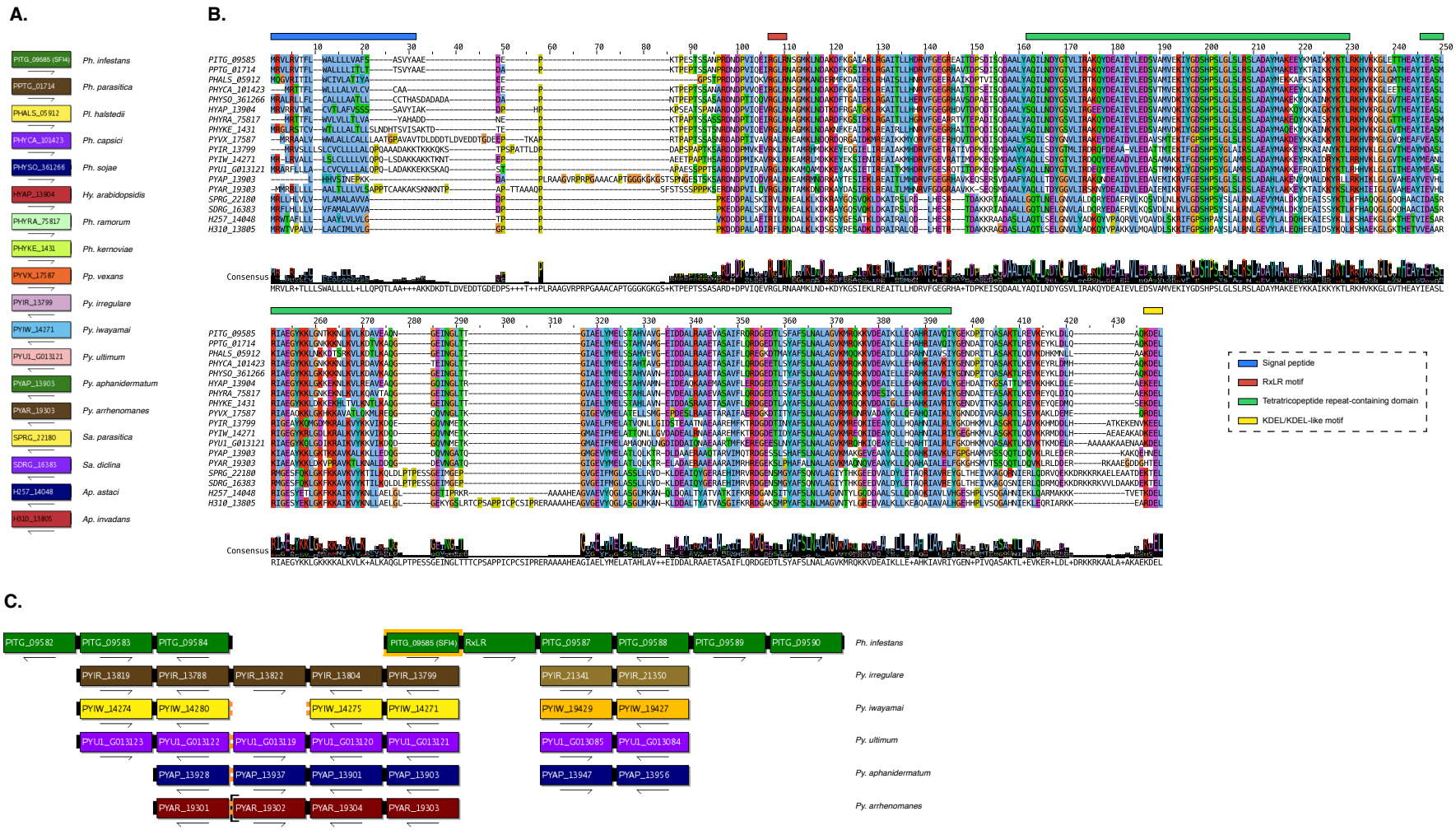


Figure 4. An ancient conserved oomycete RxLR locus. (**A**) *Ph. infestans* SFI4 (PITG_09585) has orthologs in other *Phytophthora*, *Hyaloperonospora*, *Phytopythium*, *Pythium*, *Saprolegnia* and *Aphanomyces* species. (**B**) Multiple sequence alignment of PITG_09585 and its orthologs show extensive sequence similarity and the presence of a "RxLR" motif. (**C**) Genomic context of SFI4 and its orthologs visualized in OGOB showing that it is microsyntenically conserved. For illustrative purposes, only PITG_09585 and its *Pythium* orthologs are shown. Full microsynteny of other orthologs can be examined on OGOB.

CRNs are named after their crinkling and necrosis-inducing activity. CRN proteins have highly conserved N-terminal domains containing a signal peptide and a "LxLFLAK" motif that mediates translocation into host cells (Schornack et al., 2010). The end of the N-terminus is marked by a highly conserved "HVLVxxP" motif. CRNs are modular proteins that have highly diverse C-terminal domains (Haas et al., 2009). In contrast to RxLRs, CRNs are thought to be a more ancient class of effectors as they have been detected across a wide range of oomycetes including *Albugo*, *Aphanomyces*, downy mildews, *Phytophthora* and *Pythium* (Gaulin et al., 2018; Kemen et al., 2011; Links et al., 2011; Schornack et al., 2010). CRNs were not detected in the genome of *Sa. parasitica* (Jiang et al., 2013). CRNs are also highly expanded in the genomes of some *Phytophthora* species, with 196 CRNs annotated in *Ph. infestans* and 100 in *Ph. sojae* (Haas et al., 2009). CRNs include some of the most highly expressed effector genes before and during infection (Haas et al., 2009). Several CRNs have been reported to target the host nucleus (Schornack et al., 2010). Interestingly, highly divergent CRNs were identified in the genome of the insect pathogenic *Py. guiyangense* and some were shown to be toxic in insect cells (Shen et al., 2019). Similar to RxLRs, CRNs have undergone rapid evolution, expansion, and diversification. For example, the genomes of *Ph. megakarya* and *Ph. palmivora* were where shown to encode 152 and 137 CRNs respectively, but only 30 of these are core-orthologues shared between these two closely related pathogens (Ali et al., 2017). Many annotated CRNs lack conventional N-terminal signal peptides. For example, only 58% of annotated *Ph. cactorum* CRNs possess a canonical signal peptide and only 60 out of the predicted 139 (43%) CRN proteins in *Ph. plurivora* have a signal peptide (Armitage et al., 2018; Vetukuri et al., 2018b). A large number of CRN and RxLR pseudogenes were annotated in the genome of *Ph. infestans* suggesting the prevalence of

rapid gene birth and death, possibly driven by the co-evolutionary arms race between pathogen and host, and an effort to evade host recognition (Haas et al., 2009).

Sequencing of the *Al. laibachii* genome led to the identification of a novel class of CHXC effectors (Kemen et al., 2011). Similar to the other characterized cytoplasmic effectors, the *Albugo* CHXC effectors contain a conserved "CHxC" motif within 50 amino acids of the signal peptide cleavage site that was shown to facilitate effector translocation into host cells (Kemen et al., 2011). A total of 29 CHXC effector candidates were annotated in the *Al. laibachii* genome. The "CHxC" motif was also significantly enriched in the secretome of *Al. candida* where a total of 40 CHXC effector candidates were identified, none of which had homologs in *Phytophthora* or *Hyaloperonospora* (Links et al., 2011).

7. Oomycete OMICS Studies

Genome sequencing of oomycete species has also facilitated other high throughput "OMICs" analyses such as proteomics and transcriptomics. Such analyses have revealed insights into oomycete infection mechanisms, secretome content, cell structure, metabolism, and nutritional strategies. Additionally, proteomic and transcriptomic data have been invaluable for improving the quality of oomycete genomes via the annotation and validation of predicted genes.

7.1. Oomycete Proteomics Studies

The majority of proteomics studies to date have focused on *Phytophthora* species. Comparative proteomic analysis between mycelium and germinating cysts of *Ph. ramorum* and *Ph. sojae* was carried out to identify proteins involved in vegetative and infection stages (Savidor et al., 2008). In total, 3,897 proteins were identified from *Ph. ramorum* and 2,970 from *Ph. sojae*. An average of 14% of identified proteins contained a signal peptide, 42 of which from *Ph. sojae* and 46 from *Ph. ramorum* contained an RxLR motif within the first 30 – 60 residues suggesting they belong to the RxLR effector family. 686 proteins were identified as being differentially expressed between germinating cysts and mycelium in *Ph. ramorum*. This number was 513 for *Ph. sojae* (Savidor et al., 2008). Proteins upregulated in germinating cysts were involved in functions related to cytoskeleton, protein synthesis and the transport and metabolism of lipids (including many proteins involved in the β-oxidation pathway). Proteins upregulated in mycelium were involved in functions related to the transport and metabolism of amino acids, carbohydrates, and other small molecules. These findings revealed insights into infection, growth and nutritional strategies of *Phytophthora*

species. They suggested that germinating cysts catabolize lipids (via the β-oxidation pathway), generating the energy required for protein synthesis for germ tube formation used in infection initiation. Upon entry into the host, *Phytophthora* switches to vegetative mycelial growth, generating energy via glycolysis to synthesize amino acids and other molecules required for survival inside host tissues (Savidor et al., 2008). Comparative proteomic analysis of *Ph. pisi* and *Ph. sojae* revealed proteomic differences that may contribute to host specificity between these two closely related pathogens (Hosseini et al., 2015). A total of 2,775 proteins from *Ph. pisi* and 2,891 proteins from *Ph. sojae* were identified. Similar findings were reported with regards to nutritional strategy differences between hyphae and germinating cysts (Hosseini et al., 2015; Savidor et al., 2008). Proteomic profiling of *Ph. capsici* identified 599 proteins with significantly altered expression in response to the fungicide SYP-14288 (Cai et al., 2019). The majority of affected proteins had functions related to carbohydrate metabolism, energy metabolism, metabolism of other amino acids, amino acid metabolism, transport, and catabolism.

The *Ph. infestans* genome encodes 372 genes predicted to be protein kinases, comprising 2% of the total gene set, suggesting that phosphorylation plays an important role in the lifecycle of *Ph. infestans* (Resjö et al., 2014). A large scale phosphoproteomics analysis of six life stages of *Ph. infestans* (hyphae, sporangia, zoospores, cysts, germinated cysts, and appressoria) was performed to identify phosphorylated peptides (Resjö et al., 2014). A total of 2,922 phosphopeptides were identified. Among the phosphorylated peptides were 35 CRN-derived phosphopeptides. Many CRNs were phosphorylated at multiple sites, across several life stages indicating a potential role beyond inducing necrosis (Resjö et al., 2014). Quantitative proteomic analysis of the same six life stages identified over 10,000 peptides corresponding to 2,061 proteins (Resjö et al., 2017). In particular, 59 proteins were significantly more abundant in

appressoria and germinating cysts, i.e. pre-infection stages of the life cycle. The majority of these proteins are involved in transport, energy and amino acid metabolism, redox maintenance and signaling (Resjö et al., 2017). Interestingly, this set of proteins with increased abundance in pre-infection stages also included many proteins involved in cell wall synthesis, maintenance, and adhesion. Transient silencing of three putative cell wall proteins resulted in abnormal phenotypes and an impaired ability to infect potato leaves, suggesting an important role in the infection process (Resjö et al., 2017). *Ph. infestans* cell-wall associated proteins from three life stages (germinating cysts with appressoria, sporulating mycelium, and non-sporulating mycelium) were also identified using a mass spectrometry approach, resulting in the identification of 31 proteins (Grenville-Briggs et al., 2010). All of the cell-wall associated proteins of germinating cysts with appressoria and some from mycelia were classified as effectors or MAMPs which could potentially trigger a host immune response.

Mass spectrometric analysis of *Ph. infestans* grown in seven different growth media was performed to characterize and validate proteins belonging to the secretome and extracellular proteome of *Ph. infestans* (Meijer et al., 2014). In total, 283 proteins were identified in the extracellular media confirming they are bona fide extracellular proteins, of which 227 contained a signal peptide. Among the identified proteins were 20 RxLRs, 13 CRNs, 11 elicitins, and 5 necrosis-inducing proteins (Meijer et al., 2014). Thirty one proteins were found to contain a signal peptide and a single transmembrane domain (most of which were located in the C-terminus), suggesting that they are membrane proteins that are proteolytically shed via sheddases, explaining their presence in the extracellular medium (Meijer et al., 2014). The proteomics approach facilitated the reannotation and correction of many gene models, the N-terminus of many were extended revealing signal peptides that weren't previously annotated. A similar approach was

carried out to characterize the secretome of *Ph. plurivora* when stimulated with root exudate from *Fagus sylvatica* resulting in the identification of 272 proteins, 60% of which contained signal peptides (Severino et al., 2014). The secretome was enriched with functions related to enzymatic activity such as hydrolases, oxidoreductases, and transferases. 21 proteins were found to be differentially abundant following treatment with root exudate (Severino et al., 2014).

While the majority of oomycete proteomic analyses have focused on *Phytophthora* species, several analyses have been performed for other taxa. For example, comparative proteomic analysis of *Py. insidiosum* revealed 212 proteins and 124 proteins that were found in higher or lower abundance, respectively, when grown at 37℃ compared to 25℃ (Rujirawat et al., 2018). The majority of highly upregulated proteins were predicted to be involved in transport and metabolism of amino acids, carbohydrates, lipids and nucleotides, post-translational modification, protein turnover, and chaperones. These proteins may facilitate the high-temperature tolerance of *Py. insidiosum* during infection of mammals (Rujirawat et al., 2018). Quantitative proteomic analysis of four life stages (mycelium, primary cysts, secondary cysts, and germinated cysts) of *Sa. parasitica* grown *in vitro* identified a total of 2,423 unique proteins across the four life stages (Srivastava et al., 2018). Compared to the three cyst stages, 133 proteins were found in increased abundance in the mycelium with functions related to amino acid and carbohydrate metabolism enriched. Conversely, 110 proteins were found in increased abundance in the three cyst stages compared to the mycelium with an enrichment of functions related to signal transduction, translation, ribosomal structure and biogenesis (Srivastava et al., 2018).

7.2. Oomycete Transcriptomics Studies

Transcriptomic analyses (i.e. RNA-seq) are now routinely performed alongside genome sequencing as the sequenced transcripts can be incorporated during gene calling steps and to validate existing gene models. Many transcriptomics studies have been carried out for oomycete species to identify transcriptional remodeling during infection or after exposure to chemical treatment. Comparative transcriptomics analyses between different species have also been performed to identify differences or similarities between species.

RNA-seq analysis of various life stages in two *Ph. infestans* isolates led to the mapping and validation of approximately 16,000 genes (90%) (Ah-Fong et al., 2017a). Large-scale transcriptome remodeling was detected at each life stage with the biggest differences observed during the transition from hyphae to sporangia where more than 4,200 genes were upregulated, more than 1,350 of which were detected at greater than 100-fold increased abundance. Genes encoding calcium-binding proteins, flagellar proteins, ion channels, and signaling proteins were upregulated in sporangia. Most metabolic pathways were downregulated after sporulation but later reactivated upon the germination of cysts (Ah-Fong et al., 2017a). Similarly, RNA-seq analysis of *Ph. capsici* identified a large number of stage-specific genes including effector families and metabolic pathways, revealing proteins that are important during pre-infection (Chen et al., 2013). Transcriptome analysis of *Ph. litchii* mycelia, sporangia, and zoospores identified 19,267 unigenes, of which 490 were predicted to be pathogenicity-related proteins, including 128 RxLRs, 35 elicitins, 29 NLPs and 22 CRNs (Sun et al., 2017). The *Ph. litchii* unigenes were clustered into 9,685 orthologous groups, 105 gene families did not have orthologs in the other four oomycete species under consideration in the study, suggesting that they may be novel genes.

Comparative transcriptomic analysis was performed to identify transcriptional differences between *Ph. infestans* and *Py. ultimum* during infection of potato tubers (Ah-Fong et al., 2017b). This is an interesting pair of species for comparison, as they both infect potato and have different lifestyles. *Ph. infestans* is a hemibiotroph that begins infection with a biotrophic phase whereby it requires a living host for colonization. Later in the infection process, *Ph. infestans* switches to a highly destructive necrotrophic phase. In comparison, *Py. ultimum* is a necrotroph that immediately induces host cell death upon infection. *Ph. infestans* was found to have a higher number of stage-specific genes. In particular, a much larger proportion of *Ph. infestans* genes were shown to have ≥ 4-fold expression change between early and late infection (45% in *Ph. infestans* compared to only 9% in *Py. ultimum*), this could be attributed to the switch from biotrophy to necrotrophy that *Ph. infestans* undergoes during late infection. The evolution of necrotrophy and hemibiotrophy could be ascribed to species-specific genes, expansion and contraction of orthologous gene families, and differences between the timing and level of expression of orthologs (Ah-Fong et al., 2017b). RNA-seq analysis of the saprotroph *Salisapilia sapeloensis* was performed in comparison to eight other plant pathogenic oomycetes with different lifestyles (biotrophic, hemibiotrophic and necrotrophic) (Vries et al., 2019). Results show that the saprotrophic and pathogenic oomycetes possess similar repertoires for colonization but exhibit different expression patterns. Furthermore, *Salisapilia sapeloensis* was shown to possess a smaller effector arsenal. These findings indicate that the evolution of oomycete lifestyles is influenced not only by gene content but also shifts in gene regulatory networks (Vries et al., 2019).

Transcriptomic methods have also proven useful in understanding how oomycetes respond to chemical treatments and biocontrol agents. RNA-seq of *Ph. infestans* in response to the phenazine-1-carboxylic acid (PCA) producing *Pseudomonas fluorescens*

showed that more than 200 genes were significantly differentially expressed with more than a 3-fold change following PCA treatment (Roquigny et al., 2018). Differentially expressed genes were predicted to be involved in many processes, including oxidoreduction activity, phosphorylation mechanisms and transmembrane transport activity. Transcriptional changes were also identified in effector genes even in the absence of a host plant. These findings suggest that PCA exposure results in growth repression which leads to transcriptomic changes in *Ph. infestans* (Roquigny et al., 2018). RNA-seq analysis of *Ph. parasitica* in response to the fungicide dimethomorph identified significant differential expression of 832 genes, of which 365 were up-regulated and 467 down-regulated, including many genes associated with the cell membrane and wall synthesis (Hao et al., 2019). RNA-seq of *Sa. parasitica* following treatment with copper sulfate showed than 310 genes were upregulated and 556 genes were downregulated (Hu et al., 2016). Functional annotation of differentially expressed genes indicates that copper sulfate inhibits the growth of *Sa. parasitica* by affecting multiple biological processes including energy biogenesis, metabolism and protein synthesis (Hu et al., 2016).

Tissue-specific and host induced genes from *Sa. parasitica* were identified via RNA-seq of multiple life stages as well as infected fish cell lines (Jiang et al., 2013). In particular, the results suggested that kinases play an important role during the infection process, as of the large kinome (543 kinases) of *Sa. parasitica,* 10% were upregulated greater than 4-fold in germinating cysts compared to mycelia. Similarly, the large arsenal of proteases (270) was shown to be expressed in waves at different stages during infection, indicating an important role in the infection process (Jiang et al., 2013). RNA-seq of *Ap. euteiches* infecting *Medicago truncatula* showed that adaption to plant hosts is correlated with the expression of specialized secretomes (Gaulin et al., 2018). Genes involved in carbohydrate metabolism and proteolysis were upregulated during infection.

Furthermore, three times as many secreted protein-encoding genes were upregulated during infection, the majority of which encode proteases, glycoside hydrolases and polysaccharide lyases (Gaulin et al., 2018).

Dual transcriptomic analyses can be employed to characterize the infection process of a pathogen and the response of the host to that pathogen. Dual transcriptomic time-course analysis of *Ph. palmivora* infecting *Nicotiana benthamiana* identified 2,441 differentially expressed *Ph. palmivora* genes and 6,950 differentially expressed *Ni. benthamiana* genes in response to the pathogen (Evangelisti et al., 2017). The pathogen was shown to undergo sharper transcriptional remodeling in comparison to the host plant which was characterized by steady up or downregulation. A large number of putative *Ph. palmivora* effectors were identified including 143 cell wall degrading enzymes, 140 RxLRs, 59 proteases, 42 elicitins, 28 small cysteine-rich proteins, 24 NLPs and 15 CRNs (Evangelisti et al., 2017). Genes linked to abiotic stress, biosynthesis, defense, hormone metabolism, protein modification, and transcriptional regulation were upregulated in *Ni. benthamiana*. Genes associated with cell division, cellulose biosynthesis and photosynthesis were downregulated during infection. These results suggest that *Ni. benthamiana* responds to *Ph. palmivora* infection by transcriptional remodeling and post-translational reprogramming which activates defense and stress responses (Evangelisti et al., 2017).

8. Tools for Oomycete Genomics

Compared to other taxonomic groups, there are few dedicated tools available to study oomycete genomes. However, there have been several recent developments, including EumicrobeDB (http://www.eumicrobedb.org; last accessed December 1, 2019), a database that hosts 26 oomycete genome sequences and a large number of bioinformatics tools (Panda et al., 2018). FungiDB (https://fungidb.org; last accessed December 1, 2019) also hosts genomic data for 21 oomycete species (Basenko et al., 2018). The Oomycete Gene Order Browser (OGOB) (https://ogob.ie; last accessed December 1, 2019) was recently published and hosts genomic data for 20 oomycete species, including syntenically curated orthology inferences and bioinformatics tools that facilitate comparative genomic analyses and a browser for examining syntenic conservation of protein-coding genes between multiple species (McGowan et al., 2019). Additional databases dedicated to oomycetes include *Phytophthora*-ID (http://phytophthora-id.org; last accessed December 1, 2019) (Grünwald et al., 2011) and OomyceteDB (http://oomycetedb.cgrb.oregonstate.edu; last accessed December 1, 2019), which are curated databases that host genetic marker sequences for identifying oomycete species. There are also few dedicated tools to functionally annotate oomycete genes. For example, EffectorP is a tool that can be used to predict effector proteins from fungal secretomes using a machine learning classifier (Sperschneider et al., 2018a). With the growing number of experimentally verified oomycete effectors, similar approaches could be used to build machine learning classifiers for oomycete effectors.

Identifying BUSCOs is one of the most commonly used methods for assessing genome assembly completeness (Waterhouse et al., 2018). Currently, the most specialized BUSCO dataset available for oomycete genomes is the "Alveolata-Stramenopiles" dataset. An issue with this is that it only contains 234 BUSCO proteins

based on a highly diverse set of 24 species, only seven of which are oomycetes. Other species in this dataset include the Alveolates - *Plasmodium falciparum* and *Toxoplasma gondii*. Including distantly related lineages when identifying orthologs reduces the number of universally present single copy orthologs. This may result in an inaccurate assessment of genome assembly completeness. A dedicated dataset for the oomycetes, or better, for individual oomycete lineages/genera, would provide a more accurate representation of genome and gene space completeness. For example, analysis of the 20 oomycete genomes hosted on OGOB (McGowan et al., 2019) using OrthoFinder (Emms and Kelly, 2019) identifies 643 ubiquitously conserved single-copy orthologs. Furthermore, taxonomically restricting this dataset to the genomes of 8 more closely related Peronosporales species (*Hy. arabidopsidis, Ph. capsici, Ph. infestans, Ph. kernoviae, Ph. parasitica, Ph. ramorum, Ph. sojae, and Pl. halstedii*) identifies 2,231 ubiquitously conserved single copy orthologs. Additional highly conserved markers can also be considered when assessing genome completeness. For example, the Fungal Genome Mapping Pipeline (FGMP) was recently developed to assess genome completeness of fungal genome assemblies, taking into account not only ubiquitous single-copy gene families but also ubiquitous multi-copy gene families and highly conserved non-coding regions (Cissé and Stajich, 2019). Using the OrthoFinder analysis above, there are 53 gene families that are ubiquitously multi-copy (i.e. all genomes contain more than one copy) across the 20 oomycete genomes hosted on OGOB, and 83 across the 8 Peronosporales genomes. Including ubiquitously duplicated genes in genome assembly assessment can potentially identify genome misassemblies such as collapsed duplications or repetitive regions. Similar sophisticated methods could be developed to assess genome assembly quality and completeness of oomycete genomes.

9. Oomycetes in the Post-Genomic Era

The quality of the 65 genomes reviewed herein varies greatly. Oomycete genomes can be difficult to assemble due to their relatively large size, high proportions of repeat content and high levels of heterozygosity in certain species. Genome completeness of the 65 genome assemblies, assessed using BUSCO v3 (Waterhouse et al., 2018) with the Stramenopiles/Alveolata BUSCO dataset, ranges from 99.10% to 42.80% (mean = 89.32%; median = 94.00%) (**Figure 5**), showing that most oomycete genomes are highly complete in terms of expected gene content. The percentage of single-copy BUSCOs present ranges from 98.70% to 16.70% percent (mean = 80.83%; median = 90.6%) (**Figure 5**), indicating that several genomes have a large number of BUSCOs that are present in multiple copies. While this may represent legitimate multiple copies and duplication events, such is the case in the hybrid *Py. guiyangense* which contains duplicates for 81.6% of BUSCOs (**Figure 5**), it also likely represents haplotypes that weren't fully collapsed or misassemblies of repetitive sequences that cannot be assembled fully. Most oomycete genomes have very low fragmented BUSCO scores, ranging from 0% to 4.5% (mean = 1.3%; median=0.9%) (**Figure 5**), suggesting that gene fragmentation isn't a problem. Many of the currently available oomycete genome assemblies are highly fragmented, with sequences split across thousands or even tens of thousands of short contigs or scaffolds.

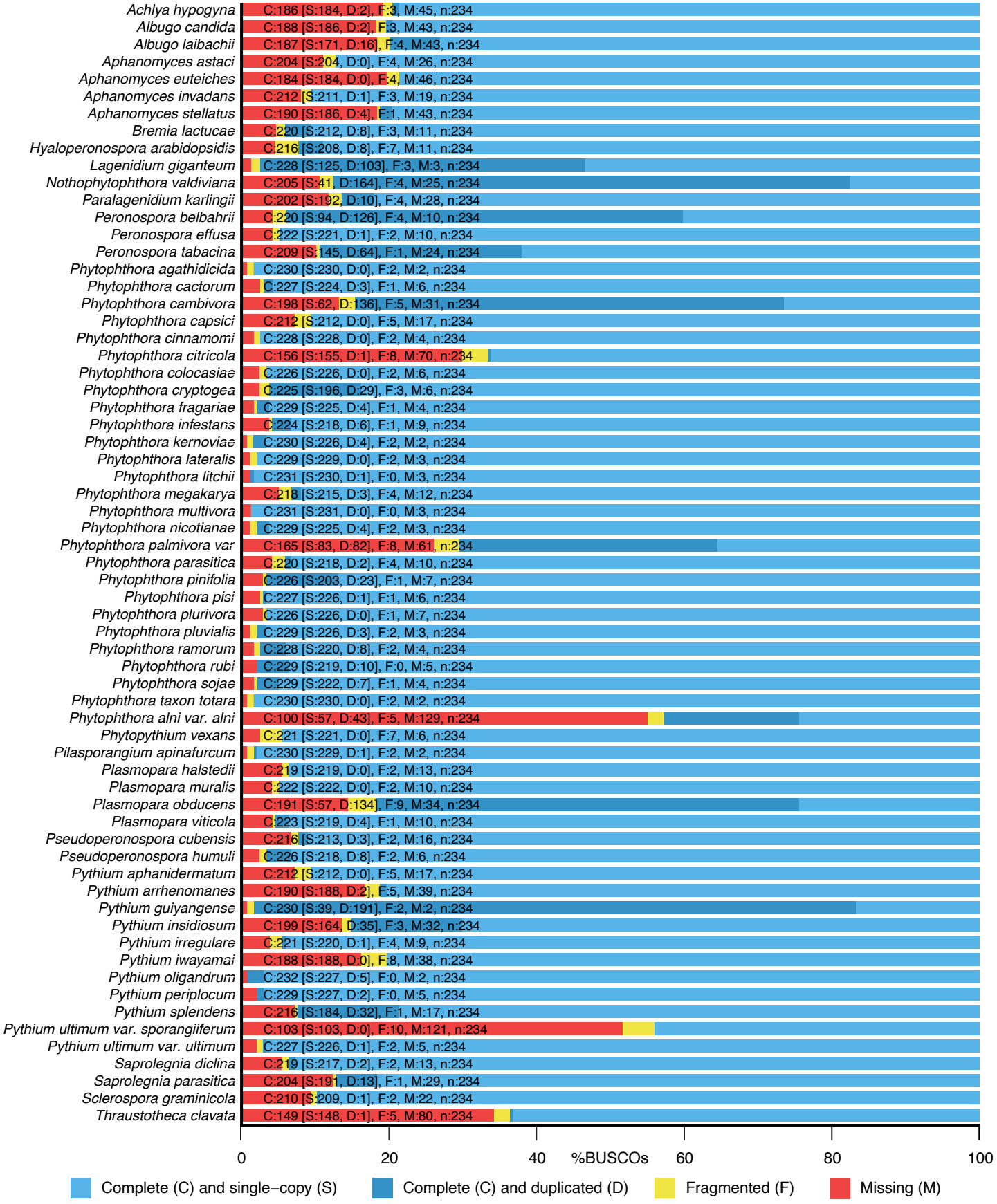


Figure 5. Distribution of BUSCO completeness scores across the 65 Oomycete genomes, assessed using the 234 BUSCOs from the "Alveolata-Stramenopiles" dataset.

The majority of oomycete genomes have been sequenced using short-read Illumina sequencing only. Illumina sequencing produces high yields of high-quality reads. However, the reads are short with a length range between 50 and 250 nucleotides. Short reads cannot span the length of repetitive regions of the genome as the repetitive regions are typically longer than the reads, resulting in highly fragmented genome assemblies with many short contigs. Long-read sequencing technologies such as PacBio and Oxford Nanopore Technologies (ONT) have the potential to generate higher-quality assemblies in terms of contiguity and completeness of repetitive regions. PacBio sequencing produces sequencing reads with average read lengths of over 10,000 bp and maximum read lengths of over 60,000 bp (Rhoads and Au, 2015). ONT can sequence full-length fragments of DNA, detecting bases as they pass through a nanopore. Long-reads from PacBio and ONT platforms are more likely to span repetitive regions leading to more contiguous oomycete genome assemblies that have longer and fewer contigs. Long-read sequencing also facilitates more comprehensive comparative genomic and syntenic analyses as larger structural variants, such as large inversions or translocations, can be detected. Long-read sequencing presents a trade-off as generated reads have a higher error rate compared to Illumina reads, and library construction and sequencing are more expensive with a higher cost per-base. Hybrid sequencing approaches can be undertaken in which long-reads or long-read assemblies are corrected and polished with high coverage, high-quality short-reads. This approach combines the benefits of long-read and short-read sequencing, producing high-quality, more complete and more contiguous assemblies.

Several oomycete genome sequencing projects have already taken advantage of long-read sequencing technologies. Sequencing of a highly virulent *Ph. ramorum* isolate (ND886) using PacBio long-read sequencing generated a haplotype-phased assembly that

is much more contagious than the reference (Pr102) assembly (302 contigs versus 2,576 contigs) (Malar C et al., 2019a). Additionally, the PacBio assembly contains a much higher proportion of repeat content than the reference assembly (48% versus 29%) showing the ability of long-read sequences to span and assemble repetitive sequences. The average read length was 10,570 bp and the longest read was 50,801 bp. The reference *Ph. ramorum* genome (Pr102) was also re-sequenced using the PacBio platform to a total of 80X coverage with an average read length of 14,000 bp, in combination with Illumina short-reads (Malar C et al., 2019b). The updated assembly is 5 Mb larger than the previous reference assembly (Tyler et al., 2006), is more contiguous (1,512 scaffolds versus 2,576 scaffolds), contains a higher proportion of repetitive content (54.4% versus 29.08%) and an additional 3,360 genes were reported (Malar C et al., 2019b). Resequencing of *Ph. capsici* using ONT MinION long-reads resulted in a more contiguous assembly (424 scaffolds) that was 95 Mb in length (~ 30 Mb larger than the reference assembly) (Cui et al., 2019). The increased assembly size corresponds to an increase in repeat content that was captured using long-read sequencing. A single MinION flow-cell produced ~10 Gb of data corresponding to 70X coverage. The N50 of the unassembled reads was 11,507 bp, the average read length was 7,114 bp and the longest read was 99,577 bp. Interestingly the authors noted that the improved assembly was generated in only nine days (Cui et al., 2019). The ONT MinION sequencing platform brings us closer to real-time field sequencing during pathogen outbreaks, allowing for rapid identification and tracking of pathogens as infection occurs.

Recent long-read sequencing efforts have also revealed novel insights into oomycete biology and genetics. Resequencing of *Ph. sojae* using ONT and PacBio long-read sequencing coupled with chromatin immunoprecipitation sequencing (ChIP-Seq) led to the identification of centromeres (Fang et al., 2020). The identified *Phytophthora*

centromeres were large, transposon rich and without nucleotide bias, showing divergence from other organisms in the SAR supergroup that have relatively short and simple centromeric sequences. Copia-like transposon (CoLT) elements that are highly enriched in the *Ph. sojae* centromeres were identified and shown to be conserved in *Ph. citricola* and *Br. lactucae*, presenting a novel feature that may be used to predict centromeres in other oomycete species (Fang et al., 2020). Telomeres were identified at single ends of 13 contigs in the long-read assembly. It is not yet accurately known how many chromosomes are present in *Ph. sojae*, due to difficulties in resolving karyotypes using pulse-field gel electrophoresis caused by large genome size and potentially similar chromosome sizes. Based on the identification of centromeres and telomeres in the updated *Ph. sojae* assembly, it is estimated that *Ph. sojae* has 12 – 14 chromosomes (Fang et al., 2020).

Resequencing multiple diverse strains of the same species followed by *de novo* genome assembly will allow for future pan-genomic analyses of oomycete species. Large numbers of isolates have already been re-sequenced for a number of oomycete species. For example, resequencing of 107 *Ph. ramorum* isolates revealed rapid evolution within lineages (Dale et al., 2019). Novel genotypic diversity within lineages was caused by mitotic recombination leading to loss of heterozygosity affecting 2,698 genes. Accelerated evolution was detected in non-core regions which are enriched in effector genes and transposons. Furthermore, positive selection was observed in 8.0% of RxLRs and 18.8% of CRNs. It is estimated that of the four lineages (EU1, EU2, NA1, and NA2), EU1 and NA1 diverged 0.75 MYA, NA2 diverged from EU1 and NA1 1.06 MYA, and EU2 diverged from the other lineages some 1.3 MYA (Dale et al., 2019). Future resequencing studies of diverse strains of individual oomycete species will facilitate pan-genomic analyses which can provide insights into strain-level variation of gene content,

genetic differences in pathogenicity aggressiveness, mechanisms of fungicide resistance and differences in host ranges. This will enable oomycete research to transition from the genomics era to the pan-genomics era.

10. Conclusions and Future Outlook

In line with recent advances in next-generation sequencing technologies and drastic decreases in the cost of whole-genome sequencing, there has been an increased pace of oomycete genome sequencing. Oomycete genomics has revealed fundamental insights into the biology and evolution of oomycetes. Based on the available genomic data, it is clear that HGT has played a significant role in shaping oomycete genome evolution, in particular, it has had a major impact in the evolution of plant pathogenicity and the convergent evolution of similar traits between oomycetes and fungi. Phylogenomic studies have helped to better understand the evolutionary relationships between oomycete species. Genome-wide cataloguing of oomycete effectors has led to the identification of large, diverse apoplastic and cytoplasmic effector families which facilitates experimentation to better understand the disease process and identify future targets for treatment.

With the increasing number of oomycete genomes there is a need for more dedicated tools to be developed to study oomycete genomics. While many genome assemblies are available for the crown oomycete orders, there is a dearth of genomic data available for more basal oomycete species. Furthermore, there is a lack of genomic data for non-pathogenic oomycetes or saprotrophic oomycetes that play key ecological roles in natural environments. Future sequencing of basal oomycetes will reveal further insights into oomycete genome evolution and the origin of oomycetes and the evolution of pathogenicity.

Acknowledgments

J.M. is funded by a postgraduate scholarship from the Irish Research Council, Government of Ireland (grant number GOIPG/2016/1112). We acknowledge the DJEI/DES/SFI/HEA Irish Centre for High-End Computing (ICHEC) for the provision of computational facilities and support.

References

Adhikari, B.N., Hamilton, J.P., Zerillo, M.M., Tisserat, N., Lévesque, C.A., Buell, C.R., 2013. Comparative Genomics Reveals Insight into Virulence Strategies of Plant Pathogenic Oomycetes. PLoS One 8.

Ah-Fong, A.M. V., Kim, K.S., Judelson, H.S., 2017a. RNA-seq of life stages of the oomycete Phytophthora infestans reveals dynamic changes in metabolic, signal transduction, and pathogenesis genes and a major role for calcium signaling in development. BMC Genomics 18, 198.

Ah-Fong, A.M. V., Shrivastava, J., Judelson, H.S., 2017b. Lifestyle, gene gain and loss, and transcriptional remodeling cause divergence in the transcriptomes of Phytophthora infestans and Pythium ultimum during potato tuber colonization. BMC Genomics 18, 764.

Ali, S.S., Shao, J., Lary, D.J., Kronmiller, B.A., Shen, D., Strem, M.D., Amoako-Attah, I., Akrofi, A.Y., Begoude, B.A.D., ten Hoopen, G.M., Coulibaly, K., Kebe, B.I., Melnick, R.L., Guiltinan, M.J., Tyler, B.M., Meinhardt, L.W., Bailey, B.A., 2017. Phytophthora megakarya and Phytophthora palmivora, Closely Related Causal Agents of Cacao Black Pod Rot, Underwent Increases in Genome Sizes and Gene Numbers by Different Mechanisms. Genome Biol. Evol. 9, 536–557.

Almagro Armenteros, J.J., Tsirigos, K.D., Sønderby, C.K., Petersen, T.N., Winther, O., Brunak, S., von Heijne, G., Nielsen, H., 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat. Biotechnol. 37, 420–423.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Armitage, A.D., Lysøe, E., Nellist, C.F., Lewis, L.A., Cano, L.M., Harrison, R.J., Brurberg, M.B., 2018. Bioinformatic characterisation of the effector repertoire of the strawberry pathogen Phytophthora cactorum. PLoS One 13, 1–24.

Basenko, E., Pulman, J., Shanmugasundram, A., Harb, O., Crouch, K., Starns, D., Warrenfeltz, S., Aurrecoechea, C., Stoeckert, C., Kissinger, J., Roos, D., Hertz-Fowler, C., 2018. FungiDB: An Integrated Bioinformatic Resource for Fungi and Oomycetes. J. Fungi 4, 39.

Baxter, L., Tripathy, S., Ishaque, N., Boot, N., Cabral, A., Kemen, E., Thines, M., Ah-Fong, A., Anderson, R., Badejoko, W., Bittner-Eddy, P., Boore, J.L., Chibucos,

M.C., Coates, M., Dehal, P., Delehaunty, K., Dong, S., Downton, P., Dumas, B., Fabro, G., Fronick, C., Fuerstenberg, S.I., Fulton, L., Gaulin, E., Govers, F., Hughes, L., Humphray, S., Jiang, R.H.Y., Judelson, H., Kamoun, S., Kyung, K., Meijer, H., Minx, P., Morris, P., Nelson, J., Phuntumart, V., Qutob, D., Rehmany, A., Rougon-Cardoso, A., Ryden, P., Torto-Alalibo, T., Studholme, D.J., Wang, Y., Win, J., Wood, J., Clifton, S.W., Rogers, J., Van den Ackerveken, G., Jones, J.D.G., McDowell, J.M., Beynon, J., Tyler, B.M., 2010. Signatures of Adaptation to Obligate Biotrophy in the Hyaloperonospora arabidopsidis Genome. Science (80-.). 330, 1549–1551.

Beakes, G.W., Glockling, S.L., Sekimoto, S., 2012. The evolutionary phylogeny of the oomycete "fungi." Protoplasma 249, 3–19.

Beakes, G.W., Thines, M., 2016. Hyphochytriomycota and Oomycota. In: Archibald, J.M., Simpson, A.G.B., Slamovits, C.H., Margulis, L., Melkonian, M., Chapman, D.J., Corliss, J.O. (Eds.), Handbook of the Protists. Springer International Publishing, Cham, pp. 1–71.

Belbahri, L., Calmin, G., Mauch, F., Andersson, J.O., 2008. Evolution of the cutinase gene family: Evidence for lateral gene transfer of a candidate Phytophthora virulence factor. Gene 408, 1–8.

Benhamou, N., le Floch, G., Vallance, J., Gerbore, J., Grizard, D., Rey, P., 2012. Pythium oligandrum: An example of opportunistic success. Microbiol. (United Kingdom) 158, 2679–2694.

Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Sayers, E.W., 2012. GenBank. Nucleic Acids Res. 41, D36–D42.

Blair, J.E., Coffey, M.D., Park, S.Y., Geiser, D.M., Kang, S., 2008. A multi-locus phylogeny for Phytophthora utilizing markers derived from complete genome sequences. Fungal Genet. Biol. 45, 266–277.

Bourret, T.B., Choudhury, R.A., Mehl, H.K., Blomquist, C.L., McRoberts, N., Rizzo, D.M., 2018. Multiple origins of downy mildews and mito-nuclear discordance within the paraphyletic genus Phytophthora. PLoS One 13, e0192502.

Burki, F., Roger, A.J., Brown, M.W., Simpson, A.G.B., 2020. The New Tree of Eukaryotes. Trends Ecol. Evol. 35, 43–55.

Cabral, A., Stassen, J.H.M., Seidl, M.F., Bautor, J., Parker, J.E., Van den Ackerveken, G., 2011. Identification of Hyaloperonospora arabidopsidis Transcript Sequences Expressed during Infection Reveals Isolate-Specific Effectors. PLoS One 6,

e19328.

Cai, M., Wang, Z., Ni, X., Hou, Y., Peng, Q., Gao, X., Liu, X., 2019. Insights from the proteome profile of Phytophthora capsici in response to the novel fungicide SYP-14288. PeerJ 7, e7626.

Capella-Gutierrez, S., Silla-Martinez, J.M., Gabaldon, T., 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973.

Chen, X.-R., Xing, Y.-P., Li, Y.-P., Tong, Y.-H., Xu, J.-Y., 2013. RNA-Seq Reveals Infection-Related Gene Expression Changes in Phytophthora capsici. PLoS One 8, e74588.

Cheung, F., Win, J., Lang, J.M., Hamilton, J., Vuong, H., Leach, J.E., Kamoun, S., André Lévesque, C., Tisserat, N., Buell, C.R., 2008. Analysis of the Pythium ultimum transcriptome using Sanger and Pyrosequencing approaches. BMC Genomics 9, 542.

Cissé, O.H., Stajich, J.E., 2019. FGMP: assessing fungal genome completeness. BMC Bioinformatics 20, 184.

Cui, C., Herlihy, J.H., Bombarely, A., McDowell, J.M., Haak, D.C., 2019. Draft Assembly of Phytophthora capsici from Long-Read Sequencing Uncovers Complexity. Mol. Plant-Microbe Interact. 32, 1559–1563.

Dale, A.L., Feau, N., Everhart, S.E., Dhillon, B., Wong, B., Sheppard, J., Bilodeau, G.J., Brar, A., Tabima, J.F., Shen, D., Brasier, C.M., Tyler, B.M., Grünwald, N.J., Hamelin, R.C., 2019. Mitotic recombination and rapid genome evolution in the invasive forest pathogen phytophthora ramorum. MBio 10, 1–19.

de Cock, A.W.A.M., Lodhi, A.M., Rintoul, T.L., Bala, K., Robideau, G.P., Abad, Z.G., Coffey, M.D., Shahzad, S., Lévesque, C.A., 2015. Phytopythium: molecular phylogeny and systematics. Persoonia - Mol. Phylogeny Evol. Fungi 34, 25–39.

Derevnina, L., Chin-Wo-Reyes, S., Martin, F., Wood, K., Froenicke, L., Spring, O., Michelmore, R., 2015. Genome Sequence and Architecture of the Tobacco Downy Mildew Pathogen Peronospora tabacina. Mol. Plant-Microbe Interact. 28, 1198–1215.

Dong, S., Raffaele, S., Kamoun, S., 2015. The two-speed genomes of filamentous pathogens: waltz with plants. Curr. Opin. Genet. Dev. 35, 57–65.

Dou, D., Kale, S.D., Wang, X., Jiang, R.H.Y., Bruce, N. a, Arredondo, F.D., Zhang, X., Tyler, B.M., 2008. RXLR-Mediated Entry of Phytophthora sojae Effector Avr1b

into Soybean Cells Does Not Require Pathogen-Encoded Machinery. Plant Cell 20, 1930–1947.

Dussert, Y., Mazet, I.D., Couture, C., Gouzy, J., Piron, M.-C., Kuchly, C., Bouchez, O., Rispe, C., Mestre, P., Delmotte, F., 2019. A High-Quality Grapevine Downy Mildew Genome Assembly Reveals Rapidly Evolving and Lineage-Specific Putative Host Adaptation Genes. Genome Biol. Evol. 11, 954–969.

Edgar, R.C., 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.

Emms, D.M., Kelly, S., 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 20, 238.

Evangelisti, E., Gogleva, A., Hainaux, T., Doumane, M., Tulin, F., Quan, C., Yunusov, T., Floch, K., Schornack, S., 2017. Time-resolved dual transcriptomics reveal early induced Nicotiana benthamiana root genes and conserved infection-promoting Phytophthora palmivora effectors. BMC Biol. 15, 39.

Fang, Y., Coelho, M.A., Shu, H., Schotanus, K., Thimmappa, B.C., Yadav, V., Chen, H., Malc, E.P., Wang, J., Mieczkowski, P.A., Kronmiller, B., Tyler, B.M., Sanyal, K., Dong, S., Nowrousian, M., Heitman, J., 2020. Long transposon-rich centromeres in an oomycete reveal divergence of centromere features in Stramenopila-Alveolata-Rhizaria lineages. PLOS Genet. 16, 1–30.

Feau, N., Taylor, G., Dale, A.L., Dhillon, B., Bilodeau, G.J., Birol, I., Jones, S.J.M., Hamelin, R.C., 2016. Genome sequences of six Phytophthora species threatening forest ecosystems. Genomics Data 10, 85–88.

Finn, R.D., Attwood, T.K., Babbitt, P.C., Bateman, A., Bork, P., Bridge, A.J., Chang, H.-Y., Dosztányi, Z., El-Gebali, S., Fraser, M., Gough, J., Haft, D., Holliday, G.L., Huang, H., Huang, X., Letunic, I., Lopez, R., Lu, S., Marchler-Bauer, A., Mi, H., Mistry, J., Natale, D.A., Necci, M., Nuka, G., Orengo, C.A., Park, Y., Pesseat, S., Piovesan, D., Potter, S.C., Rawlings, N.D., Redaschi, N., Richardson, L., Rivoire, C., Sangrador-Vegas, A., Sigrist, C., Sillitoe, I., Smithers, B., Squizzato, S., Sutton, G., Thanki, N., Thomas, P.D., Tosatto, S.C.E., Wu, C.H., Xenarios, I., Yeh, L.-S., Young, S.-Y., Mitchell, A.L., 2017. InterPro in 2017—beyond protein family and domain annotations. Nucleic Acids Res. 45, D190–D199.

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G.A., Tate, J., Bateman, A., 2016. The Pfam protein families database: towards a more

sustainable future. Nucleic Acids Res. 44, D279–D285.

Fletcher, K., Gil, J., Bertier, L.D., Kenefick, A., Wood, K.J., Zhang, L., Reyes-Chin-Wo, S., Cavanaugh, K., Tsuchida, C., Wong, J., Michelmore, R., 2019. Genomic signatures of heterokaryosis in the oomycete pathogen Bremia lactucae. Nat. Commun. 10, 1–13.

Fletcher, K., Klosterman, S.J., Derevnina, L., Martin, F., Bertier, L.D., Koike, S., Reyes-Chin-Wo, S., Mou, B., Michelmore, R., 2018. Comparative genomics of downy mildews reveals potential adaptations to biotrophy. BMC Genomics 19, 8–10.

Gaastra, W., Lipman, L.J.A., De Cock, A.W.A.M., Exel, T.K., Pegge, R.B.G., Scheurwater, J., Vilela, R., Mendoza, L., 2010. Pythium insidiosum: An overview. Vet. Microbiol. 146, 1–16.

Gao, R., Cheng, Y., Wang, Y., Wang, Y., Guo, L., Zhang, G., 2015. Genome Sequence of Phytophthora fragariae var. fragariae, a Quarantine Plant-Pathogenic Fungus. Genome Announc. 3, 6–7.

Gaulin, E., Pel, M.J.C., Camborde, L., San-Clemente, H., Courbier, S., Dupouy, M.-A., Lengellé, J., Veyssiere, M., Le Ru, A., Grandjean, F., Cordaux, R., Moumen, B., Gilbert, C., Cano, L.M., Aury, J.-M., Guy, J., Wincker, P., Bouchez, O., Klopp, C., Dumas, B., 2018. Genomics analysis of Aphanomyces spp. identifies a new class of oomycete effector associated with host adaptation. BMC Biol. 16, 43.

Grayburn, W.S., Hudspeth, D.S.S., Gane, M.K., Hudspeth, M.E.S., 2004. The mitochondrial genome of Saprolegnia ferax : organization, gene content and nucleotide sequence. Mycologia 96, 981–989.

Grenville-Briggs, L.J., Anderson, V.L., Fugelstad, J., Avrova, A.O., Bouzenzana, J., Williams, A., Wawra, S., Whisson, S.C., Birch, P.R.J., Bulone, V., van West, P., 2008. Cellulose Synthesis in Phytophthora infestans Is Required for Normal Appressorium Formation and Successful Infection of Potato. Plant Cell 20, 720–738.

Grenville-Briggs, L.J., Avrova, A.O., Hay, R.J., Bruce, C.R., Whisson, S.C., van West, P., 2010. Identification of appressorial and mycelial cell wall proteins and a survey of the membrane proteome of Phytophthora infestans. Fungal Biol. 114, 702–723.

Grünwald, N.J., Martin, F.N., Larsen, M.M., Sullivan, C.M., Press, C.M., Coffey, M.D., Hansen, E.M., Parke, J.L., 2011. Phytophthora-ID.org: A Sequence-Based Phytophthora Identification Tool. Plant Dis. 95, 337–342.

Haas, B.J., Kamoun, S., Zody, M.C., Jiang, R.H.Y., Handsaker, R.E., Cano, L.M., Grabherr, M., Kodira, C.D., Raffaele, S., Torto-Alalibo, T., Bozkurt, T.O., Ah-Fong, A.M. V., Alvarado, L., Anderson, V.L., Armstrong, M.R., Avrova, A., Baxter, L., Beynon, J., Boevink, P.C., Bollmann, S.R., Bos, J.I.B., Bulone, V., Cai, G., Cakir, C., Carrington, J.C., Chawner, M., Conti, L., Costanzo, S., Ewan, R., Fahlgren, N., Fischbach, M.A., Fugelstad, J., Gilroy, E.M., Gnerre, S., Green, P.J., Grenville-Briggs, L.J., Griffith, J., Grünwald, N.J., Horn, K., Horner, N.R., Hu, C.-H., Huitema, E., Jeong, D.-H., Jones, A.M.E., Jones, J.D.G., Jones, R.W., Karlsson, E.K., Kunjeti, S.G., Lamour, K., Liu, Z., Ma, L., MacLean, D., Chibucos, M.C., McDonald, H., McWalters, J., Meijer, H.J.G., Morgan, W., Morris, P.F., Munro, C.A., O'Neill, K., Ospina-Giraldo, M., Pinzón, A., Pritchard, L., Ramsahoye, B., Ren, Q., Restrepo, S., Roy, S., Sadanandom, A., Savidor, A., Schornack, S., Schwartz, D.C., Schumann, U.D., Schwessinger, B., Seyer, L., Sharpe, T., Silvar, C., Song, J., Studholme, D.J., Sykes, S., Thines, M., van de Vondervoort, P.J.I., Phuntumart, V., Wawra, S., Weide, R., Win, J., Young, C., Zhou, S., Fry, W., Meyers, B.C., van West, P., Ristaino, J., Govers, F., Birch, P.R.J., Whisson, S.C., Judelson, H.S., Nusbaum, C., 2009. Genome sequence and analysis of the Irish potato famine pathogen Phytophthora infestans. Nature 461, 393–398.

Hao, K., Lin, B., Nian, F., Gao, X., Wei, Z., Luo, G., Lu, Y., Lan, M., Yang, J., Wu, G., 2019. RNA-seq analysis of the response of plant-pathogenic oomycete Phytophthora parasitica to the fungicide dimethomorph. Rev. Argent. Microbiol. 51, 268–277.

Hardham, A.R., 2005. Phytophthora cinnamomi. Mol. Plant Pathol.

Horner, N.R., Grenville-Briggs, L.J., van West, P., 2012. The oomycete Pythium oligandrum expresses putative effectors during mycoparasitism of Phytophthora infestans and is amenable to transformation. Fungal Biol. 116, 24–41.

Hosseini, S., Resjö, S., Liu, Yongfeng, Durling, M., Heyman, F., Levander, F., Liu, Yanhong, Elfstrand, M., Funck Jensen, D., Andreasson, E., Karlsson, M., 2015. Comparative proteomic analysis of hyphae and germinating cysts of Phytophthora pisi and Phytophthora sojae. J. Proteomics 117, 24–40.

Hu, K., Ma, R.-R., Cheng, J.-M., Ye, X., Sun, Q., Yuan, H.-L., Liang, N., Fang, W.-H., Li, H.-R., Yang, X.-L., 2016. Analysis of Saprolegnia parasitica Transcriptome following Treatment with Copper Sulfate. PLoS One 11, e0147445.

Jiang, R.H.Y., de Bruijn, I., Haas, B.J., Belmonte, R., Löbach, L., Christie, J., van den Ackerveken, G., Bottin, A., Bulone, V., Díaz-Moreno, S.M., Dumas, B., Fan, L., Gaulin, E., Govers, F., Grenville-Briggs, L.J., Horner, N.R., Levin, J.Z., Mammella, M., Meijer, H.J.G., Morris, P., Nusbaum, C., Oome, S., Phillips, A.J., van Rooyen, D., Rzeszutek, E., Saraiva, M., Secombes, C.J., Seidl, M.F., Snel, B., Stassen, J.H.M., Sykes, S., Tripathy, S., van den Berg, H., Vega-Arreguin, J.C., Wawra, S., Young, S.K., Zeng, Q., Dieguez-Uribeondo, J., Russ, C., Tyler, B.M., van West, P., 2013. Distinctive Expansion of Potential Virulence Genes in the Genome of the Oomycete Fish Pathogen Saprolegnia parasitica. PLoS Genet. 9, e1003272.

Jiang, R.H.Y., Tyler, B.M., 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.Y., Tyler, B.M., Whisson, S.C., Hardham, A.R., Govers, F., 2006b. Ancient Origin of Elicitin Gene Clusters in Phytophthora Genomes. Mol. Biol. Evol. 23, 338–351.

Judelson, H.S., Shrivastava, J., Manson, J., 2012. Decay of Genes Encoding the Oomycete Flagellar Proteome in the Downy Mildew Hyaloperonospora arabidopsidis. PLoS One 7, e47624.

Jupe, J., Stam, R., Howden, A.J.M., Morris, J.A., Zhang, R., Hedley, P.E., Huitema, E., 2013. Phytophthora capsici-tomato interaction features dramatic shifts in gene expression associated with a hemi-biotrophic lifestyle. Genome Biol. 14.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14, 587–589.

Kamoun, S., 2003. Molecular genetics of pathogenic oomycetes. Eukaryot. Cell 2, 191–9.

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

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., Jones, J.D.G., 2011. Gene gain and loss during evolution of obligate parasitism in the white rust pathogen of Arabidopsis thaliana. PLoS Biol. 9.

Krings, M., Taylor, T.N., Dotzler, N., 2011. The fossil record of the Peronosporomycetes (Oomycota). Mycologia 103, 445–457.

Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E.L.., 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305, 567–80.

Kushwaha, S.K., Vetukuri, R.R., Grenville-Briggs, L.J., 2017a. Draft Genome Sequence of the Mycoparasitic Oomycete Pythium periplocum Strain CBS 532.74. Genome Announc. 5, e00556-18.

Kushwaha, S.K., Vetukuri, R.R., Grenville-Briggs, L.J., 2017b. Draft Genome Sequence of the Mycoparasitic Oomycete Pythium oligandrum Strain CBS 530.74. Genome Announc. 5, e00556-18.

Lamour, K.H., Mudge, J., Gobena, D., Hurtado-Gonzales, O.P., Schmutz, J., Kuo, A., Miller, N.A., Rice, B.J., Raffaele, S., Cano, L.M., Bharti, A.K., Donahoo, R.S., Finley, S., Huitema, E., Hulvey, J., Platt, D., Salamov, A., Savidor, A., Sharma, R., Stam, R., Storey, D., Thines, M., Win, J., Haas, B.J., Dinwiddie, D.L., Jenkins, J., Knight, J.R., Affourtit, J.P., Han, C.S., Chertkov, O., Lindquist, E.A., Detter, C., Grigoriev, I. V, Kamoun, S., Kingsmore, S.F., 2012. Genome sequencing and mapping reveal loss of heterozygosity as a mechanism for rapid adaptation in the vegetable pathogen Phytophthora capsici. Mol. Plant-Microbe Interact. 25, 1350–1360.

Lassiter, E.S., Russ, C., Nusbaum, C., Zeng, Q., Saville, A.C., Olarte, R.A., Carbone, I., Hu, C.-H., Seguin-Orlando, A., Samaniego, J.A., Thorne, J.L., Ristaino, J.B., 2015. Mitochondrial genome sequences reveal evolutionary relationships of the Phytophthora 1c clade species. Curr. Genet. 61, 567–577.

Lee, S.-J., Kelley, B.S., Damasceno, C.M.B., John, B. St., Kim, B.-S., Kim, B.-D., Rose, J.K.C., 2006. A Functional Screen to Characterize the Secretomes of Eukaryotic Pathogens and Their Hosts In Planta. Mol. Plant-Microbe Interact. 19, 1368–1377.

Leonard, G., Labarre, A., Milner, D.S., Monier, A., Soanes, D., Wideman, J.G., Maguire, F., Stevens, S., Sain, D., Grau-Bové, X., Sebé-Pedrós, A., Stajich, J.E., Paszkiewicz, K., Brown, M.W., Hall, N., Wickstead, B., Richards, T.A., 2018. Comparative genomic analysis of the ‘pseudofungus’ Hyphochytrium catenoides. Open Biol. 8, 170184.

Letunic, I., Bork, P., 2007. Interactive Tree Of Life (iTOL): An online tool for

phylogenetic tree display and annotation. Bioinformatics 23, 127–128.

Lévesque, 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 Biol. 11, R73.

LéVesque, C.A., De Cock, A.W.A.M., 2004. Molecular phylogeny and taxonomy of the genus Pythium. Mycol. Res. 108, 1363–1383.

Links, M.G., Holub, E., Jiang, R.H., Sharpe, A.G., Hegedus, D., Beynon, E., Sillito, D., Clarke, W.E., Uzuhashi, S., Borhan, M.H., 2011. De novo sequence assembly of Albugo candida reveals a small genome relative to other biotrophic oomycetes. BMC Genomics 12, 503.

Malar C, M., Yuzon, J.D., Das, S., Das, A., Panda, A., Ghosh, S., Tyler, B.M., Kasuga, T., Tripathy, S., 2019a. Haplotype-Phased Genome Assembly of Virulent Phytophthora ramorum Isolate ND886 Facilitated by Long-Read Sequencing Reveals Effector Polymorphisms and Copy Number Variation. Mol. Plant-Microbe Interact. 32, 1047–1060.

Malar C, M., Yuzon, J.D., Panda, A., Kasuga, T., Tripathy, S., 2019b. Updated Assembly of Phytophthora ramorum pr102 Isolate Incorporating Long Reads from PacBio Sequencing. Mol. Plant-Microbe Interact. 32, 1472–1474.

Martin, F.N., Bensasson, D., Tyler, B.M., Boore, J.L., 2007. Mitochondrial genome sequences and comparative genomics of Phytophthora ramorum and P. sojae. Curr. Genet. 51, 285–296.

Martin, F.N., Blair, J.E., Coffey, M.D., 2014. A combined mitochondrial and nuclear multilocus phylogeny of the genus Phytophthora. Fungal Genet. Biol. 66, 19–32.

Matari, N.H., Blair, J.E., 2014. A multilocus timescale for oomycete evolution estimated under three distinct molecular clock models. BMC Evol. Biol. 14, 101.

McCarthy, C.G.P., Fitzpatrick, D.A., 2016. Systematic Search for Evidence of Interdomain Horizontal Gene Transfer from Prokaryotes to Oomycete Lineages.

mSphere 1, 1–18.

McCarthy, C.G.P., Fitzpatrick, D.A., 2017. Phylogenomic Reconstruction of the Oomycete Phylogeny Derived from 37 Genomes. mSphere 2, e00095-17.

McGowan, J., Byrne, K.P., Fitzpatrick, D.A., 2019. Comparative Analysis of Oomycete Genome Evolution Using the Oomycete Gene Order Browser (OGOB). Genome Biol. Evol. 11, 189–206.

McGowan, J., Fitzpatrick, D.A., 2017. Genomic, Network, and Phylogenetic Analysis of the Oomycete Effector Arsenal. mSphere 2, e00408-17.

Meijer, H.J.G., Mancuso, F.M., Espadas, G., Seidl, M.F., Chiva, C., Govers, F., Sabido, E., 2014. Profiling the Secretome and Extracellular Proteome of the Potato Late Blight Pathogen Phytophthora infestans. Mol. Cell. Proteomics 13, 2101–2113.

Misner, I., Blouin, N., Leonard, G., Richards, T.A., Lane, C.E., 2015. The Secreted Proteins of Achlya hypogyna and Thraustotheca clavata Identify the Ancestral Oomycete Secretome and Reveal Gene Acquisitions by Horizontal Gene Transfer. Genome Biol. Evol. 7, 120–135.

Nayaka, S.C., Shetty, H.S., Satyavathi, C.T., Yadav, R.S., Kishor, P.B.K., Nagaraju, M., Anoop, T.A., Kumar, M.M., Kuriakose, B., Chakravartty, N., Katta, A.V.S.K.M., Lachagari, V.B.R., Singh, O.V., Sahu, P.P., Puranik, S., Kaushal, P., Srivastava, R.K., 2017. Draft genome sequence of Sclerospora graminicola, the pearl millet downy mildew pathogen. Biotechnol. Reports 16, 18–20.

Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274.

Oome, S., Raaymakers, T.M., Cabral, A., Samwel, S., Böhm, H., Albert, I., Nürnberger, T., Van den Ackerveken, G., 2014. Nep1-like proteins from three kingdoms of life act as a microbe-associated molecular pattern in Arabidopsis. Proc. Natl. Acad. Sci. U. S. A. 111, 16955–60.

Orsomando, G., Lorenzi, M., Raffaelli, N., Dalla Rizza, M., Mezzetti, B., Ruggieri, S., 2001. Phytotoxic Protein PcF, Purification, Characterization, and cDNA Sequencing of a Novel Hydroxyproline-containing Factor Secreted by the Strawberry Pathogen Phytophthora cactorum. J. Biol. Chem. 276, 21578–21584.

Panda, A., Sen, D., Ghosh, A., Gupta, A., C., M.M., Prakash Mishra, G., Singh, D., Ye, W., Tyler, B.M., Tripathy, S., 2018. EumicrobeDBLite: a lightweight genomic resource and analytic platform for draft oomycete genomes. Mol. Plant Pathol. 19,

227–237.

Paquin, B., Laforest, M.-J., Forget, L., Roewer, I., Wang, Z., Longcore, J., Lang, B.F., 1997. The fungal mitochondrial genome project: evolution of fungal mitochondrial genomes and their gene expression. Curr. Genet. 31, 380–395.

Raaymakers, T.M., Van den Ackerveken, G., 2016. Extracellular Recognition of Oomycetes during Biotrophic Infection of Plants. Front. Plant Sci. 7, 1–12.

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

Rahman, A., Góngora-Castillo, E., Bowman, M.J., Childs, K.L., Gent, D.H., Martin, F.N., Quesada-Ocampo, L.M., 2019. Genome Sequencing and Transcriptome Analysis of the Hop Downy Mildew Pathogen Pseudoperonospora humuli Reveal Species-Specific Genes for Molecular Detection. Phytopathology 109, 1354–1366.

Resjö, S., Ali, A., Meijer, H.J.G., Seidl, M.F., Snel, B., Sandin, M., Levander, F., Govers, F., Andreasson, E., 2014. Quantitative label-free phosphoproteomics of six different life stages of the late blight pathogen phytophthora infestans reveals abundant phosphorylation of members of the CRN effector family. J. Proteome Res. 13, 1848–1859.

Resjö, S., Brus, M., Ali, A., Meijer, H.J.G., Sandin, M., Govers, F., Levander, F., Grenville-Briggs, L., Andreasson, E., 2017. Proteomic Analysis of Phytophthora infestans Reveals the Importance of Cell Wall Proteins in Pathogenicity. Mol. Cell. Proteomics 16, 1958–1971.

Rhoads, A., Au, K.F., 2015. PacBio Sequencing and Its Applications. Genomics. Proteomics Bioinformatics 13, 278–289.

Richards, T.A., Dacks, J.B., Jenkinson, J.M., Thornton, C.R., Talbot, N.J., 2006. Evolution of Filamentous Plant Pathogens: Gene Exchange across Eukaryotic Kingdoms. Curr. Biol. 16, 1857–1864.

Richards, T.A., Soanes, D.M., Jones, M.D.M., Vasieva, O., Leonard, G., Paszkiewicz, K., Foster, P.G., Hall, N., Talbot, N.J., 2011. Horizontal gene transfer facilitated the evolution of plant parasitic mechanisms in the oomycetes. Proc. Natl. Acad. Sci. 108, 15258–15263.

Rizzo, D.M., Garbelotto, M., Hansen, E.M., 2005. Phytophthora ramorum: Integrative Research and Management of an Emerging Pathogen in California and Oregon Forests. Annu. Rev. Phytopathol. 43, 309–335.

Roquigny, R., Novinscak, A., Arseneault, T., Joly, D.L., Filion, M., 2018. Transcriptome alteration in Phytophthora infestans in response to phenazine-1-carboxylic acid production by Pseudomonas fluorescens strain LBUM223. BMC Genomics 19, 474.

Rujirawat, T., Patumcharoenpol, P., Lohnoo, T., Yingyong, W., Kumsang, Y., Payattikul, P., Tangphatsornruang, S., Suriyaphol, P., Reamtong, O., Garg, G., Kittichotirat, W., Krajaejun, T., 2018. Probing the Phylogenomics and Putative Pathogenicity Genes of Pythium insidiosum by Oomycete Genome Analyses. Sci. Rep. 8, 4135.

Rujirawat, T., Patumcharoenpol, P., Lohnoo, T., Yingyong, W., Lerksuthirat, T., Tangphatsornruang, S., Suriyaphol, P., Grenville-Briggs, L.J., Garg, G., Kittichotirat, W., Krajaejun, T., 2015. Draft Genome Sequence of the Pathogenic Oomycete Pythium insidiosum Strain Pi-S, Isolated from a Patient with Pythiosis. Genome Announc. 3, e00574-15.

Savidor, A., Donahoo, R.S., Hurtado-Gonzales, O., Land, M.L., Shah, M.B., Lamour, K.H., McDonald, W.H., 2008. Cross-species Global Proteomics Reveals Conserved and Unique Processes in *Phytophthora sojae* and *Phytophthora ramorum*. Mol. Cell. Proteomics 7, 1501–1516.

Savory, F.R., Leonard, G., Richards, T.A., 2015. The Role of Horizontal Gene Transfer in the Evolution of the Oomycetes. PLOS Pathog. 11, e1004805.

Savory, F.R., Milner, D.S., Miles, D.C., Richards, T.A., 2018. Ancestral Function and Diversification of a Horizontally Acquired Oomycete Carboxylic Acid Transporter. Mol. Biol. Evol. 35, 1887–1900.

Schornack, S., van Damme, M., Bozkurt, T.O., Cano, L.M., Smoker, M., Thines, M., Gaulin, E., Kamoun, S., Huitema, E., 2010. Ancient class of translocated oomycete effectors targets the host nucleus. Proc. Natl. Acad. Sci. U. S. A. 107, 17421–17426.

Seidl, M.F., Van den Ackerveken, G., 2019. Activity and Phylogenetics of the Broadly Occurring Family of Microbial Nep1-Like Proteins. Annu. Rev. Phytopathol. 57, 367–386.

Seidl, M.F., Van Den Ackerveken, G., Govers, F., Snel, B., 2012. Reconstruction of oomycete genome evolution identifies differences in evolutionary trajectories leading to present-day large gene families. Genome Biol. Evol. 4, 199–211.

Severino, V., Farina, A., Fleischmann, F., Dalio, R.J.D., Di Maro, A., Scognamiglio,

M., Fiorentino, A., Parente, A., Osswald, W., Chambery, A., 2014. Molecular profiling of the Phytophthora plurivora secretome: a step towards understanding the cross-talk between plant pathogenic oomycetes and their hosts. PLoS One 9, e112317.

Sharma, R., Xia, X., Cano, L.M., Evangelisti, E., Kemen, E., Judelson, H., Oome, S., Sambles, C., van den Hoogen, D.J., Kitner, M., Klein, J., Meijer, H.J.G., Spring, O., Win, J., Zipper, R., Bode, H.B., Govers, F., Kamoun, S., Schornack, S., Studholme, D.J., Van den Ackerveken, G., Thines, M., 2015. Genome analyses of the sunflower pathogen Plasmopara halstedii provide insights into effector evolution in downy mildews and Phytophthora. BMC Genomics 16, 741.

Shen, D., Tang, Z., Wang, C., Wang, J., Dong, Y., Chen, Y., Wei, Y., Cheng, B., Zhang, M., Grenville-Briggs, L.J., Tyler, B.M., Dou, D., Xia, A., 2019. Infection mechanisms and putative effector repertoire of the mosquito pathogenic oomycete Pythium guiyangense uncovered by genomic analysis. PLoS Genet. 15, 1–26.

Sperschneider, J., Dodds, P.N., Gardiner, D.M., Manners, J.M., Singh, K.B., Taylor, J.M., 2015. Advances and Challenges in Computational Prediction of Effectors from Plant Pathogenic Fungi. PLOS Pathog. 11, e1004806.

Sperschneider, J., Dodds, P.N., Gardiner, D.M., Singh, K.B., Taylor, J.M., 2018a. Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. Mol. Plant Pathol. 19, 2094–2110.

Sperschneider, J., Dodds, P.N., Singh, K.B., Taylor, J.M., 2018b. ApoplastP: prediction of effectors and plant proteins in the apoplast using machine learning. New Phytol. 217, 1764–1778.

Spies, C.F.J., Grooters, A.M., Lévesque, C.A., Rintoul, T.L., Redhead, S.A., Glockling, S.L., Chen, C., de Cock, A.W.A.M., 2016. Molecular phylogeny and taxonomy of Lagenidium -like oomycetes pathogenic to mammals. Fungal Biol. 120, 931–947.

Srivastava, V., Rezinciuc, S., Bulone, V., 2018. Quantitative Proteomic Analysis of Four Developmental Stages of Saprolegnia parasitica. Front. Microbiol. 8, 1–13.

Studholme, D.J., McDougal, R.L., Sambles, C., Hansen, E., Hardy, G., Grant, M., Ganley, R.J., Williams, N.M., 2015. Genome sequences of six Phytophthora species associated with forests in New Zealand. Genomics Data 7, 54–56.

Studholme, D.J., Panda, P., Sanfuentes Von Stowasser, E., González, M., Hill, R., Sambles, C., Grant, M., Williams, N.M., McDougal, R.L., 2019. Genome sequencing of oomycete isolates from Chile supports the New Zealand origin of

Phytophthora kernoviae and makes available the first Nothophytophthora sp. genome. Mol. Plant Pathol. 20, 423–431.

Sun, J., Gao, Z., Zhang, X., Zou, X., Cao, L., Wang, J., 2017. Transcriptome analysis of Phytophthora litchii reveals pathogenicity arsenals and confirms taxonomic status. PLoS One 12, e0178245.

Tabima, J.F., Kronmiller, B.A., Press, C.M., Tyler, B.M., Zasada, I.A., Grünwald, N.J., 2017. Whole Genome Sequences of the Raspberry and Strawberry Pathogens Phytophthora rubi and P. fragariae. Mol. Plant-Microbe Interact. 30, 767–769.

Tangphatsornruang, S., Ruang-areerate, P., Sangsrakru, D., Rujirawat, T., Lohnoo, T., Kittichotirat, W., Patumcharoenpol, P., Grenville-Briggs, L.J., Krajaejun, T., 2016. Comparative mitochondrial genome analysis of Pythium insidiosum and related oomycete species provides new insights into genetic variation and phylogenetic relationships. Gene 575, 34–41.

Thines, M., Kamoun, S., 2010. Oomycete–plant coevolution: recent advances and future prospects. Curr. Opin. Plant Biol. 13, 427–433.

Thines, M., Sharma, R., Rodenburg, S.Y.A., Gogleva, A., Judelson, H.S., Xia, X., Hoogen, J. van den, Kitner, M., Klein, J., Neilen, M., Ridder, D. de, Seidl, M.F., Ackerveken, G. Van den, Govers, F., Schornack, S., Studholme, D.J., 2019. The genome of Peronospora belbahrii reveals high heterozygosity, a low number of canonical effectors and CT-rich promoters. bioRxiv 721027.

Tian, M., Win, J., Song, J., van der Hoorn, R., van der Knaap, E., Kamoun, S., 2006. A Phytophthora infestans Cystatin-Like Protein Targets a Novel Tomato Papain-Like Apoplastic Protease. PLANT Physiol. 143, 364–377.

Torto, T., Rauser, L., Kamoun, S., 2002. The pipg1 gene of the oomycete Phytophthora infestans encodes a fungal-like endopolygalacturonase. Curr. Genet. 40, 385–390.

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.Y., Aerts, A., Arredondo, F.D., Baxter, L., Bensasson, D., Beynon, J.L., Chapman, J., Damasceno, C.M.B., Dorrance, A.E., Dou, D., Dickerman, A.W., Dubchak, I.L., Garbelotto, M., Gijzen, M., Gordon, S.G., Govers, F., Grunwald, N.J., Huang, W., Ivors, K.L., Jones, R.W., Kamoun, S., Krampis, K., Lamour, K.H., Lee, M.-K., McDonald, W.H., Medina, M., Meijer, H.J.G., Nordberg, E.K., Maclean, D.J., Ospina-Giraldo, M.D., Morris, P.F., Phuntumart, V., Putnam, N.H., Rash, S., Rose, J.K.C., Sakihama, Y.,

Salamov, A.A., Savidor, A., Scheuring, C.F., Smith, B.M., Sobral, B.W.S., Terry, A., Torto-Alalibo, T.A., Win, J., Xu, Z., Zhang, H., Grigoriev, I. V, Rokhsar, D.S., Boore, J.L., 2006. Phytophthora genome sequences uncover evolutionary origins and mechanisms of pathogenesis. Science 313, 1261–6.

Urban, M., Cuzick, A., Rutherford, K., Irvine, A., Pedro, H., Pant, R., Sadanadan, V., Khamari, L., Billal, S., Mohanty, S., Hammond-Kosack, K.E., 2017. PHI-base: a new interface and further additions for the multi-species pathogen–host interactions database. Nucleic Acids Res. 45, D604–D610.

van den Berg, A.H., McLaggan, D., Diéguez-Uribeondo, J., van West, P., 2013. The impact of the water moulds Saprolegnia diclina and Saprolegnia parasitica on natural ecosystems and the aquaculture industry. Fungal Biol. Rev. 27, 33–42.

Vetukuri, R.R., Kushwaha, S.K., Sen, D., Whisson, S.C., Lamour, K.H., Grenville-Briggs, L.J., 2018a. Genome sequence resource for the oomycete taro pathogen phytophthora colocasiae. Mol. Plant-Microbe Interact. 31, 903–905.

Vetukuri, R.R., Tripathy, S., Mathu, M.C., Panda, A., Kushwaha, S.K., Chawade, A., Andreasson, E., Grenville-Briggs, L.J., Whisson, S.C., 2018b. Draft genome sequence for the tree pathogen phytophthora plurivora. Genome Biol. Evol. 10, 2432–2442.

Vilela, R., Humber, R.A., Taylor, J.W., Mendoza, L., 2019. Phylogenetic and physiological traits of oomycetes originally identified as Lagenidium giganteum from fly and mosquito larvae. Mycologia 111, 408–422.

Vries, S. de, Vries, J. de, Archibald, J.M., Slamovits, C.H., 2019. Comparative analyses of saprotrophy in Salisapilia sapeloensis and diverse plant pathogenic oomycetes reveal lifestyle-specific gene expression. bioRxiv 656496.

Wang, S., Boevink, P.C., Welsh, L., Zhang, R., Whisson, S.C., Birch, P.R.J., 2017. Delivery of cytoplasmic and apoplastic effectors from Phytophthora infestans haustoria by distinct secretion pathways. New Phytol. 216, 205–215.

Wang, S., McLellan, H., Bukharova, T., He, Q., Murphy, F., Shi, J., Sun, S., van Weymers, P., Ren, Y., Thilliez, G., Wang, H., Chen, X., Engelhardt, S., Vleeshouwers, V., Gilroy, E.M., Whisson, S.C., Hein, I., Wang, X., Tian, Z., Birch, P.R.J., Boevink, P.C., 2019. Phytophthora infestans RXLR effectors act in concert at diverse subcellular locations to enhance host colonization. J. Exp. Bot. 70, 343–356.

Wang, S., Welsh, L., Thorpe, P., Whisson, S.C., Boevink, P.C., Birch, P.R.J., 2018. The

Phytophthora infestans Haustorium Is a Site for Secretion of Diverse Classes of Infection-Associated Proteins. MBio 9, 1–14.

Waterhouse, R.M., Seppey, M., Simão, F.A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E. V, Zdobnov, E.M., 2018. BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. Mol. Biol. Evol. 35, 543–548.

Wawra, S., Belmonte, R., Löbach, L., Saraiva, M., Willems, A., van West, P., 2012. Secretion, delivery and function of oomycete effector proteins. Curr. Opin. Microbiol. 15, 685–691.

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., 2007. A translocation signal for delivery of oomycete effector proteins into host plant cells. Nature 450, 115–118.

Whitaker, J.W., McConkey, G.A., Westhead, D.R., 2009. The transferome of metabolic genes explored: analysis of the horizontal transfer of enzyme encoding genes in unicellular eukaryotes. Genome Biol. 10, R36.

Win, J., Krasileva, K. V., Kamoun, S., Shirasu, K., Staskawicz, B.J., Banfield, M.J., 2012. Sequence Divergent RXLR Effectors Share a Structural Fold Conserved across Plant Pathogenic Oomycete Species. PLoS Pathog. 8, e1002400.

Yang, X., Hong, C., 2018. Differential Usefulness of Nine Commonly Used Genetic Markers for Identifying Phytophthora Species. Front. Microbiol. 9, 2334.

Ye, W., Wang, Yang, Shen, D., Li, D., Pu, T., Jiang, Z., Zhang, Z., Zheng, X., Tyler, B.M., Wang, Yuanchao, 2016. Sequencing of the Litchi Downy Blight Pathogen Reveals It Is a Phytophthora Species With Downy Mildew-Like Characteristics. Mol. Plant-Microbe Interact. 29, 573–583.

Yin, J., Gu, B., Huang, G., Tian, Y., Quan, J., Lindqvist-Kreuze, H., Shan, W., 2017. Conserved RXLR Effector Genes of Phytophthora infestans Expressed at the Early Stage of Potato Infection Are Suppressive to Host Defense. Front. Plant Sci. 8, 1–11.

Yin, L., An, Y., Qu, J., Li, X., Zhang, Y., Dry, I., Wu, H., Lu, J., 2017. Genome sequence of Plasmopara viticola and insight into the pathogenic mechanism. Sci. Rep. 7, 46553.

Yuan, X., Feng, C., Zhang, Z., Zhang, C., 2017. Complete Mitochondrial Genome of Phytophthora nicotianae and Identification of Molecular Markers for the

Oomycetes. Front. Microbiol. 8, 1–14.

Zerillo, M.M., Adhikari, B.N., Hamilton, J.P., Buell, C.R., Lévesque, C.A., Tisserat, N., 2013. Carbohydrate-Active Enzymes in Pythium and Their Role in Plant Cell Wall and Storage Polysaccharide Degradation. PLoS One 8, e72572.

Zheng, X., McLellan, H., Fraiture, M., Liu, X., Boevink, P.C., Gilroy, E.M., Chen, Y., Kandel, K., Sessa, G., Birch, P.R.J., Brunner, F., 2014. Functionally Redundant RXLR Effectors from Phytophthora infestans Act at Different Steps to Suppress Early flg22-Triggered Immunity. PLoS Pathog. 10, e1004057.

Chapter 2

Genomic, Network, and Phylogenetic Analysis of the Oomycete Effector Arsenal

This chapter has been published in the journal mSphere.

McGowan, J., Fitzpatrick D.A., 2017. Genomic, Network, and Phylogenetic Analysis of the Oomycete Effector Arsenal. mSphere 2, e00408-17.

Abstract

The oomycetes are a class of microscopic, filamentous eukaryotes within the Stramenopiles-Alveolate-Rhizaria (SAR) supergroup and include ecologically significant animal and plant pathogens. Oomycetes secrete large arsenals of effector proteins that degrade host cell components, manipulate host immune responses and induce necrosis, enabling parasitic colonization. This study investigates the expansion and evolution of effectors in 37 oomycete species, from 4 oomycete orders including Albuginales, Peronosporales, Pythiales and Saprolegniales species. Our results highlight the large expansions of effector protein families in *Phytophthora* species, including glycoside hydrolases, pectinases and necrosis-inducing proteins. Species specific expansions were detected, including expansions of chitinases in *Aphanomyces astaci* and *Pythium oligandrum*. Novel effectors which may be involved in suppressing animal immune responses were also identified in *Aphanomyces astaci* and *Pythium insidiosum*. Type 2 necrosis-inducing proteins with an unusual phylogenetic history were also located in a number of oomycete species. We also investigated the RxLR effector complement of all 37 species and as expected observed large expansions in *Phytophthora* species. Our results provide in depth sequence information on all putative RxLR effectors from all 37 species. This work is an up to date *in silico* catalogue of the effector arsenal of the oomycetes based on the 37 genomes currently available.

Introduction

The oomycetes are a class of diverse eukaryotic microorganisms which includes some of the most devastating pathogens of plants, mammals, fish and fungi (1). Previously they were thought to be fungi, due to similar morphology, ecological roles, modes of nutrition and filamentous growth (2). Molecular analyses has placed the oomycetes into the Stramenopiles-Alveolata-Rhizaria (SAR) Eukaryotic Supergroup with close relationships to the diatoms and brown algae (3). Within the oomycete class, there are a number of highly diverse orders including the Saprolegniales, Peronosporales, Albuginales and Pythiales, that exhibit different lifestyles and can have either very specific or very broad host ranges.

More than 60% of known oomycetes are pathogens of plants (4) and have a devastating effect on many agriculturally important crops and ornamental plants. Members of the Saprolegniales order predominantly exhibit saprotrophic lifestyles and include the animal and plant pathogens *Aphanomyces* (5) as well as the fish pathogenic *Saprolegnia* genus, known as "cotton moulds" (6, 7). The Peronosporales order consists largely of phytopathogens and includes the hemibiotrophic genus *Phytophthora* (the "plant destroyers"). *Phytophthora* species include the notorious phytopathogen *Phytophthora infestans* which is the causative agent of late potato blight, a disease reported to cause billions of dollars' worth of damage worldwide annually (8). Also included in the Peronosporales are the genera *Hyaloperonospora* and *Plasmopara* which are closely related to *Phytophthora* species (11) (**Figure 1**). These two genera contain species that cause downy mildew in a number of economically important plants (12–14). In contrast to *Phytophthora* species, they are obligate biotrophs that are completely dependent on their host for survival. Other obligate oomycete biotrophs include the *Albugo* species ("white blister rusts") which are members of the Albuginales order (15,

16) (**Figure 1**). The Pythiales order includes the genus *Pythium*, members of which are necrotrophs that cause root rot in many terrestrial plants. Exceptions include *Pythium insidiosum*, a pathogen causing pythiosis in mammals (17) and *Pythium oligandrum*, a pathogen of other oomycetes and fungi (18). *Pythium* species are divided into 10 clades (A – J) (19) (**Figure 1**).

Oomycetes are notorious for secreting a large arsenal of effector proteins (20). Effectors facilitate infection by manipulating host cell components, exploiting host nutrients, triggering defence responses and inducing necrosis (21). Oomycete effectors can be categorized into two classes (apoplastic or cytoplasmic) depending on where they localize. Apoplastic effectors are secreted by the pathogens and exert their pathogenic activity in the host's extracellular environment (22). Oomycete apoplastic effectors include a large number of hydrolytic enzymes, which are involved in the degradation of host cell components, enabling penetration of host cells. These include cutinases, glycoside hydrolases, pectinases and proteases among others. Some oomycete species, such as *Phytophthora*, also encode extracellular toxin families such as necrosis-inducing proteins and Pcf family toxins (23). Host species are known to secrete protective proteases in an effort to degrade pathogen effectors, for example P69B and P69C are subtilisin-like serine proteases secreted by tomatoes in response to *Phytophthora* species protease (24). To counteract this, *Phytophthora* species secrete protease inhibitors to block the host defence (25).

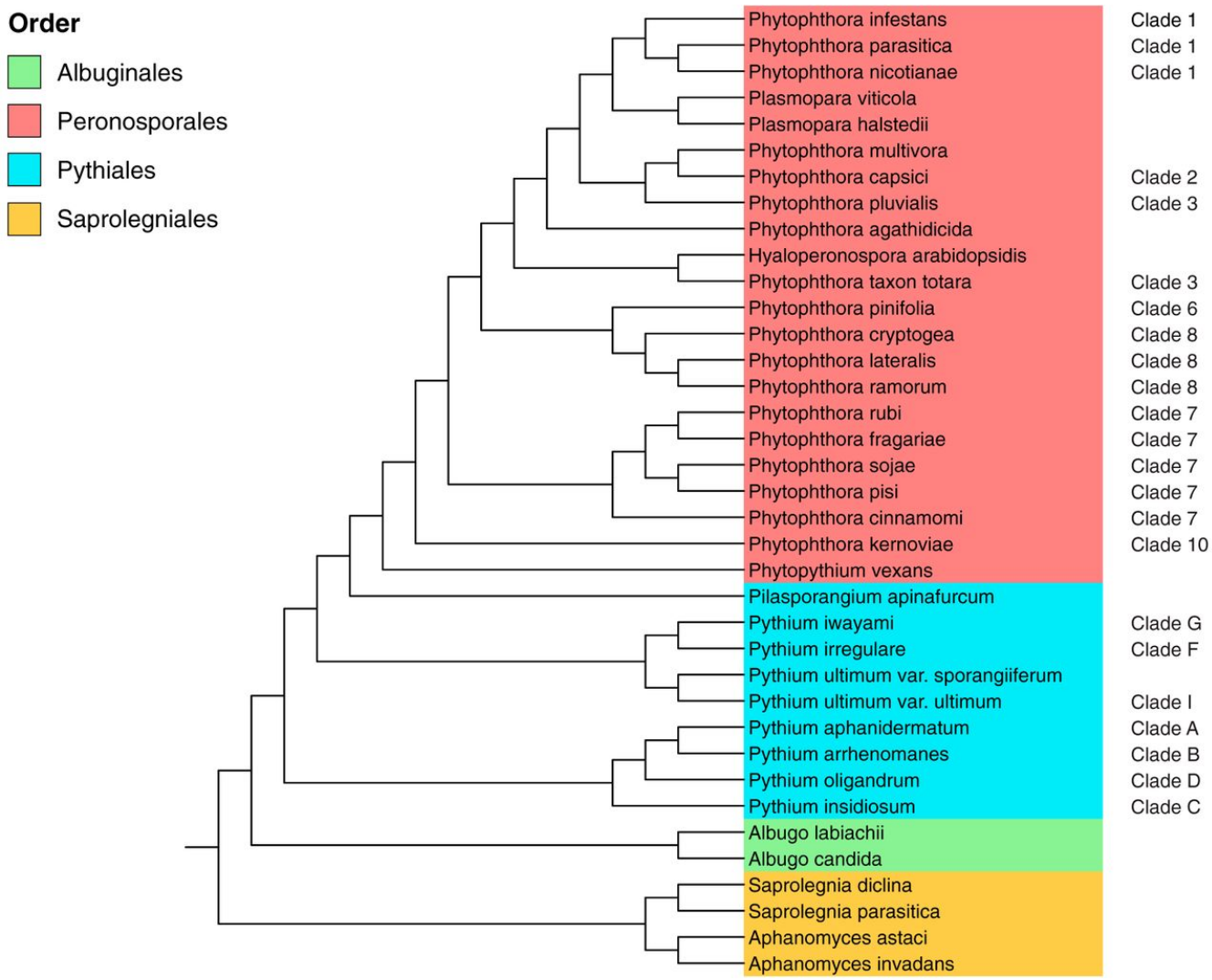


Figure 1. Phylogeny of the 37 oomycete species from 4 oomycete orders considered in this study. *Phytophthora* clades as designated by Blair et al. (2008) and *Pythium* clades as designated by de Cock et al. (2015) in red and blue, respectively. Adapted from reference 11.

In contrast to apoplastic effectors, cytoplasmic effectors are secreted and translocated into the host cell where they exhibit their pathogenic activity. Two types of cytoplasmic effectors dominate the oomycete secretome - RxLR effectors and Crinklers. RxLR effectors (RxLRs) are so named because they contain a highly conserved "RxLR" motif in their N-terminal domain (8, 26). This motif is followed by a downstream 'EER' motif in many RxLRs. Studies have shown that the 'RxLR' motif acts as a translocation signal, marking the protein for trafficking into the host cell (27). The mechanisms of the RxLR motif are thought to be similar to that of the 'Pexel' translocation motif found in effectors of the malaria parasite *Plasmodium falciparum* (27–29). Some RxLRs can enter host cells independent of any additional pathogen encoded machinery (30). RxLRs have been described as a rapidly evolving superfamily whereby each member is related and shares a common ancestor (31). They have very conserved N-terminals and divergent C-terminals although conserved WY folds are observed in some (8, 32). Genes encoding RxLR effectors are mainly found in gene-sparse regions of the genome which contains a high frequency of transposons (33). This could account for the rapid evolution of RxLR effectors. Large expansions of RxLR effectors have been observed in *Phytophthora* species, with some species reportedly encoding several hundred putative RxLRs (8, 26).

Other well characterised oomycete effectors include Crinkler proteins (CRNs), named for their crinkling and necrosis-inducing activity, are composed of a highly conserved N-terminal domain containing a signal peptide and a "LxLFLAK" motif which mediates translocation into the host cell (34). The end of the N-terminal is marked by a highly conserved "HVLVxxP" motif, which separates the N-terminal and C-terminal. They are modular, rapidly evolving proteins that consist of a diverse collection of C-terminal domains (8, 23). The CRNs of some oomycetes carry a modified version of the "LxLFLAK" motif (15). CRNs have been reported to localise to, target and accumulate

in host nuclear components (34, 35). In comparison to RxLRs, CRNs are thought to be a more ancient class of cytoplasmic effectors as they have been found to be distributed across a wide range of oomycete orders (23) including Albuginales (15, 16), Peronosporales (8, 26, 36) and Pythiales (37, 38).

In this study, we catalogue the effector repertoire amongst 590,896 protein coding genes from 37 publicly available genome sequences for the oomycete class, including *Albugo*, *Aphanomyces*, *Hyaloperonospora*, *Phytophthora*, *Phytopythium*, *Pilasporangium*, *Plasmopara*, *Pythium* and *Saprolegnia* species (**Table 1**). Numerous bioinformatic techniques were employed to identify and catalogue putative proteins which may be involved in pathogenesis. A mix of network and phylogenetic methods were utilised to analyse their evolutionary history. Our results have identified novel effector families that appear to be unique to particular oomycete lineages, including *Ap. astaci* proteins which might have the potential to cleave immunoglobulin. Consistent with previous studies, we have identified a significant expansion of effectors in *Phytophthora* species, including glycoside hydrolases and necrosis-inducing proteins. We have detected expansions of chitin degrading enzymes in *Ap. astaci* and *Py. oligandrum*. We have also identified multiple type 2 necrosis-inducing proteins in a number of oomycete species.

Table 1. Taxonomic and genomic information for the 37 oomycete species analysed here[a]

Species	Clade	Order	Host	Size (Mb)	Genes	Candidate Effectors	References
Al. candida	n/a	Albuginales	Plants	34.40	13,310	162	(15)
Al. laibachii	n/a	Albuginales	Plants	32.76	13,804	143	(16)
H. arabidopsidis	n/a	Peronosporales	Plants	78.89	14,321	292	(36)
Ph. agathidicida	Clade 5	Peronosporales	Plants	37.23	14,110*	573	(39)
Ph. capsici	Clade 2	Peronosporales	Plants	64.00	19,805	755	(40)
Ph. cinnamomi	Clade 7	Peronosporales	Plants	53.96	12,942*	459	(39)
Ph. cryptogea	Clade 8	Peronosporales	Plants	63.80	11,876*	430	(41)
Ph. fragariae	Clade 7	Peronosporales	Plants	73.68	13,361*	268	(42)
Ph. infestans	Clade 1	Peronosporales	Plants	228.50	17,797	1,249	(8)
Ph. kernoviae	Clade 10	Peronosporales	Plants	43.00	10,650	329	(43)
Ph. lateralis	Clade 8	Peronosporales	Plants	43.17	11,635	400	(44)
Ph. multivora	Clade 2	Peronosporales	Plants	40.32	15,006	637	(39)

Table 1. Continued.

Species	Clade	Order	Host	Size (Mb)	Genes	Candidate Effectors	References
Ph. nicotianae	Clade 1	Peronosporales	Plants	76.50	10,521	559	(45)
Ph. parasitica	Clade 1	Peronosporales	Plants	82.39	27,942	983	INRA-310
Ph. pinifolia	Clade 6	Peronosporales	Plants	94.60	19,533*	358	(41)
Ph. pluvialis	Clade 3	Peronosporales	Plants	53.62	18,426*	632	(39)
Ph. pisi	Clade 7	Peronosporales	Plants	58.81	15,495*	596	PRJEB6298
Ph. ramorum	Clade 8	Peronosporales	Plants	65.00	15,743	762	(26)
Ph. rubi	Clade 7	Peronosporales	Plants	74.65	15,462*	311	PRJNA244739
Ph. sojae	Clade 7	Peronosporales	Plants	82.60	26,584	1,181	(26)
Ph. taxon totara	Clade 3	Peronosporales	Plants	55.58	16,691	585	(39)
Pl. halstedii	n/a	Peronosporales	Plants	75.32	15,469	435	(14)
Pl. viticola	n/a	Peronosporales	Plants	74.73	12,048*	331	PRJNA329579
P. vexans	n/a	Peronosporales	Plants	33.90	11,958	265	(38)

Table 1. Continued.

Species	Clade	Order	Host	Size (Mb)	Genes	Candidate Effectors	References
Pi. apinafurcum	n/a	Pythiales	Plants	37.58	13,184*	388	PRJDB3797
Py. aphanidermatum	Clade A	Pythiales	Plants	34.26	12,312	348	(38)
Py. arrhenomanes	Clade B	Pythiales	Plants	44.67	13,805	354	(38)
Py. insidiosum	Clade C	Pythiales	Animals	53.23	19,290	558	(46)
Py. irregulare	Clade F	Pythiales	Plants	42.97	13,805	357	(38)
Py. iwayamai	Clade G	Pythiales	Plants	43.20	14,875	330	(38)
Py. oligandrum	Clade D	Pythiales	Fungi	35.90	14,292*	413	(47)
Py. ultimum var. sporangiiferum	Clade I	Pythiales	Plants	37.70	14,096	255	(38)

Table 1. Continued.

Species	Clade	Order	Host	Size (Mb)	Genes	Candidate Effectors	References
Py. ultimum var. ultimum	Clade I	Pythiales	Plants	42.80	15,323	386	(37)
Ap. astaci	n/a	Saprolegniales	Animals	75.84	26,259	582	APO3
Ap. invadans	n/a	Saprolegniales	Animals	71.40	20,816	290	9901
S. diclina	n/a	Saprolegniales	Animals	62.88	18,229	378	PRJNA168273
S. parasitica	n/a	Saprolegniales	Animals	53.13	20,121	385	(48)

[a]Protein counts manually generated from assembly data are marked with an asterisk. References are to the genome publications where possible or otherwise to NCBI BioProject identifiers or the Broad Institute strain identifier. Adapted from reference 11.

Results and Discussion

The Oomycete Secretome

Oomycete pathogens secrete effector proteins and degradative enzymes to facilitate host colonisation through altered physiology (38). Using the proteome data available we undertook an *in-silico* prediction analyses to determine the number of proteins in each species that may be secreted. Our dataset consisted of the predicted proteomes of 37 oomycete species. This included 18 *Phytophthora*, 8 *Pythium*, 2 *Albugo*, 2 *Aphanomyces*, 2 *Plasmopara*, 2 *Saprolegnia*, 1 *Hyaloperonospora*, 1 *Pilasporangium* and 1 *Phytopythium* species (**Table 1**).

Proteins predicted to contain signal peptides were located using SignalP v3. SignalP v3 was chosen over earlier and later versions of the software as previous studies have found v3 the most sensitive in detecting oomycete signal peptides (49). Proteins that contained a transmembrane domain after the signal peptide cleavage site were discarded as these proteins are not likely to be secreted and instead retained in the plasma membrane.

A previous analysis of 13 stramenopiles (including 11 species in our dataset) by Adhikari *et al* , found that between 6.19% to 10.34% of the proteins in each species were secreted (38). Our analysis showed that of the 590,896 proteins tested, 5.25% (30,996) are predicted to be secreted, from a relative low of 2.11% (291) in the obligate biotroph *Al. labiachii* to a relative high of 7.93% (834) in the necrotroph *Ph. nicotianae* (**SupplTable_1**). We observed that in all cases, the percentages of secreted proteins in our analysis differs from what was observed in the analysis of Adhikari *et al*. These differences cannot be accounted for by differences in the number of proteins per species as these were consistent between both analyses as was the methodology used. For example, Adhikari *et al* predict that 10.34% of the *Ph. ramorum* genome is secreted

compared to 7.65% in our analysis (**SupplTable_1**). Similarly, there are large discrepancies between the percentages of secreted proteins for *Ph. sojae* (10.24 % *vs.* 6.24%), *Py. irregular* (6.95% *vs.* 5.03%) and the largest discrepancy was seen between the comparisons for *H. arabidopsidis* (8.81% *vs.* 3.68%). For transparency all scripts used to annotate secreted proteins are publicly available (see methods).

Our results show there is significant difference (chi-square test $p < 0.01$) in the number of predicted proteins between two of the closely related Saprolegniales species (*Ap. astaci* and *Ap. invadans*) where 4.26% (1,119) of the *Ap. astaci* proteome is predicted to be secreted compared to 2.93% (609) in *Ap. invadans* (**SupplTable_1**). The predicted proteome for both of these species differs in size (26,529 *vs.* 20,816 predicted proteins) indicating that an expansion in secreted proteins is partially responsible for the difference observed in the size of the proteome.

Overall, the correlation between the number of predicted secreted proteins and the overall number of proteins per species shows a moderate ($R^2 = 0.455$) and significant correlation ($p < 0.00001$, Pearson correlation test). The correlation between predicted secreted proteins and genome size is weak however ($R^2 = 0.1065$) and not significant ($p > 0.05$).

Secretome Enrichment Analysis

We have documented which biological functions are enriched in the predicted secretomes of individual species (**Figure 2**). This was achieved by comparing the frequency of GO terms and Pfam domains in the secretome compared to the non-secreted proportion of the proteome using Fisher's exact test corrected for false discovery rate (FDR) (50). InterProScan was used to functionally annotate all proteins with GO terms and Pfam domains (51).

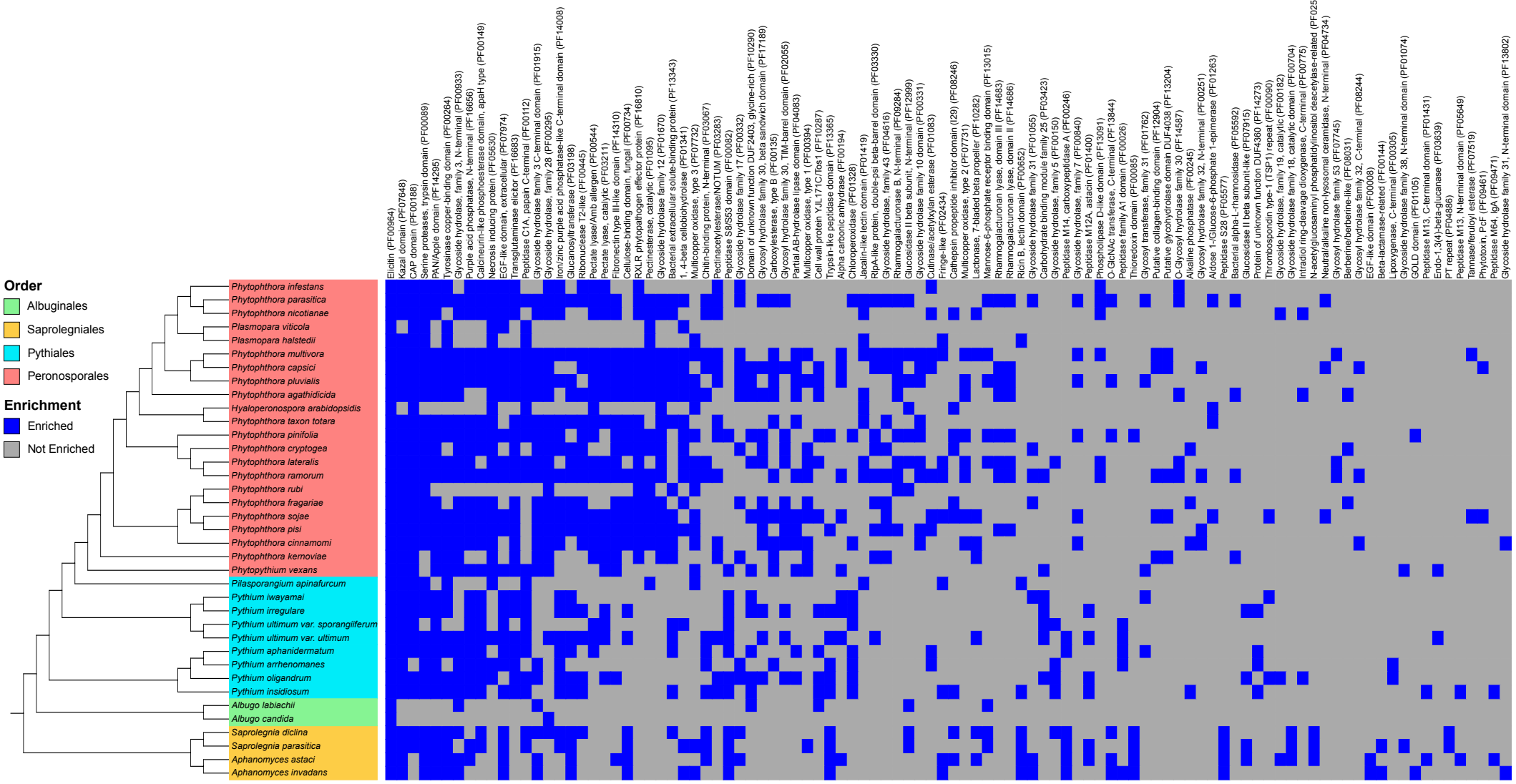


Figure 2. Heatmap of enriched Pfam terms in oomycete secretomes. Terms are ordered on the number of the times they are observed. Only terms enriched in the secretome of two or more species are shown. Terms that are statistically enriched are coloured blue in the heatmap. Grey colouring indicates the term is not enriched.

Comparing the putative secretomes of the 37 species in this analysis we see that the number of Pfam domains enriched in the secretome relative to the non-secreted portion of the proteome varies from a low of 2 domains in *Al. candida* to a high of 56 in *Ap. astaci* (**SupplTable_2**). The enrichment analysis shows that in all 37 species, the Elicitin Pfam domain (PF000964) is enriched (**Figure 2** and **SupplTable_2**). Elicitins are structurally conserved extracellular proteins in *Phytophthora* and *Pythium* species (37). They are known to bind lipids and sequester sterols from plants, thereby overcoming the inability of *Phytophthora* and *Pythium* species to synthesize sterols (52). Similarly, the serine protease inhibitor, Kazal-like domain (PF07648) is enriched in 33 of the 37 species (**Figure 2** and **SupplTable_2**). The Kazal-like domain has been implicated in the infection process of *Phytophthora* species and act as apoplastic effectors (24). Another widely distributed domain that is enriched in a large number of oomycete secretomes (33 of 37) is the cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 proteins (CAP) (**Figure 2** and **SupplTable_2)**. CAP-domain-containing secreted proteins are produced by non-vertebrate eukaryotes and prokaryotes and have been implicated in both virulence and immunity functions (53, 54), however currently little is known about the molecular mode of action of such proteins. Of the 37 species analysed, 31 show enrichment for chymotrypsin (PF00089) which most probably has a role to play in extracellular proteolysis. Similarly, 30 of the 37 species show enrichment for the PAN/Apple domain (PF14295). Previous analyses have shown that carbohydrate-binding module (CBM) containing proteins that recognise and bind saccharide ligands from *Ph. parasitica* are associated with two PAN/Apple domains. The PAN/Apple domain is known to interact with both protein and carbohydrates (55). Knockdown of *Ph. parasitica* CBM affects adhesion to cellulose substrates including plant cell walls (56). Domains involved in the possible degradation of plant cell walls were also found to be enriched in

the secretomes of many of the species investigated. For example, pectin degradation (PF00544, PF03283 and PF03211), Glycoside hydrolases (PF00933, PF00915, PF00295, PF17189, PF02055) and Cellulose binding (PF00374) are all enriched across a range of species (**Figure 2** and **SupplTable_2)**. Well known effectors including necrosis inducing protein (PF05630) is found to be enriched in *Phytophthora* and Pythiales species while the RxLR phytopathogen domain (PF16810) is enriched in the secretome of the majority of *Phytophthora* species (**Figure 2** and **SupplTable_2)**.

With respect to GO term secretome enrichment within our dataset, we observe results that corroborate our Pfam enrichment analysis. For example enriched GO terms associated with plant cell wall degradation such as pectin activity (GO:0030570 & GO:0030599), cellulose activity (GO:0030248 & GO:0008810), polygalacturonase activity (GO:004650), glucan catabolism (GO:0009251), cellulose catabolism (GO:0030245), cellulose metabolism (GO:0030243) and carbohydrate binding (GO:0030246) are widely distributed. Cell wall organisation (GO:0071555), modification (GO:0042545) and biogenesis (GO:0071554) are all enriched in the majority of *Phytophthora* species (**SupplTable_2** and **Suppl_Figure1**). Hydrolase activity acting on glycosyl bond (GO:0004553, GO:0016787 and GO:0016798) are found to be enriched in all 37 species (**SupplTable_2** and **Suppl_Figure1**). Other terms that are ubiquitously enriched include defence response (GO:0006952), pathogenesis (GO:0009405), interspecies interactions (GO:0044419) and multiorganism process (GO:0051704) (**SupplTable_2** and **SupplFigure_1**).

Previous researchers have undertaken similar analyses on some of the species within our dataset including *Ph. infestans* (57) and six *Pythium* species (38). We found broad agreement between our results and those previously reported for the *Ph. infestans* analysis. For example, we also observed enrichment in GO terms associated with

carbohydrate metabolic processes (GO:0005975 and GO:0016052), sugar metabolism (GO:0006040), sugar binding (GO:0030246), sugar modification (GO:0008810, GO:0004650, GO:0030570 and others), pathogenesis (GO:0009405), defence response (GO:0006952), proteolysis (GO:0006508) and serine peptidase activity (GO:0004252 and GO:0008236) (**SupplTable_2**). Similarly, we also observed Pfam domains associated with pectin degradation (PF03283, PF00544 & PF03211), elicitins (PF00964), Kazal-type domains (PF07648) and necrosis inducing protein (PF05630) (**SupplTable_2**). With respect to our results and those previously reported for *Pythium* species the level of agreement is not as strong. We did observe enrichment in GO terms associated with pathogenesis, proteolysis, carbohydrate metabolic process (GO:0006508), hydrolyase activity (GO:0004553) and glycosyl hydrolyase activity (GO:16798) but failed to detect enrichment in terms such a nucleotide binding (GO:000166), integral to membrane (GO:0016021), transmembrane transport (GO:0055085) and RNA processing (GO:0006396) as previously reported (38).

The Oomycete Effector Arsenal

We set out to investigate the abundance of effectors in oomycete species. The list of effectors considered in our analysis (**SupplTable_3**) is based on a number of previous studies which describe pathogenic proteins from oomycete species, including *Plasmopara* (14), *Phytophthora* (26) and *Pythium* (37) species. Any protein identified as having a pathogenicity related domain was classified as a putative effector. These were combined with our secretome analysis to determine if the effector was predicted to be secreted or not. The overall effector content of each species was quantified to detect expansions. The following counts exclude RxLR effectors as they are treated in more depth in a following section.

In total 13,751 proteins were identified as having domains that could be implicated in pathogenicity (**SupplTable_3**). Overall 6,250 (~45%) of these were predicted to be secreted by SignalP (**SupplTable_3**). Our results show that *Phytophthora* proteomes generally contain the highest frequency of effectors, with *Ph. sojae* (766) and *Ph. infestans* (673) possessing the largest arsenals of effectors representing 2.9%, 3.8% of their total proteome respectively (**SupplTable_3**). *Albugo* species were found to contain the smallest number of effectors. A trend was identified whereby hemibiotrophic species, such as *Phytophthora*, typically possess more effectors than saprotrophic species (members of the Saprolegniales order) and necrotrophs, such as *Pythium* species. Obligate biotrophs, including *Al. candida*, *Al. laibachii* and *H. arabidopsidis*, feature the fewest number of effectors, 183, 166 and 215 respectively (**SupplTable_3**). Exceptions to this trend exist, most notably *Ap. astaci* (636), *Py. insidiosum* (534) and *Py. oligandrum* (405) all of which contain large repertoires of effectors compared to other closely related oomycete species with similar lifestyles. For example, *Ap. astaci* was found to contain 635 effectors in comparison to the closely related *Ap. invadans* which contains 311 (**SupplTable_3**). A number of oomycete effectors are discussed in more detail in the following sections.

Necrosis-Inducing Proteins (NLPs)

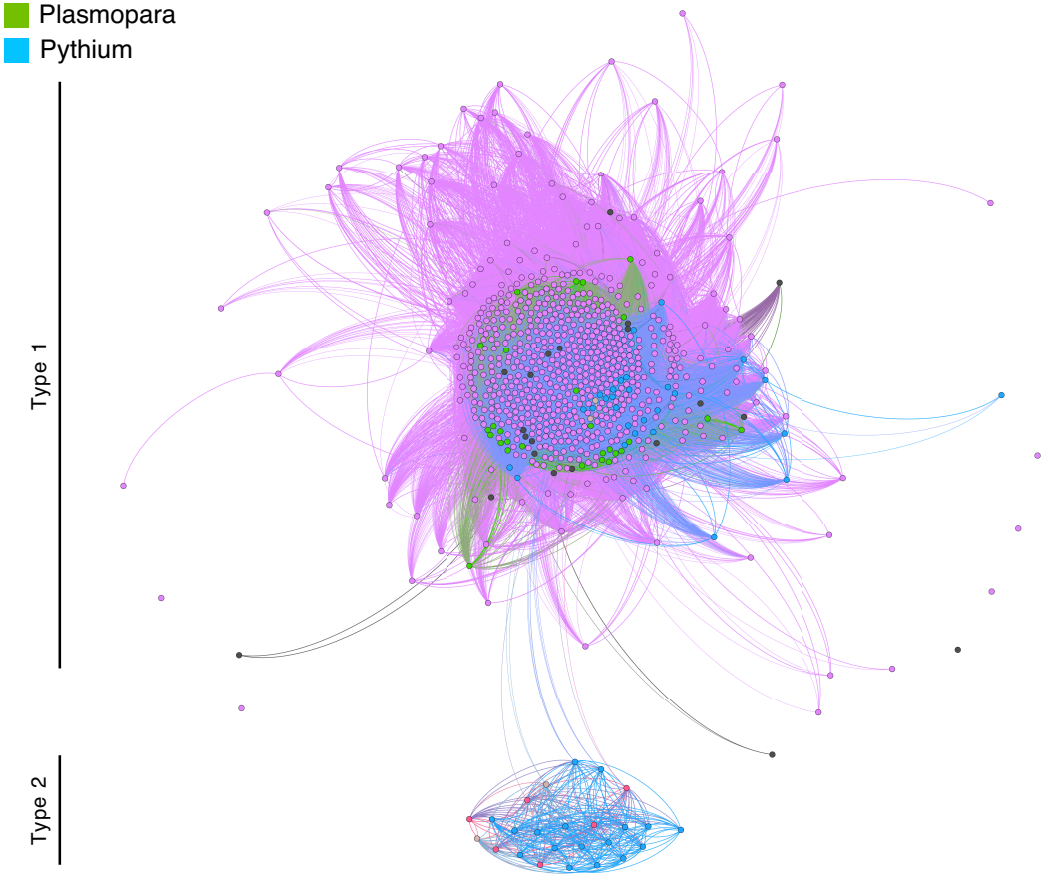
Necrosis-inducing proteins (NLPs) are apoplastic effectors found in bacteria, fungi and oomycetes (58). The mechanisms by which NLPs act are not fully understood but they are known to induce necrosis, trigger ethylene accumulation and elicit immune responses in dicotyledons (59). A number of NLPs have previously been reported to be non-cytotoxic but instead act as microbe associated molecular patterns (MAMPs) which are recognised by the plant hosts and result in the activation of the plant immune system (60).

Our InterProScan analysis detected 771 proteins with signatures of NLPs, 499 (67%) of these NLPs were predicted to contain signal peptides by SignalP (**SupplTable_3**). Our results show that NLPs are absent from *Albugo*, *Aphanomyces* and *Saprolegnia* species but are present in all Peronosporales and Pythiales species. Most *Pythium* species have less than seven copies. *Py. insidiosum* is the only *Pythium* species in our dataset to lack NLPs. NLPs are highly expanded in *Phytophthora* species. In particular, *Ph. ramorum* has 69 copies, *Ph. parasitica* has 74 copies and *Ph. sojae* has 80 copies (**SupplTable_3).**

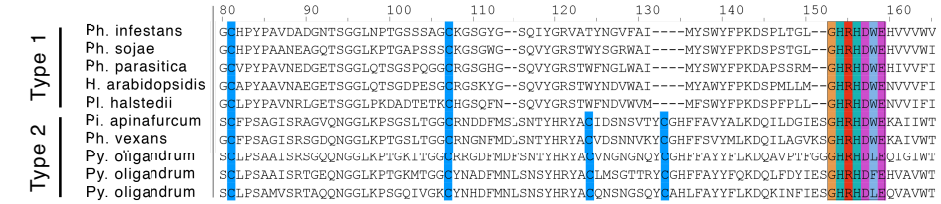
NLPs can be divided into two types: type 1 NLPs and type 2 NLPs (61). Both types share a conserved amino acid motif "GHRHDWE". They are distinguished by the presence of pairs of cysteine residues. Type 1 NLPs have one pair of conserved cysteines, while type 2 NLPs have two pairs of conserved cysteine residues (59). Each pair of cysteines could potentially form disulphide bridges, providing additional stability. Type 1 NLPs are found in bacteria, fungi and oomycetes. Type 2 NLPs have been located in bacteria and fungi but were originally thought to be absent from oomycetes (59, 61). However, work by Horner *et al* has shown that *Py. oligandrum* contains a Type 2 NLP with similarity to a homolog from the proteobacterial plant pathogen *Pectobacterium atrosepticum* (62).

To further investigate the oomycete NLPs in our dataset, we constructed a homology network of all 771 NLPs in our dataset (**Figure 3A**). An interactive version of the network is available online at https://oomycetes.github.io. In the network, the degree of a node is the number of edges it has connecting it to other nodes, therefore the degree is the number of homologs the protein has in our network. The overall NLP network (**Figure 3A**) has an average degree of 522.5, revealing that many of these NLPs are homologous to each other. 740 NLPs are grouped in a large, dense cluster (**Figure 3A)**.

A.


Genus
Hyaloperonospora
Phytophthora
Phytopythium
Pilasporangium
Plasmopara
Pythium
Type 1
Type 2

B.


80 90 100 110 120 130 140 150 160
Type 1
Ph. infestans
Ph. sojae
Ph. parasitica
H. arabidopsidis
Pl. halstedii
Type 2
Pi. apinafurcum
Ph. vexans
Py. oligandrum
Py. oligandrum
Py. oligandrum

C.


Taxonomy
Bacteria
Fungi
Oomycetes
Actinobacteria
Proteobacteria
Saccharopolyspora erythraea
Stackebrandtia nassauensis
Amycolatopsis orientalis
Streptomyces davawensis
Streptomyces hygroscopicus subsp. jinggangensis
Actinoplanes missouriensis
Actinoplanes friuliensis
Streptomyces griseus subsp. griseus
Streptomyces sp. Sirex
Streptomyces fulvissimus
Frankia alni
Streptomyces coelicolor
Streptomyces scabiei
Streptomyces davawensis
Streptomyces albus
Streptomyces venezuelae
Streptomyces pratensis
Streptomyces sp.
Streptomyces rapamycinicus
Streptomyces violaceusniger
Frankia sp.
Streptomyces bingchenggensis
Glomerella graminicola
Fusarium oxysporum f. sp. lycopersici
Podospora anserina
Podospora anserina
Verticillium alboatrum
Verticillium dahliae VdLs
Fusarium graminearum species complex
Nectria haematococca
Fusarium oxysporum f. sp. lycopersici
Gibberella moniliformis
Metarhizium anisopliae
Verticillium dahliae
Verticillium alboatrum
Nectria haematococca
Verticillium alboatrum
Magnaporthe grisea
Magnaporthe grisea
Sporotrichum thermophile
Chaetomium globosum
Thielavia terrestris
Podospora anserina
Neurospora crassa
Neurospora tetrasperma
Neurospora discreta
Sordaria macrospora
Sordaria macrospora
Podospora anserina
Verticillium alboatrum
Verticillium dahliae
Podospora anserina
Trichoderma virens
Drechslera maydis
cochliobolus sativus
Pythium oligandrum (07472)
Pythium oligandrum (07487)
Pythium oligandrum (07488)
Pythium oligandrum (13373)
Pythium oligandrum (13268)
Pythium oligandrum (07485)
Pythium oligandrum (07483)
Pythium oligandrum (07484)
Pythium oligandrum (07489)
Pythium oligandrum (07481)
Pythium oligandrum (07478)
Pythium oligandrum (07476)
Pythium oligandrum (07477)
Pythium oligandrum (07479)
Pythium oligandrum (07482)
Pythium oligandrum (07480)
Teredinibacter turnerae
Dickeya zeae
Dickeya dadantii
Pectobacterium carotovorum subsp. carotovorum
Pectobacterium atrosepticum
Pectobacterium sp.
Pectobacterium wasabiae
Pilasporangium apinafurcum (11165)
Pilasporangium apinafurcum (00472)
Pythium vexans (08006)
Pilasporangium apinafurcum (00473)
Pythium vexans (03426)
Pilasporangium apinafurcum (00474)
Pilasporangium apinafurcum (06227)
Pilasporangium apinafurcum (06228)
Tree scale: 1

Figure 3. Analysis of oomycete necrosis-inducing proteins (NLP). **(A)** Homology network of 771 oomycete NIPs. Identified type 2 NIPs are arranged in a small, strongly connected cluster of 25 proteins from *Phytopythium vexans*, *Pilasporangium apinafurcum* and *Pythium oligandrum*. **(B)** Multiple sequence alignment of five type 1 NLPs and five type 2 NLPs. Both types share a conversed "GHRHDWE" NIP motif. Type 2 NLPs have an additional pair of cysteine residues. **(C)** Maximum-likelihood phylogeny of 87 NLPs containing 29 bacterial proteins, 33 fungal proteins and 25 oomycete proteins. Bootstrap values greater than 50% are shown.

The network included six singletons that did not share significant sequence similarity with any other protein in the network (**Figure 3A**). They included one *H. arabidopsidis*, one *Ph. cryptogea,* one *Ph. kernoviae* and three *Ph. parasitica* proteins. Interestingly, a smaller cluster of 25 NLPs was identified **(Figure 3A)**. Only five of these proteins were homologous to other proteins outside of this cluster, this lack of homology is illustrated by few edges between both clusters **(Figure 3A)**. The 25 proteins included 17 *Py. oligandrum*, 6 *Pilasporangium apinafurcum* and 2 *Phytopythium vexans* putative NLPs. One of the *Py. oligandrum* proteins is an ortholog of the Type 2 NLP (GenBank: EV243877) previously reported by Horner *et al* (62). A multiple sequence alignment of a selection of proteins from both clusters was carried out (**Figure 3B**). Inspection of this alignment revealed no significant sequence similarity between the two clusters except for the presence of the conserved "GHRHDWE" motif and conserved cysteine residues (**Figure 3B**). The NLPs from the large cluster contained two conserved cysteine residues indicating that they are type 1 NLPs. Proteins from the smaller cluster of 25 proteins contained four conserved cysteine residues indicating that they are type 2 NLPs. Apart from the shared NLP motif and conserved cysteine residues, there is no significant sequence similarity between the two types of NLPs.

We set out to further investigate the evolutionary history of the putative oomycete Type 2 NLPs. BLASTp searches of the 25 type 2 NLPs against the NCBI databases revealed top hits with proteobacterial species. To reconstruct the evolutionary history of these proteins we used the 25 Type 2 NLPs as bait sequences in a BLASTp homology search (E value cut-off of 10^{-20}) against a local database of 8,805,033 proteins, with broad taxon sampling across prokaryotes and eukaryotes (63). This search identified 87 homologous proteins, including 25 oomycete proteins (included in our dataset), 33 fungal proteins and 29 bacterial proteins. We aligned and manually edited the resulting

homologs to give a final alignment of 460 amino acids and generated a maximum-likelihood phylogeny with PhyML using the WAG model of substitution **(Figure 3C)**. This phylogeny places all fungal proteins into a single clan (41% Bootstrap support). All oomycete proteins are located in a single clan with 100% bootstrap support (**Figure 3C**). Within this large oomycete containing clan there is also a sister clan of Proteobacteria proteins with 70% bootstrap support (**Figure 3C**). There is also a completely separate Actinobacterial clan with 93% bootstrap support (**Figure 3C**).

The phylogenetic distribution of these Type 2 NLPs is interesting. All of the fungal homologs are from the subphylum Pezizomycotina. Two bacterial phyla are represented, the Actinobacteria and the Proteobacteria. However, the Proteobacterial proteins are inferred to be more closely related to their oomycete homologs than to the Actinobacterial homologs (**Figure 3C**). One possible scenario is that horizontal gene transfer (HGT) has occurred during the evolutionary history of these proteins. However, due to patch phyletic distribution we cannot confidently infer the direction of gene transfer or indeed if HGT has definitely occurred.

Immunoglobulin A Peptidases

Immunoglobulin A peptidases are a family of hydrolytic enzymes that cleave immunoglobulin A (IgA) and have been implicated to be important virulence factors in bacterial infections of humans (64). InterProScan analysis revealed 40 oomycete proteins containing 'IgA peptidase M64' domains (**SupplTable_3**), 21 (53%) of which were predicted to contain signal peptides (**SupplTable_3**). Expansions of these proteins are present in the genomes of both *Ap. astaci* and *Py. insidiosum*. 10 proteins containing this domain were found in *Ap. astaci* (6 with signal peptides) and 6 in *Py. insidiosum* (4 with signal peptides) (**SupplTable_3**). Other *Pythium* species in our dataset contained between one and five copies of the protein. No proteins with this domain were found in any *Albugo*, *Phytophthora*, *Plasmopara* or *Saprolegnia* species (**SupplTable_3**). Therefore, within our dataset, IgA peptidases are unique to *Aphanomyces* species and the Pythiales order.

We aligned and manually curated all 40 IgA peptidase containing proteins using MUSCLE and JalView respectively to give a final alignment of 770 amino acids. A maximum-likelihood phylogeny of all IgA peptidase containing proteins was generated using a WAG substitution model and 100 bootstrap replicates (**Figure 4**). Our phylogenetic reconstruction shows multiple species-specific duplication events which have led to the expansion of IgA peptidases in *Ap. astaci* and *Py. insidiosum* (**Figure 4**). The IgA peptidases are grouped into two separate monophyletic clades. The first contains all Pythiales homologs while the second contains all *Aphanomyces homologs* (**Figure 4**).

Ap. astaci and *Py. insidiosum* are pathogens of crayfish and mammals respectively (17, 65). It is tempting to speculate that these proteins possibly provide a defence mechanism for these species, allowing them to suppress the immune response of their animal hosts. However, it should be noted that a number of *Pythium* species with plant

hosts were also shown to contain multiple copies of IgA peptidases, including five in *Py. iwayami* and four in both *Py. arrhenomanes* and *Py. ultimum var. sporangiiferum* **(SupplTable_3).**

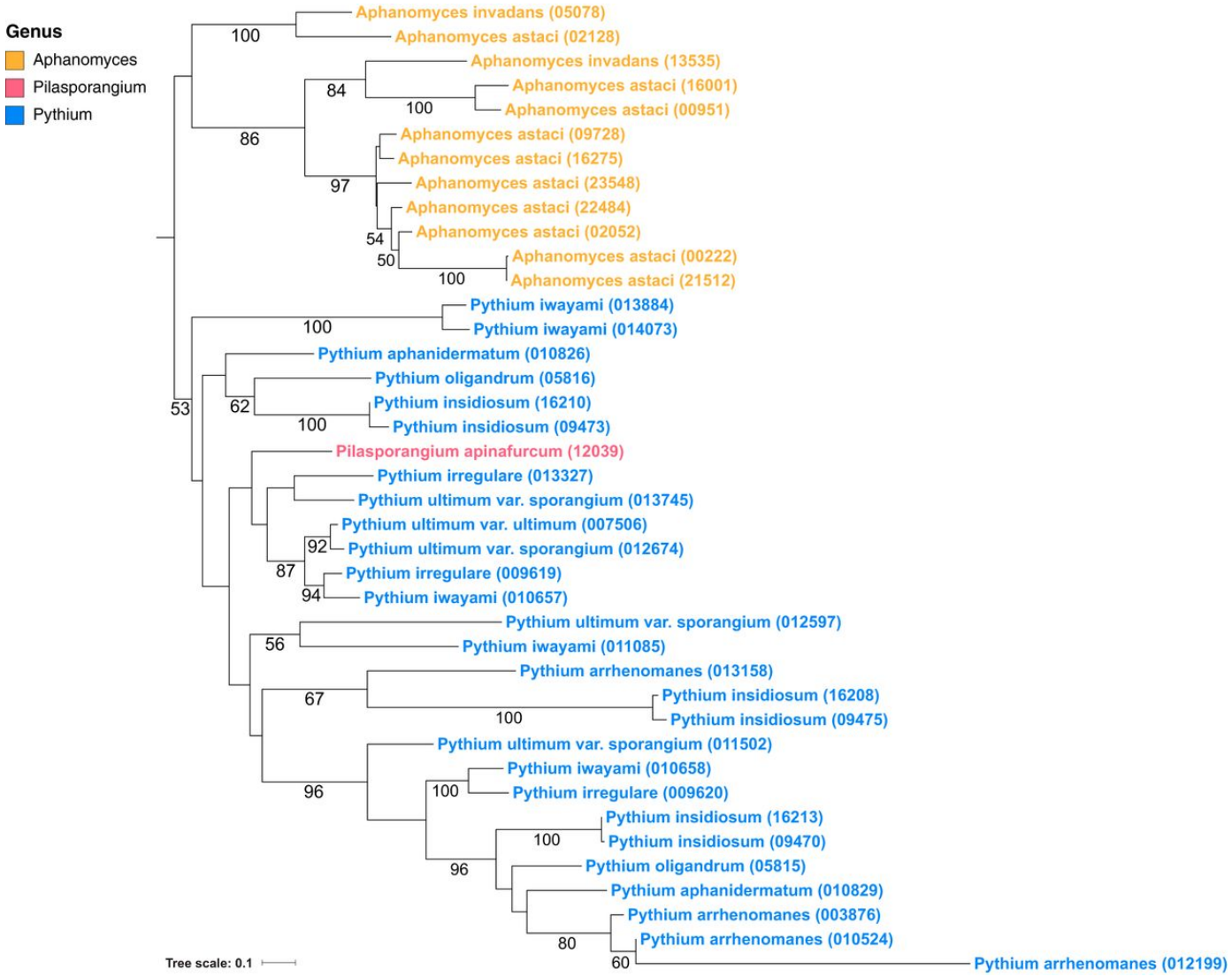


Figure 4. Maximum-likelihood phylogenetic reconstruction of oomycete IgA peptidases. Bootstrap values greater than 50% are shown.

Glycoside Hydrolases

Glycoside hydrolases (GH), or glycosyl hydrolases, are hydrolytic proteins that cleave glycosidic bonds in complex sugars. They may be involved in the breakdown of cellular components, for example, the cell membrane and cell wall. We set out to investigate the extent of expansions within GH families across oomycete species. A total of 4521 proteins with GH domains were located in our InterProScan analysis (**SupplTable_3 & SupplTable_4**). Expansions of GHs have been observed in oomycetes previously (38, 48). Identified GHs were distributed across 47 different GH families. SignalP predicted 1928 (42.6%) GHs to be secreted. *Phytophthora* species possess the highest number of GH proteins, with an average of 162, followed by *Aphanomyces* and *Saprolegnia* species, with averages of 106 and 98 respectively (**SupplTable_4**). *Al. candida* and *Al. laibachii* feature smaller sets of GH proteins with less than 60 each.

A number of GH families were found to be species or genus specific. For example, GH families 4, 26, 29, 42 and 48 are restricted to *Ph. kernoviae* while GH families 8, 24 and 44 are unique to *Ph. rubi*. Similarly, GH family 36 is found only in *Ph. kernoviae* and *Ph. rubi*. GH families 20 and 27 are found only in members of the Saprolegniales order. GH families 18, 19 and 48 have known chitinase activity (66). Expansions of these families were revealed in *Ap. astaci*, *Py. oligandrum* and *Saprolegnia* species. They are investigated in more detail below **(SupplTable_4)**.

Chitinases

Chitin is the second most abundant biopolymer in nature and is an important structural component of invertebrate exoskeletons and fungal cell walls (67). Chitinases are enzymes that degrade chitin. In fungi, chitinases play an important role in hyphal growth, spore germination and cell wall remodelling (68). Oomycete cell walls, however, contain

no or very small amounts of chitin (69). Thus, it would appear that chitinases encoded by oomycete species play roles primarily in the degradation of exogenous chitin, for example, the breakdown of chitin in host cell walls.

Several GH families are known to have chitinase activity (66). These include GH families 18, 19, 23 and 48. A total of 281 proteins in our dataset were reported to have putative chitinase activity in our InterProScan analysis, 112 (40%) of these are predicted to be secreted **(SupplTable_3)**. The chitinases identified were spread across three GH families, 18, 19 and 48. No members of GH family 23 were located in our dataset. Only one protein was identified as a member of GH family 48, this belonged to *Ph. kernoviae*. However, it was not predicted to be secreted.

A large expansion of 50 chitinases were detected in *Ap. astaci* (**SupplTable_3**), this is significantly more than any other oomycete, 38 of these proteins belong to GH family 18 and the remaining 12 belong to GH family 19 (**SupplTable_4**). *Ap. astaci* is a pathogen of crayfish, growing on and within the crayfish cuticle (65), which is composed of chitin. This expansion of chitinases may reveal a successful adaption of *Ap. astaci* to its crayfish host, allowing the pathogen to penetrate through the chitin layers of crayfish cuticles. In agreement with previous work (48), our analysis also revealed a large number of chitinases in the genome of *S. diclina* and *S. parasitica* (25 and 23 respectively) (**SupplTable_3**). Additionally, large numbers of chitinases were located in *Al. laibachii* and *Py. oligandrum*. 16 chitinases were identified in *Al. laibachii* and 15 in *Py. oligandrum* (**SupplTable_3**). This is particularly significant for *Py. oligandrum* which is a pathogen of fungi (18). Some fungal cell walls are composed of up to 20% chitin (70). Possessing a large repertoire of chitinase enzymes may facilitate the pathogen to breakdown host fungal cells. Other oomycetes in the dataset typically had less than 10 chitinases.

We constructed a homology network of all 281 oomycete chitinases in our dataset to investigate their evolutionary history (**SupplFigure_2**). An interactive version of the network is available online at https://oomycetes.github.io. In the networks, nodes represent proteins and edges represent sequence similarity between two proteins. Our network consists of a number of disconnected clusters (no edges/homology to other clusters). We identified one singleton (a node with no edges, i.e. a protein with no homologs) which was the GH family 48 protein belonging to *Ph. kernoviae*. All GH family 19 proteins were placed into a single, strongly connected cluster (**SupplFigure_2**). No *Albugo* proteins were found in this cluster. GH family 18 proteins were divided into five clusters. Saprolegnia proteins dominate the network, in particular *Ap. astaci* whose proteins make up almost 18% of the network. Thus, our network analysis shows that oomycete chitinases are divided into a number of subclasses that are not homologous. This is consistent with the division of chitinases into 5 classes (71).

Our analysis also revealed 31 *Ap. astaci* proteins with N-terminal chitin-binding domains (**SupplTable_3**), 25 (81%) of these were predicted to be secreted (**SupplTable_3**). Other oomycetes contained less than 10 of these proteins, with *Al. candida*, *Al. laibachii*, *Ap. invadans* and *Ph. rubi* containing no proteins with this domain. BLAST searches of these proteins against the NCBI databases did not reveal any hits with annotated proteins, therefore their function is unknown, it is possible that the N-terminal domain facilitates attachment to the chitin exoskeleton of its crayfish host.

Proteases

A large number of proteins with hydrolytic activity were reported, including glycoside hydrolases, proteases and pectin modifying proteins (**SupplTable_3**). These are thought to be involved in the degradation of host cells (8). Several protease families were found

to make up a large part of oomycete secretomes including aspartyl proteases, papain family cysteine proteases, subtilases, trypsins and trypsin-like proteases. Members of the Saprolegniales order contained the most proteases, with each member possessing over 150 proteins with predicted protease activity **(SupplTable_3**). *Pythium* species also had a large number of proteases, with an average of 139 proteases, higher than the *Phytophthora* average of 88 (**SupplTable_3**). A large number of these proteases were reported to be secreted in our SignalP analysis, indicating that they may be involved in the breakdown of host cells. For example, *Saprolegnia* species have an average of 86 secreted proteases*, Pythium* species have an average of 55 and *Phytophthora* have an average of 24 (**SupplTable_3**). Obligate biotrophs such as *Al. laibachii*, *Al. candida* and *H. arabidopsidis* feature the fewest proteases, with each species possessing less than 60 proteases each. Expansions of aspartyl proteases were detected in *Py. insidiosum* (95), *Ph. ramorum* (70) and *Ph. sojae* (68). *Ap. astaci* contains 161 proteins with predicted trypsin or trypsin-like activity, the largest number of any species in our dataset. SignalP predicted 93 (58%) of these to be secreted (**SupplTable_3**). Every other oomycete had less than 61 copies. Our analysis also identified an expansion of proteins with subtilase domains in *Aphanomyces*, *Pythium* and *Saprolegnia* species. Again, *Ap. astaci* contains a large repertoire of 94 proteins with subtilase domains compared to *Phytophthora* species, all of which has less than 15 subtilases (**SupplTable_3**). Our results show that the *Ap. astaci* proteome harbours a vast number of proteases, significantly more so than any other oomycete. The large arsenal of hydrolytic enzymes in *Ap. astaci*, and other Saprolegniales members, may play an important role in the degradation host cells.

Oomycete Pectin Modifying Proteins

Pectin is a major component of plant cell walls, making up to 35% of primary walls in higher plants, and plays important roles in plant defence, development and growth (72). A total of 1048 oomycete proteins were found to contain domains that are involved in the modification or degradation of pectin, including 693 pectate lyases, 226 pectin esterases and 129 pectin acetylesterases (**SupplTable_3**). Pectate lyases are involved in the cleavage of pectin and result in fruit softening and rot *via* degradation of the plant cell wall (73). Proteins with pectate lyase domains were abundant in *Phytophthora* and *Pythium* species (**SupplTable_3**). They are completely absent in *Al. laibachii*, *Ap. astaci* and *Ap. invadans*. One copy was found in each of *Al. candida*, *S. diclina* and *S. parasitica* but the *S. diclina* copy was not predicted to be secreted. Pectinesterases, or pectin methylesterases, catalyse the de-esterification of pectin and are used by plants in a wide range of biological processes including cell wall remodelling, fruit ripening, pollen growth and root development (74). However, they can also be exploited by pathogens to invade plant tissues (75). Our results show that pectin esterases are present only in members of the Peronosporales order in our dataset (**SupplTable_3**). The majority of these proteins contain signal peptides (**SupplTable_3**), indicating that they are effectors that potentially cause damage to the cell walls of their hosts. Similarly, pectin acetylesterases can be exploited by pathogens to catalyse the deacetylation of pectin, making the pectin backbone more accessible to pectin-degrading enzymes such as pectate lyases (76). Proteins with the pectin acetylesterase domain were found in *Phytophthora*, *Phytopythium, Pilasporangium*, *Plasmopara* and *Pythium* species (**SupplTable_3**). Species with more pectate lyases typically had more pectin acetylesterases this correlation is strong and significant ($R^2 = 0.600$, $p < 0.00001$, Pearson correlation test).

Cutinases

Cutin is one of the main components of plant cuticles (77). The plant cuticle acts as a physical barrier, allowing plant cells to tolerate external environmental stresses and also protect interior plant tissues from invading pathogens (78). Cutinases are extracellular enzymes that hydrolyse cutin and have been identified in bacteria, fungi and oomycetes (79). They can be used by plant pathogens to compromise the structural integrity of the plant cuticle, allowing them to penetrate and infect inner plant tissues. Enzymatic digestion of cutin has been proven to be an essential initial step in the infection process of plant pathogens (80).

Overall, we have identified 122 proteins with cutinase domains in our dataset, 79 (65%) of which were predicted to be secreted (**SupplTable_3**). Cutinases were completely absent in *Aphanomyces* and *Saprolegnia* species. *Pythium* species have previously been reported to lack cutinases (81). We found this to be true for the majority of *Pythium* species in our dataset, however we identified nine cutinases in *Py. aphanidermatum* and seven cutinases in *Py. arrhenomanes* (**SupplTable_3**). Both *Albugo* species in our dataset also contained multiple copies. *Phytophthora pinifolia* was the only Peronosporales member to lack cutinases.

Toxin Families

We identified a number of toxin families in our dataset including necrosis-inducing proteins (see previous section) and the phytotoxic protein family (PcF). Relative to other effectors, the overall number of PcFs detected in our dataset was low. PcFs are known to induce necrosis (82). In total 31 proteins with signatures of PcF proteins were identified and were unique to several members of the Peronosporales, 19 (61%) of these were predicted to be secreted (**SupplTable_3**). *Phytophthora infestans* contains 14 PcF

proteins, 5 of which are predicted to be secreted (**SupplTable_3**). *Ph. capsici* contains six PcF proteins, *Ph. sojae* has four, *Ph. parasitica* has two and *Ph. multivora*, *Ph. pisi* and *Ph. ramorum* have one copy each. All of these are predicted to be secreted (**SupplTable_3**). The proteome of *Ph. lateralis* was also reported to have one PcF protein but it is not predicted to be secreted. *Hyaloperonospora arabidopsidis* was the only non-*Phytophthora* species reported to have a PcF protein (**SupplTable_3**). However, it was not predicted to be secreted. This finding suggests that PcF proteins are unique to *Phytophthora* and closely related species, suggesting that they may have arisen in the last common ancestor of the Peronosporales order.

Crinklers

A combination of regular expression searches and hidden Markov models were utilised to identify Crinkler effectors (CRNs). After manual inspection and removal of suspected false positives, a total of 899 CRNs were identified in our dataset. Our results highlight that *Phytophthora* species possess large expansions of CRNs, more so than other genera. In particular, we found 91, 92 and 167 CRNs in *Ph. capsici*, *Ph. sojae* and *Ph. infestans*, respectively (**SupplTable_3**). *Pilasporangium apinafurcum* and *Pl. viticola* also feature large numbers of CRNs relative to other species, having 52 and 65 CRNs each respectively. CRN numbers were sparse in the Albuginales and Saprolegniales orders, with some species only possessing one copy (*Ap. invadans*, *S. diclina* and *S. parasitica*) (**SupplTable_3**). Thus, it would appear that CRNs play an important role in plant infection for *Phytophthora* species. Only 177 CRNs were predicted to be secreted in our SignalP analysis (**SupplTable_3**). However, previous reports have indicated that a large number of CRNs may be secreted *via* unconventional protein secretion systems that can't be predicted *in silico* (83).

Oomycete Protease Inhibitors

Plants and plant pathogens are constantly undergoing an evolutionary arms race with one another (84). One such example of this is the production and secretion of proteases by plants to degrade pathogen effectors. To counteract this, plant pathogens, including oomycetes, have coevolved to secrete protease inhibitors (4). These protease inhibitors block the defensive mechanism of plant proteases. We have identified a number of secreted oomycete protease inhibitors. The most abundant were Kazal-type protease inhibitors (507 in total, 355 were predicted to be secreted) and cathepsin propeptide inhibitors (155 in total, 95 were predicted to be secreted) (**SupplTable_3**). Counts of cathepsin propeptide inhibitors were consistent across our oomycete dataset with most species possessing less than 5 copies (**SupplTable_3**). Kazal-type protease inhibitors were more abundant and not evenly distributed, for example expansions were recorded in several species, including *Py. insidiosum* (43 copies), *Ph. sojae* (39 copies) and *Ph. infestans* (33 copies). Most other species contained less than 15 Kazal-type protease inhibitors. *Albugo* species have less than five copies each (**SupplTable_3**).

RxLR Effectors

Numerous *in silico* analyses using varying bioinformatics strategies have been employed in previous analyses to identify candidate RxLR effectors in oomycetes. The most liberal was first described by Win *et al*, where all possible open reading frames are examined for the presence of a signal peptide within the first 30 amino acid residues followed by an RxLR motif between residues 30 and 60 (Win method) (85). Extensions to this method have been developed (27, 85) with searches for a downstream EER motif (or a loose match), the EER motif is present in numerous validated RxLR effectors (regular expression or regex) (27). Hidden Markov Model (HMM) profiles derived from

alignments of RxLR-EER effectors have also been implemented successfully (27). The initial genome analyses that described the *Ph. infestans* RxLR complement utilised all three approaches above as well as additional criteria such as exhibiting sequence homology to previously known RxLRs or belonging to protein families where the majority of proteins are deemed putative RxLRs based on Win, regex or HMM criteria. Depending on the method or combination of methods utilised it is possible to detect a broad range of potential RxLR effectors (8). As RxLR effectors are most abundant and well characterised in *Phytophthora* species (8), a strict *Phytophthora* biased approach was taken to identify candidate oomycete RxLR effectors. To be classified as a putative RxLR effector, proteins or ORFs had to meet one of the following RxLR criteria, Win method, HMMsearch, Regex or Homologous (see methods).

The four RxLR criteria were tested on the predicted proteomes of each of the 37 oomycete species in our dataset. Utilising predicted proteins adds an additional criteria layer but may potentially miss open reading frames that were missed by gene callers during annotation. It is obvious some of the criteria above may detect a large number of false positives. For example, a number of the 563 ORFS counted as putative RxLRs in *Ph. infestans* were found by Blast homology alone and do not contain a RxLR domain, using these as "bait" in a BlastP search against other oomycetes will locate non RxLR domain containing homologs. However, we have noticed that when the RxLR repertoire of model genomes such as *Ph. infestans*, *Ph. sojae* and *Ph. ramorum* are referenced in the literature they normally consider homologs of known RxLRs located *via* BlastP alone as putative RxLRs. Our analysis extends this as we use all proteins from these model genomes as bait sequences in our analysis. We also searched for the presence of repeating sequence motifs termed "W", "Y" and "L" that are found towards the C-terminus of a number of *Phytophthora* cytoplasmic effectors (31). These domains form an alpha-helical

fold known as the WY fold that may provide structural flexibility (32). However, due to the sequence divergence observed in *Phytophthora* RxLRs it seems likely alternative folds most likely exist (86).

Putative RxLRs along with sequence information and the criteria that were satisfied in classifying the proteins as putative RxLRs are listed in **SupplTable_6.** For completeness and to allow comparisons to previous analyses we also searched all putative ORFs that were non-overlapping and longer than 70 amino acids with the criteria above (**SupplTable_7)**. We do not discuss the results of these counts here and instead concentrate on counts related to predicted protein coding genes. The reason for this is to maintain consistency, as our secretome analysis also considered putative protein coding genes and did not look at all possible open reading frames.

In total, 4131 proteins in our dataset matched one or more of the RxLR criteria, ranging from a low of 6 proteins in *Al. candida* to a high of 603 in *Ph. infestans* (**SupplTable_5**). Unsurprisingly, the vast majority (3,600 or 87%) of the 4,131 proteins are located in Peronosporales species (**SupplTable_5**). A homology network of the RxLRs like proteins was generated to investigate the evolutionary relationships within the putative RxLRs (**Figure 5**). An interactive and searchable version of the network is available online at https://oomycetes.github.io. The online version of the network permits users to query the network based on protein ID and retrieve sequence information as well as performing BlastP search against the NCBI database. Furthermore, users can filter proteins based on genus or the RxLR criteria used (Win, regex, HMM or BLAST). Proteins can also be viewed based on the presence or absence of the WY fold.

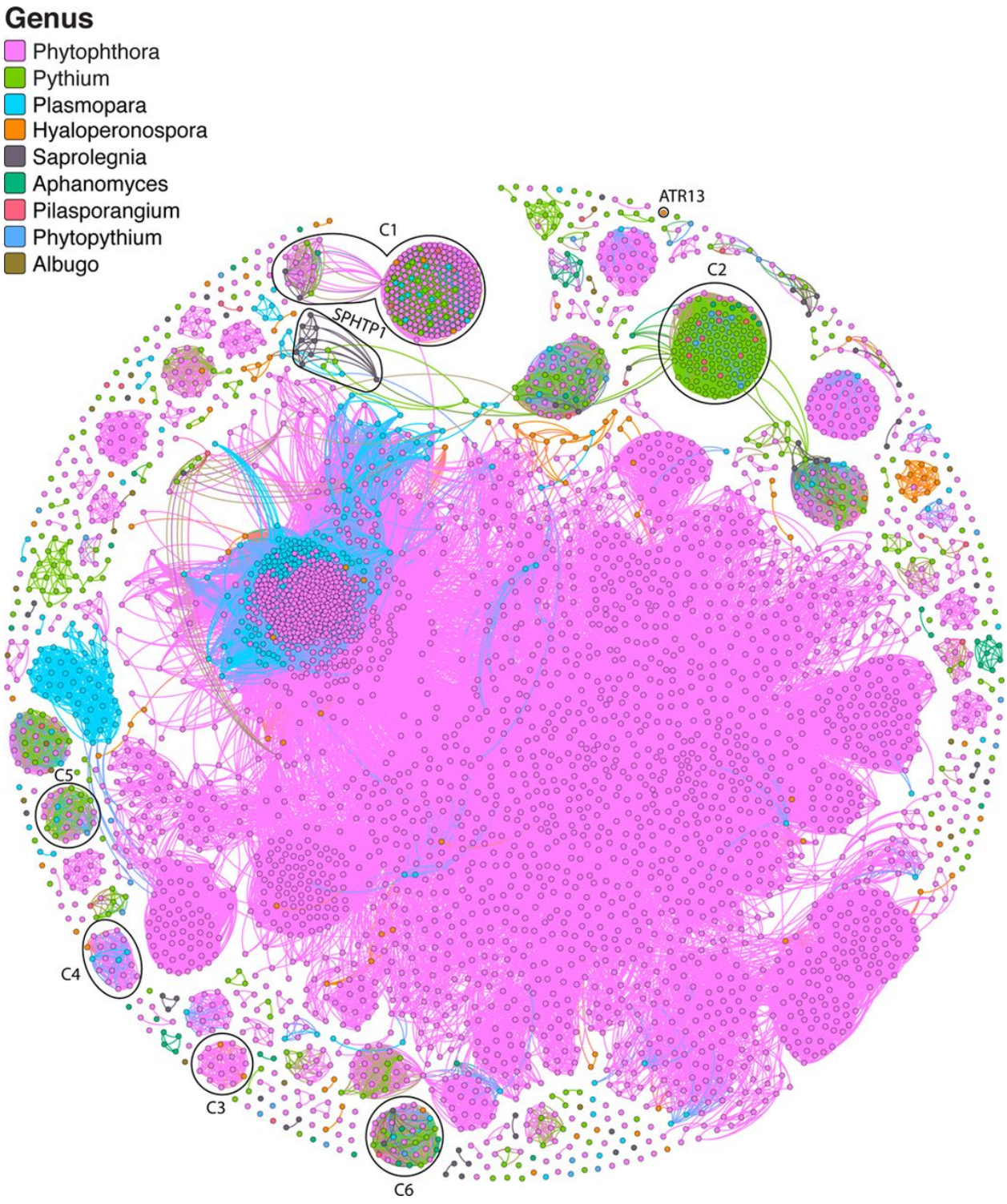


Figure 5. Homology network of 4,131 putative RxLR effectors. Each protein is represented by a node. An edge joining two nodes represents shared sequence similarity between the two proteins (*E* value cut off of 10^{-10}). Many disconnected clusters of RxLR effectors are present, indicating no significant homology between clusters. Nodes are coloured by genus. Clusters discussed in text are labelled. An interactive version of this network is available online at https://oomycetes.github.io/rxlrs.html.

The RxLR network is composed of 4131 nodes, each node corresponds to an individual protein and 184,302 edges, edges link nodes if they are homologous. The putative RxLRs were clustered into 357 connected components (**Figure 5)**. Nodes had an average degree of 89 (i.e. on average, each protein in the network has 89 homologs). The fact that the network is comprised of a number of disconnected clusters (individual clusters of connected nodes that are not connected to other clusters) shows there are no significant sequence similarities between some clusters (**Figure 5**).

A significant proportion (1,852 or 44.8%) of the 4,131 proteins only have one line of evidence labelling them as RxLRs. For example, 1,045 (25.3%) are located based on BlastP homology alone (therefore the RxLR domain is absent), 683 (16.5%) based on the Win method alone, 25 (0.6%) based on the HMM alone and 99 (2.4%) on Regex alone (**Figure 6 & SupplFigure_3**). Of the 1045 proteins that were located by sequence homology exclusively (no evidence of RxLR motif), 981 of these are found in species from the Peronosporales order and may not be functional RxLRs, although 287 of these homologs were found to contain the WY fold (**SupplTable_6**). Conversely 967 (23.4%) of all proteins met the four RxLR criteria and all of these were *Phytophthora* proteins (**Figure 6, SupplFigure_3** & **SupplTable_6**). Furthermore 401 of these 967 proteins also contained the WY fold. Overall of the 4131 proteins, 1123 were found to contain the WY fold and all proteins are from Peronosporales species (**SupplTable_6**).

On closer examination of the similarity network and the criteria responsible for labelling proteins as putative RxLRs, it is obvious that a number of clusters are potentially false positives. For example, C1, contains a number of Peronosporales and Pythiales proteins, excluding a single edge to a large *Phytophthora* cluster, it forms a distinct cluster which shares no similarity to any other cluster in the network. The vast majority of

proteins in this cluster have been labelled putative RxLRs based on BlastP alone (**SupplFigure_3** and **SupplTable_6**).

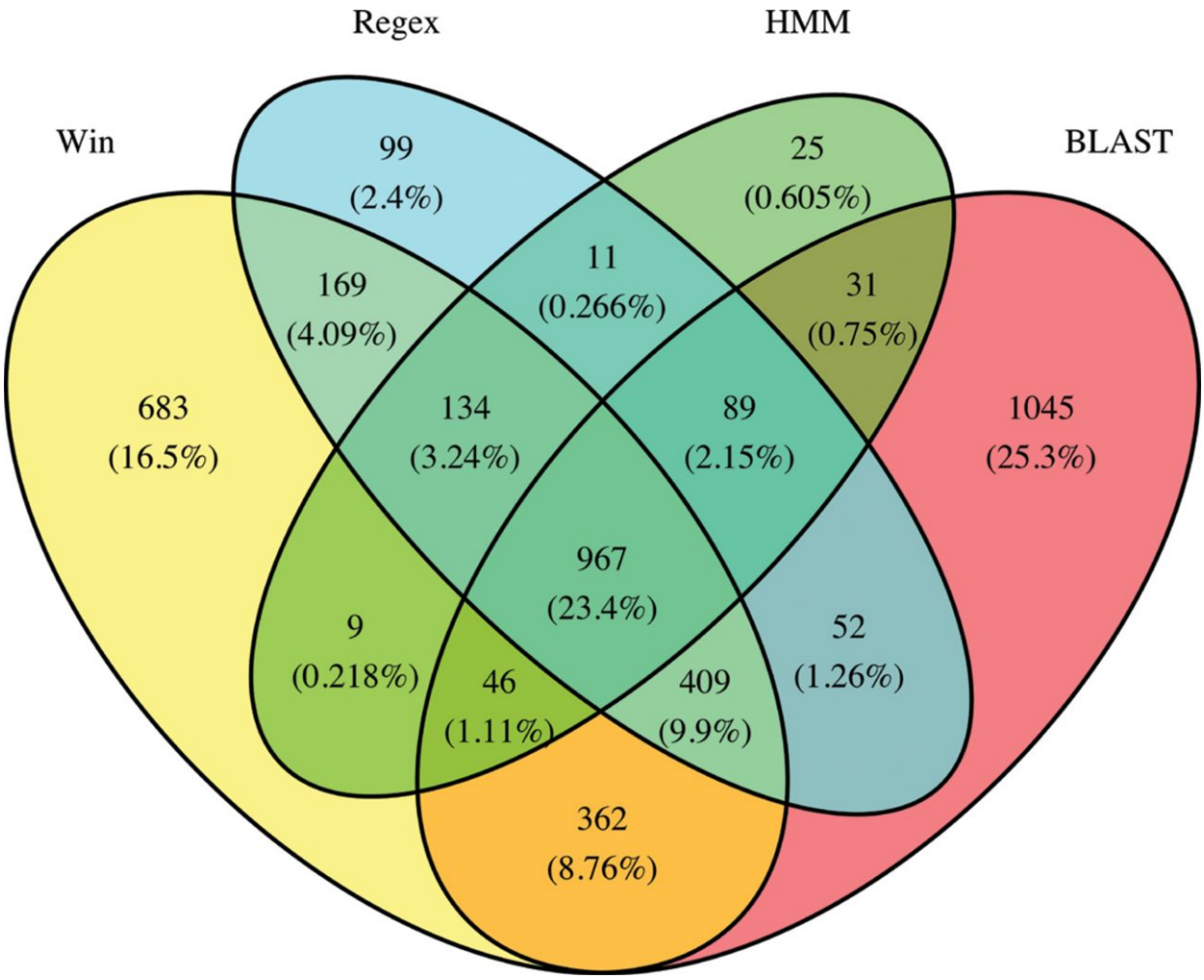


Figure 6. Venn diagram showing the overlap between the 4 RxLR criteria (Win, regex, HMM and Blast) used in this analysis. 967 Phytophthora proteins are labelled as putative RxLRs based on all 4 criteria.

Similarly, C2 is composed of proteins that are also mostly from *Pythium* species (**SupplFigure_3**). All of these proteins have been labelled as RxLR based on the Win method exclusively. C3 is composed of Peronosporales proteins only but all of these have been detected using BlastP alone (**SupplFigure_3**). There are many other examples of potential false positive RxLR clusters including C4 (detected using Win and BlastP) and C5 (detected using Win exclusively) (**SupplFigure_3** and **SupplTable_6**). It is important to note however that the RxLR search performed here, labelled a *S. parasitica* protein (SAPA|10778) as being a putative RxLR. This putative RxLR effector protein (SPHTP1) has been shown to be translocated into fish cells and may play an important role in Saprolegniosis (87). The uptake of SPHTP1is mediated by an interaction with tyrosine-O-sulfate–modified cell-surface molecules (88) and not *via* phospholipids, as is the case for RxLR-effectors from oomycete plant pathogens (89). Examining our RxLR network we see that SPHTP1 lies in a cluster which is composed primarily of *Saprolegnia* proteins with some of these proteins having homology to *Plasmopara* and *Pythium* proteins (**Figure 5)**. *S. parasitica* contains 5 homologs of this protein while *S. diclina* contains 3. Similarly, our RxLR analysis labelled a *H. arabidopsidis* protein (HYAP|12966 or ATR13) as being a putative RxLR again based on the Win method alone. ATR13 is found as a singleton with no homology to any other putative RxLR in our network which is unsurprising as high levels of polymorphism have been reported for this protein. It has been shown that ATR13 has the ability to translocate into host cells (90). Similarly, another previously described RxLR effector from *H. arabidopsidis,* ATR1 (HYAP|01864) was also labelled as a putative RxLR based on Win and Regex criteria (**SupplTable_6)**.

C6 contains 33 RxLRs, with proteins from every genus in our dataset except *Albugo*. All of these proteins have homology to reference RxLRs (**SupplTable_6**) and

are classified RxLRs based on the Win or regex criteria. The group forms a clique (a subnetwork where each member is connected to every other member), showing that every protein in the group is homologous to every other protein. On average these proteins have sequence similarity of 52%. The position of the RxLR motif is ubiquitously conserved (not shown). However, the majority of these proteins contain a KDEL endoplasmic retention (ER) motif at the C-terminal as previously reported for the *Py. oligandrum* protein represented in this cluster (62). Proteins with this ER motif are not predicted to be secreted and are therefore likely false positives (62). Overall 75 putative of the 4131 putative RxLRs contain the KDEL motif (or a variation) (**SupplTable_6**).

Of the 4131 putative RxLRs, 3011 were located based on Win, regex or HMM meaning 1120 have been labelled due to homology to a reference RxLR alone or they may also contain a KDEL retention motif so may not be secreted. Ignoring these 1120 proteins we observe that *Ph. infestans* is predicted to contain 470 RxLRs a figure that is consistent with the 563 putative RxLR figure reported based on ORFs (8). Furthermore 95 of these 563 ORFs were labelled based on homology alone therefore giving an overall number of 468 if non-RxLR containing homologs are excluded. Similarly, the comparison between our *Ph. sojae* RxLR prediction and previous studies is consistent (338 *vs.* 312).

Conclusion

The first oomycete genome sequences were published in 2006 (26) and at the time of writing there are 37 oomycete genome assemblies publicly available. Due to their importance as pathogens of economically important crops and animals, along with the ongoing advances and reductions in costs associated with next generation sequencing technologies, this number is expected to increase dramatically over the coming years. In this analysis we have performed an inventory of known oomycete effectors in all available genome sequences. As well as quantifying their occurrences we have in a number of cases also investigated their evolutionary history.

This genome wide survey provides an up to date inventory of previously described effectors in the Class Oomycota. It is by no means a complete list and we are cognisant that many additional effectors will be characterised in the coming years especially with improved host/pathogen interaction omics studies. However, it does provide the current overview of the arsenal of known effectors in this economically important class of animal and plant pathogens.

We have also examined the presence and absence of glycoside hydrolases and found a diverse range of families across the oomycote class. The majority (54%) of these are secreted. Glycoside hydrolase families 18, 19 and 48 all have chitinase activity. A homology network (**SupplFigure_2**) showed that there is no sequence homology between family members 18 and 19. Interestingly, family 18 can be subdivided into 4 distinct clusters indicating that while family members have the same enzymatic function they do not share sequence similarity and confirm that family 18 has different subclasses. Our analysis has also detected the presence of Immunoglobulin A peptidases in Pythiales species and *Aphanomyces* species (**Figure 4**). Some of these species are animal pathogens

and it is possible that their presence may be important in suppressing the immune response of their animal hosts.

We have also catalogued the putative RxLR repertoire of all 37 oomycete species. Our results are consistent with others in showing that *Phytophthora* species have undergone expansions in these proteins. We are aware that this analysis may be reporting false positives particularly for species outside the Peronosporales order. However, a number of previously described RxLR effectors such as *H. arabidopsidis* ATR1 & ATR13 were detected as was SPHTP1 *S. parasitica*. We are also aware that we may be underreporting the number of RxLRs as we do not analyse all possible open reading frames and instead concentrate on predicted protein coding genes, this decision was taken as our secretome analysis also used protein coding genes. Furthermore, using open reading frames could in itself lead to the reporting of false positives by reporting pseudogenes or non-coding regions of the genome. The confirmation of all putative RxLRs is beyond the scope of this in silico catalogue however detailed sequence information on all proteins is provided in supplementary information.

Materials and Methods

Dataset Assembly

The predicted proteomes for 23 oomycete species were obtained from public databases (**Table 1**). Predicted proteomes for another 14 oomycete species (10 *Phytophthora* species, *Pl. viticola*, *Pi. apinafurcum*, *Py. insidiosum* and *Py. oligandrum*) were generated from publicly available data using AUGUSTUS (91) (**Table 1**). Templates for *ab initio* protein prediction were generated for AUGUSTUS using assembly and EST data from a number of reference oomycete species. *Ph. capsici* was used as a reference for *Ph. agathidicida*, *Ph. multivora*, *Ph. pluvialis* and *Ph. taxon totara*. *Ph. sojae* was used as a reference for *Ph. cinnamomi*, *Ph. cryptogea*, *Ph. fragariae*, *Ph. pinifolia*, *Ph. pisi*, and *Ph. rubi*. *Pl. halstedii* was used as a reference for *Pl. viticola*. *Py. ultimum var. ultimum* was used as a reference for the two *Pythium* species. GeneMark-ES (92) was used in addition to AUGUSTUS for predicting proteins of *Pi. apinafurcum*. The final dataset contained 590,896 proteins from 37 predicted oomycete proteomes (**Table 1**). All proteomes used in this analysis as well as pipelines and scripts for identification of effectors, secretome and RxLRs are available at https://github.com/oomycetes/oomycetes.github.io/tree/master/SupplementaryMaterial.

Identification of Putative Effectors

InterProScan 5 (51) was run on all 590,896 predicted oomycete proteins in our dataset. Any proteins reported by InterProScan as having a Pfam domain that could be implicated in pathogenesis were classified as potential effectors (**SupplTable_3**). The list of pathogenic Pfam domains considered were based on a number of previous analyses (14, 26, 37).

Identification of Putatively Secreted Proteins

For all 590,896 proteins, transmembrane domain prediction was carried out using THMM (93) and signal peptides were predicted using SignalP v3 (94). Proteins that had HMM S probability ≥ 0.9, NN Ymax score ≥ 0.5, NN D-score ≥ 0.5 with predicted localization "Secreted" and no transmembrane domain after signal peptide cleavage site were considered to be putatively secreted.

Enrichment Analyses

Enrichment for particular Pfam and GO terms was undertaken in Blast2GO (95) by comparing the frequency of GO terms and Pfam domains in the secretome compared to the non-secreted proportion of the proteome using Fisher's exact test corrected for false discovery rate (FDR) (50). Only Pfam domains or GO terms with enrichment p-value < 0.05 are reported.

Identification of Putative RxLR Effectors

Proteins were classified as putative RxLR effectors if they passed one of the following four criteria:

1) **Win method:** The protein contained a signal peptide in residues 1-30 followed by an RxLR motif within residues 30-60 (85).
2) **HMMsearch**: A Hidden Markov Model was run on all proteins predicted to be secreted to detect the RxLR motif using the 'cropped.hmm' HMM profile constructed by Whisson *et al* (27). This accounts for variations in the RxLR and EER motifs. Matches with a bit score > 0 were retained.
3) **Regex:** The protein must contain a signal peptide between residues 10-40, an RxLR motif within the following 100 residues followed by the EER motif within

40 residues downstream of the RxLR (allowing replacements of E to D and R to K) (27). The regular expression used was '^.{10,40}.{1,96}R.LR.{1,40}[ED][ED][KR]'.

4) **Homologous:** The set of 1,207 putative *Phytophthora* RxLRs effectors were downloaded for *Ph. infestans*, *Ph. ramorum* and *Ph. sojae* (8). These were used as reference RxLRs in the RxLR homology search. Proteins located *via* a BlastP search (E value cut off 10^{-20}) to a reference *Phytophthora* RxLR were considered putative RxLRs.

A HMM search was run on all candidate RxLRs to determine if they contain the WY fold using the HMM developed by Boutemy *et al.* (86).

Identification of Crinkler Effectors

String searches were performed to account for variations in the Crinkler "LxLFLAK" motif. First, a search was carried out with the regular expression '^.{30,70}LFLA[RK]'. All-hits were aligned using MUSCLE (v3.8.31) (96). A hidden Markov model (HMM) was built for the alignment using HMMER (3.1) (97). The HMM was searched against our entire dataset using HMMsearch to identify homologs. A second string search was carried out using the regular expression '^.{30,70}LYLA[RK]'. Again, hits from this search were aligned using MUSCLE and a HMM was built and searched against our dataset. The two results were merged as our candidate Crinkler effector set. These candidates were manually inspected and any proteins that did not contain obvious an "LxLFLAK-like" motif were excluded.

Generation of Homology Networks

Homology networks were generated for a number of protein families. In each instance, an all-vs-all BLASTp search (98) was run against each member of the family with an E value cut-off of 10^{-10} for the chitinase and NLP networks and 10^{-5} for the RxLR network. Each protein was represented by a node in the network. Two proteins were connected by an undirected edge if they were identified as homologous in our all-vs-all BLASTp search. The network was visualised and annotated in Gephi (99) and arranged using the Fruchterman Reingold layout (100). Network statistics were calculated within Gephi. Online interactive versions of all networks are available at https://oomycetes.github.io. Protein/node information is available to view in the network. The networks can be filtered to hide/show particular proteins by protein ID, species, genus or order.

Maximum-likelihood Phylogenetic Reconstruction of Effector Families

For the NLP phylogeny, the 25 Type 2 NLPs were used as bait sequences in a BLASTp (98) homology search (E value cut-off of 10^{-20}) against a local database of 8,805,033 proteins (63). All homologs were aligned using MUSCLE (96) and manually edited giving a final alignment of 460 amino acids. Molelgenerator inferred that the optimum model of substitution was the WAG model of substitution (101). A maximum likelihood phylogeny was reconstructed in PhyML using this model and 100 bootstrap replicates were undertaken (102). The final tree was visualised and annotated with iTOL (103).

For the IgA peptidase containing proteins, all 40 were aligned using MUSCLE and edited with to give a final alignment of 770 amino acids. Molelgenerator inferred that the optimum model of substitution was the WAG model of substitution. A maximum-likelihood phylogeny of all IgA peptidase containing proteins was generated using a WAG substitution model in PhyML and 100 bootstrap replicates were undertaken. The final tree was visualised and annotated with iTOL (103).

Acknowledgements

The authors wish to acknowledge the DJEI/DES/SFI/HEA Irish Centre for High-End Computing (ICHEC) for the provision of computational facilities and support. We would also like to thank Charley McCarthy for the provision of datasets. JM is funded by an Irish Research Council, Government of Ireland Postgraduate scholarship. Grant number GOIPG/2016/1112

Supplementary Data

Supplementary data for this chapter is available from the publisher website: https://msphere.asm.org/content/2/6/e00408-17/figures-only.

Supplementary data is also available at the following GitHub repository:

https://github.com/jamiemcg/ThesisSupplementaryMaterial

Supplementary Figures

Suppl_Figure1: Heatmap of enriched GO terms in oomycete secretomes. Terms are ordered on the number of the times they are observed. **A**) corresponds to biological process and **B**) corresponds to molecular function.

Suppl_Figure2: Homology network of 281 oomycete chitinases. Each protein is represented by a node. An edge joining two nodes represents shared sequence similarity between the two proteins (E value cut off of 10^{-10}). All members of GH family 19 are contained in a single cluster. Members of GH family 18 are separated into five clusters. One singleton is present in the network, a GH 48 family protein belonging to *Ph. kernoviae*. Nodes are coloured by order. An interactive version of this network is available online at https://oomycetes.github.io/chitinases.html.

Suppl_Figure3: Homology network of 4131 putative oomycete RxLR effectors. Identical to Figure 5 except Nodes are coloured by the number of RxLR criteria that labelled a particular node as being a putative RxLR. An interactive version of this network is available online at https://oomycetes.github.io/rxlrs.html.

Supplementary Tables

SupplTable_1: Overall counts of predicted secreted oomycetes proteins per species. Signal peptides were predicted using SignalP 3.

SupplTable_2: Pfam domains and GO terms enriched in the secretome of cognate oomycete genomes. Only domains or terms significantly enriched are shown.

SupplTable_3: Overall counts of oomycetes proteins with pathogenicity related domains. Proteins were functionally annotated using InterProScan 5. Counts of secreted oomycete proteins with pathogenicity related domains are also given.

SupplTable_4: Counts of oomycete glycoside hydrolase family proteins. InterProScan 5 was used to functionally annotate proteins and identify glycoside hydrolases. In total 4521 glycoside hydrolases were identified, distributed across 47 glycoside hydrolase families. Corresponding Pfam IDs are marked.

SupplTable_5: Overall counts (4131) of putative RxLR proteins detected per species. Proteins detected by Win, regex, HMM or BLAST are given. Column G gives counts for proteins per species that have been detected by Win, regex or HMM therefore proteins found by homology searches alone. This count also omits proteins that were reported to contain a putative retention motif.

SupplTable_6: List of all putative RxLR proteins detected in this analysis. Sequence information, RxLR position and cleavage site is also included. The methodology used for labelling the protein as a putative RxLR is given in columns E-H.

SupplTable_7: List of all putative RxLR ORFs detected in this analysis. Sequence information, RxLR position and cleavage site is also included. The methodology used for labelling the ORF as a putative RxLR is given in columns C-G.

References

1. Beakes GW, Glockling SL, Sekimoto S. 2012. The evolutionary phylogeny of the oomycete “fungi.” Protoplasma 249:3–19.
2. Richards TA, Dacks JB, Jenkinson JM, Thornton CR, Talbot NJ. 2006. Evolution of Filamentous Plant Pathogens: Gene Exchange across Eukaryotic Kingdoms. Curr Biol 16:1857–1864.
3. Burki F. 2014. The eukaryotic tree of life from a global phylogenomic perspective. Cold Spring Harb Perspect Biol 6:1–17.
4. Thines M, Kamoun S. 2010. Oomycete–plant coevolution: recent advances and future prospects. Curr Opin Plant Biol 13:427–433.
5. Diéguez-Uribeondo J, García MA, Cerenius L, Kozubíková E, Ballesteros I, Windels C, Weiland J, Kator H, Söderhäll K, Martín MP. 2009. Phylogenetic relationships among plant and animal parasites, and saprotrophs in Aphanomyces (Oomycetes). Fungal Genet Biol 46:365–376.
6. Hulvey JP, Padgett DE, Bailey JC. 2007. Species boundaries within Saprolegnia (Saprolegniales, Oomycota) based on morphological and DNA sequence data. Mycologia 99:421–429.
7. Banfield MJ, Kamoun S. 2013. Hooked and Cooked: A Fish Killer Genome Exposed. PLoS Genet 9:2–3.
8. Haas BJ, Kamoun S, Zody MC, Jiang RHY, Handsaker RE, Cano LM, Grabherr M, Kodira CD, Raffaele S, Torto-Alalibo T, Bozkurt TO, Ah-Fong AM V., Alvarado L, Anderson VL, Armstrong MR, Avrova A, Baxter L, Beynon J, Boevink PC, Bollmann SR, Bos JIB, Bulone V, Cai G, Cakir C, Carrington JC, Chawner M, Conti L, Costanzo S, Ewan R, Fahlgren N, Fischbach MA, Fugelstad J, Gilroy EM, Gnerre S, Green PJ, Grenville-Briggs LJ, Griffith J, Grünwald NJ, Horn K, Horner NR, Hu C-H, Huitema E, Jeong D-H, Jones AME, Jones JDG, Jones RW, Karlsson EK, Kunjeti SG, Lamour K, Liu Z, Ma L, MacLean D, Chibucos MC, McDonald H, McWalters J, Meijer HJG, Morgan W, Morris PF, Munro CA, O’Neill K, Ospina-Giraldo M, Pinzón A, Pritchard L, Ramsahoye B, Ren Q, Restrepo S, Roy S, Sadanandom A, Savidor A, Schornack S, Schwartz DC, Schumann UD, Schwessinger B, Seyer L, Sharpe T, Silvar C, Song J, Studholme DJ, Sykes S, Thines M, van de Vondervoort PJI, Phuntumart V, Wawra S, Weide R, Win J, Young C, Zhou S, Fry W, Meyers BC, van West

P, Ristaino J, Govers F, Birch PRJ, Whisson SC, Judelson HS, Nusbaum C. 2009. Genome sequence and analysis of the Irish potato famine pathogen Phytophthora infestans. Nature 461:393–398.

9. Cooke DE, Drenth a, Duncan JM, Wagels G, Brasier CM. 2000. A molecular phylogeny of Phytophthora and related oomycetes. Fungal Genet Biol 30:17–32.
10. Blair JE, Coffey MD, Park SY, Geiser DM, Kang S. 2008. A multi-locus phylogeny for Phytophthora utilizing markers derived from complete genome sequences. Fungal Genet Biol 45:266–277.
11. McCarthy CGP, Fitzpatrick DA. 2017. Phylogenomic Reconstruction of the Oomycete Phylogeny Derived from 37 Genomes. mSphere 2:e00095-17.
12. Coates ME, Beynon JL. 2010. Hyaloperonospora Arabidopsidis as a pathogen model. Annu Rev Phytopathol 48:329–345.
13. Gascuel Q, Martinez Y, Boniface M-C, Vear F, Pichon M, Godiard L. 2015. The sunflower downy mildew pathogen Plasmopara halstedii. Mol Plant Pathol 16:109–122.
14. Sharma R, Xia X, Cano LM, Evangelisti E, Kemen E, Judelson H, Oome S, Sambles C, van den Hoogen DJ, Kitner M, Klein J, Meijer HJG, Spring O, Win J, Zipper R, Bode HB, Govers F, Kamoun S, Schornack S, Studholme DJ, Van den Ackerveken G, Thines M. 2015. Genome analyses of the sunflower pathogen Plasmopara halstedii provide insights into effector evolution in downy mildews and Phytophthora. BMC Genomics 16:741.
15. Links MG, Holub E, Jiang RH, Sharpe AG, Hegedus D, Beynon E, Sillito D, Clarke WE, Uzuhashi S, Borhan MH. 2011. De novo sequence assembly of Albugo candida reveals a small genome relative to other biotrophic oomycetes. BMC Genomics 12:503.
16. Kemen E, Gardiner A, Schultz-Larsen T, Kemen AC, Balmuth AL, Robert-Seilaniantz A, Bailey K, Holub E, Studholme DJ, MacLean D, Jones JDG. 2011. Gene gain and loss during evolution of obligate parasitism in the white rust pathogen of Arabidopsis thaliana. PLoS Biol 9.
17. Gaastra W, Lipman LJA, De Cock AWAM, Exel TK, Pegge RBG, Scheurwater J, Vilela R, Mendoza L. 2010. Pythium insidiosum: An overview. Vet Microbiol 146:1–16.
18. Benhamou N, le Floch G, Vallance J, Gerbore J, Grizard D, Rey P. 2012. Pythium oligandrum: An example of opportunistic success. Microbiol (United

Kingdom) 158:2679–2694.

19. de Cock AWAM, Lodhi AM, Rintoul TL, Bala K, Robideau GP, Abad ZG, Coffey MD, Shahzad S, Lévesque CA. 2015. Phytopythium: molecular phylogeny and systematics. Persoonia - Mol Phylogeny Evol Fungi 34:25–39.
20. Wawra S, Belmonte R, Löbach L, Saraiva M, Willems A, van West P. 2012. Secretion, delivery and function of oomycete effector proteins. Curr Opin Microbiol 15:685–691.
21. Kamoun S. 2006. A catalogue of the effector secretome of plant pathogenic oomycetes. Annu Rev Phytopathol 44:41–60.
22. Giraldo MC, Valent B. 2013. Filamentous plant pathogen effectors in action. Nat Rev Microbiol 11:800–814.
23. Jiang RHY, Tyler BM. 2012. Mechanisms and Evolution of Virulence in Oomycetes. Annu Rev Phytopathol 50:295–318.
24. Tian M, Huitema E, da Cunha L, Torto-Alalibo T, 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.
25. Tian M, Win J, Song J, van der Hoorn R, van der Knaap E, Kamoun S. 2006. A Phytophthora infestans Cystatin-Like Protein Targets a Novel Tomato Papain-Like Apoplastic Protease. PLANT Physiol 143:364–377.
26. Tyler BM, Tripathy S, Zhang X, Dehal P, Jiang RHY, Aerts A, Arredondo FD, Baxter L, Bensasson D, Beynon JL, Chapman J, Damasceno CMB, Dorrance AE, Dou D, Dickerman AW, Dubchak IL, Garbelotto M, Gijzen M, Gordon SG, Govers F, Grunwald NJ, Huang W, Ivors KL, Jones RW, Kamoun S, Krampis K, Lamour KH, Lee M-K, McDonald WH, Medina M, Meijer HJG, Nordberg EK, Maclean DJ, Ospina-Giraldo MD, Morris PF, Phuntumart V, Putnam NH, Rash S, Rose JKC, Sakihama Y, Salamov AA, Savidor A, Scheuring CF, Smith BM, Sobral BWS, Terry A, Torto-Alalibo TA, Win J, Xu Z, Zhang H, Grigoriev I V, Rokhsar DS, Boore JL. 2006. Phytophthora genome sequences uncover evolutionary origins and mechanisms of pathogenesis. Science 313:1261–6.
27. Whisson SC, Boevink PC, Moleleki L, Avrova AO, Morales JG, Gilroy EM, Armstrong MR, Grouffaud S, van West P, Chapman S. 2007. A translocation signal for delivery of oomycete effector proteins into host plant cells. Nature 450:115–118.
28. Bhattacharjee S, Hiller NL, Liolios K, Win J, Kanneganti TD, Young C, Kamoun

S, Haldar K. 2006. The malarial host-targeting signal is conserved in the irish potato famine pathogen. PLoS Pathog 2:453–465.

29. Haldar K, Kamoun S, Hiller NL, Bhattacharje S, van Ooij C. 2006. Common infection strategies of pathogenic eukaryotes. Nat Rev Microbiol 4:922–931.
30. Dou D, Kale SD, Wang X, Jiang RHY, Bruce N a, Arredondo FD, Zhang X, Tyler BM. 2008. RXLR-Mediated Entry of Phytophthora sojae Effector Avr1b into Soybean Cells Does Not Require Pathogen-Encoded Machinery. Plant Cell 20:1930–1947.
31. Jiang RHY, Tripathy S, Govers F, Tyler BM. 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.
32. Win J, Krasileva K V., Kamoun S, Shirasu K, Staskawicz BJ, Banfield MJ. 2012. Sequence Divergent RXLR Effectors Share a Structural Fold Conserved across Plant Pathogenic Oomycete Species. PLoS Pathog 8:e1002400.
33. Anderson RG, Deb D, Fedkenheuer K, Mcdowell JM. 2015. Recent Progress in RXLR Effector Research. Mol Plant-Microbe Interact 28:1063–1072.
34. Schornack S, van Damme M, Bozkurt TO, Cano LM, Smoker M, Thines M, Gaulin E, Kamoun S, Huitema E. 2010. Ancient class of translocated oomycete effectors targets the host nucleus. Proc Natl Acad Sci U S A 107:17421–17426.
35. Stam R, Jupe J, Howden AJM, Morris JA, Boevink PC, Hedley PE, Huitema E. 2013. Identification and Characterisation CRN Effectors in Phytophthora capsici Shows Modularity and Functional Diversity. PLoS One 8:1–13.
36. Baxter L, Tripathy S, Ishaque N, Boot N, Cabral A, Kemen E, Thines M, Ah-Fong A, Anderson R, Badejoko W, Bittner-Eddy P, Boore JL, Chibucos MC, Coates M, Dehal P, Delehaunty K, Dong S, Downton P, Dumas B, Fabro G, Fronick C, Fuerstenberg SI, Fulton L, Gaulin E, Govers F, Hughes L, Humphray S, Jiang RHY, Judelson H, Kamoun S, Kyung K, Meijer H, Minx P, Morris P, Nelson J, Phuntumart V, Qutob D, Rehmany A, Rougon-Cardoso A, Ryden P, Torto-Alalibo T, Studholme DJ, Wang Y, Win J, Wood J, Clifton SW, Rogers J, Van den Ackerveken G, Jones JDG, McDowell JM, Beynon J, Tyler BM. 2010. Signatures of Adaptation to Obligate Biotrophy in the Hyaloperonospora arabidopsidis Genome. Science (80-) 330:1549–1551.
37. Lévesque CA, Brouwer H, Cano L, Hamilton JP, Holt C, Huitema E, Raffaele S, Robideau GP, Thines M, Win J, Zerillo MM, Beakes GW, Boore JL, Busam D,

Dumas B, Ferriera S, Fuerstenberg SI, Gachon CM, Gaulin E, Govers F, Grenville-Briggs L, Horner N, Hostetler J, Jiang RH, Johnson J, Krajaejun T, Lin H, Meijer HJ, Moore B, Morris P, Phuntmart V, Puiu D, Shetty J, Stajich JE, Tripathy S, Wawra S, van West P, Whitty BR, Coutinho PM, Henrissat B, Martin F, Thomas PD, Tyler BM, De Vries RP, Kamoun S, Yandell M, Tisserat N, Buell CR. 2010. Genome sequence of the necrotrophic plant pathogen Pythium ultimum reveals original pathogenicity mechanisms and effector repertoire. Genome Biol 11:R73.

38. Adhikari BN, Hamilton JP, Zerillo MM, Tisserat N, Lévesque CA, Buell CR. 2013. Comparative Genomics Reveals Insight into Virulence Strategies of Plant Pathogenic Oomycetes. PLoS One 8.
39. Studholme DJ, McDougal RL, Sambles C, Hansen E, Hardy G, Grant M, Ganley RJ, Williams NM. 2015. Genome sequences of six Phytophthora species associated with forests in New Zealand. Genomics Data 7:54–56.
40. Lamour KH, Mudge J, Gobena D, Hurtado-Gonzales OP, Schmutz J, Kuo A, Miller NA, Rice BJ, Raffaele S, Cano LM, Bharti AK, Donahoo RS, Finley S, Huitema E, Hulvey J, Platt D, Salamov A, Savidor A, Sharma R, Stam R, Storey D, Thines M, Win J, Haas BJ, Dinwiddie DL, Jenkins J, Knight JR, Affourtit JP, Han CS, Chertkov O, Lindquist EA, Detter C, Grigoriev I V, Kamoun S, Kingsmore SF. 2012. Genome sequencing and mapping reveal loss of heterozygosity as a mechanism for rapid adaptation in the vegetable pathogen Phytophthora capsici. Mol Plant-Microbe Interact 25:1350–1360.
41. Feau N, Taylor G, Dale AL, Dhillon B, Bilodeau GJ, Birol I, Jones SJM, Hamelin RC. 2016. Genome sequences of six Phytophthora species threatening forest ecosystems. Genomics Data 10:85–88.
42. Gao R, Cheng Y, Wang Y, Wang Y, Guo L, Zhang G. 2015. Genome Sequence of Phytophthora fragariae var. fragariae, a Quarantine Plant-Pathogenic Fungus. Genome Announc 3:6–7.
43. Sambles C, Schlenzig A, O'Neill P, Grant M, Studholme DJ. 2015. Draft genome sequences of Phytophthora kernoviae and Phytophthora ramorum lineage EU2 from Scotland. Genomics Data 6:193–194.
44. Quinn L, O'Neill PA, Harrison J, Paskiewicz KH, Mccracken AR, Cooke LR, Grant MR, Studholme DJ. 2013. Genome-wide sequencing of Phytophthora lateralis reveals genetic variation among isolates from Lawson cypress

(Chamaecyparis lawsoniana) in Northern Ireland. FEMS Microbiol Lett 344:179–185.

45. Liu H, Ma X, Yu H, Fang D, Li Y, Wang X, Wang W, Dong Y, Xiao B. 2016. Genomes and virulence difference between two physiological races of Phytophthora nicotianae. Gigascience 5:3.
46. Rujirawat T, Patumcharoenpol P, Lohnoo T, Yingyong W, Lerksuthirat T, Tangphatsornruang S, Suriyaphol P, Grenville-Briggs LJ, Garg G, Kittichotirat W, Krajaejun T. 2015. Draft Genome Sequence of the Pathogenic Oomycete Pythium insidiosum Strain Pi-S, Isolated from a Patient with Pythiosis. Genome Announc 3:e00574-15.
47. Berger H, Yacoub A, Gerbore J, Grizard D, Rey P, Sessitsch A, Compant S. 2016. Draft Genome Sequence of Biocontrol Agent Pythium oligandrum Strain Po37, an Oomycota. Genome Announc 4:e00215-16.
48. Jiang RHY, de Bruijn I, Haas BJ, Belmonte R, Löbach L, Christie J, van den Ackerveken G, Bottin A, Bulone V, Díaz-Moreno SM, Dumas B, Fan L, Gaulin E, Govers F, Grenville-Briggs LJ, Horner NR, Levin JZ, Mammella M, Meijer HJG, Morris P, Nusbaum C, Oome S, Phillips AJ, van Rooyen D, Rzeszutek E, Saraiva M, Secombes CJ, Seidl MF, Snel B, Stassen JHM, Sykes S, Tripathy S, van den Berg H, Vega-Arreguin JC, Wawra S, Young SK, Zeng Q, Dieguez-Uribeondo J, Russ C, Tyler BM, van West P. 2013. Distinctive Expansion of Potential Virulence Genes in the Genome of the Oomycete Fish Pathogen Saprolegnia parasitica. PLoS Genet 9:e1003272.
49. Sperschneider J, Williams AH, Hane JK, Singh KB, Taylor JM. 2015. Evaluation of Secretion Prediction Highlights Differing Approaches Needed for Oomycete and Fungal Effectors. Front Plant Sci 6:1–14.
50. Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc 57:289–300.
51. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S-Y, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236–1240.
52. Ponchet M, Panabières F, Milat ML, Mikes V, Montillet JL, Suty L, Triantaphylides C, Tirilly Y, Blein JP. 1999. Are elicitins cryptograms in plant-Oomycete communications? Cell Mol Life Sci.

53. Gibbs GM, Roelants K, O'Bryan MK. 2008. The CAP superfamily: Cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 proteins - Roles in reproduction, cancer, and immune defense. Endocr Rev.
54. Schneiter R, Di Pietro A. 2013. The CAP protein superfamily: function in sterol export and fungal virulence. Biomol Concepts 4:519–525.
55. Zhou H, Casas-Finet JR, Coats RH, Kaufman JD, Stahl SJ, Wingfield PT, Rubin JS, Bottaro DP, Byrd RA. 1999. Identification and dynamics of a heparin-binding site in hepatocyte growth factor. Biochemistry 38:14793–14802.
56. Gaulin E, Jauneau A, Villalba F, Rickauer M, Esquerré-Tugayé M-T, Bottin A. 2002. The CBEL glycoprotein of Phytophthora parasitica var-nicotianae is involved in cell wall deposition and adhesion to cellulosic substrates. J Cell Sci 115:4565–75.
57. Raffaele S, Win J, Cano LM, Kamoun S. 2010. Analyses of genome architecture and gene expression reveal novel candidate virulence factors in the secretome of Phytophthora infestans. BMC Genomics 11:637.
58. Feng BZ, Zhu XP, Fu L, Lv RF, Storey D, Tooley P, Zhang XG. 2014. Characterization of necrosis-inducing NLP proteins in Phytophthora capsici. BMC Plant Biol 14:126.
59. Oome S, Van den Ackerveken G. 2014. Comparative and functional analysis of the widely occurring family of Nep1-like proteins. Mol Plant Microbe Interact 27:1–51.
60. Oome S, Raaymakers TM, Cabral A, Samwel S, Böhm H, Albert I, Nürnberger T, Van den Ackerveken G. 2014. Nep1-like proteins from three kingdoms of life act as a microbe-associated molecular pattern in Arabidopsis. Proc Natl Acad Sci U S A 111:16955–60.
61. Gijzen M, Nürnberger T. 2006. Nep1-like proteins from plant pathogens: Recruitment and diversification of the NPP1 domain across taxa. Phytochemistry 67:1800–1807.
62. Horner NR, Grenville-Briggs LJ, van West P. 2012. The oomycete Pythium oligandrum expresses putative effectors during mycoparasitism of Phytophthora infestans and is amenable to transformation. Fungal Biol 116:24–41.
63. McCarthy CGP, Fitzpatrick DA. 2016. Systematic Search for Evidence of Interdomain Horizontal Gene Transfer from Prokaryotes to Oomycete Lineages. mSphere 1:1–18.

64. Mistry D, Stockley RA. 2006. IgA1 protease. Int J Biochem Cell Biol 38:1244–1248.
65. Unestam T, Weiss DW. 1970. The host-parasite relationship between freshwater crayfish and the crayfish disease fungus Aphanomyces astaci: responses to infection by a susceptible and a resistant species. J Gen Microbiol 60:77–90.
66. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42:490–495.
67. Seidl V. 2008. Chitinases of filamentous fungi: a large group of diverse proteins with multiple physiological functions. Fungal Biol Rev 22:36–42.
68. Duo-Chuan L. 2006. Review of fungal chitinases. Mycopathologia 161:345–360.
69. Mélida H, Sandoval-Sierra J V., Diéguez-Uribeondo J, Bulone V. 2013. Analyses of extracellular carbohydrates in oomycetes unveil the existence of three different cell wall types. Eukaryot Cell 12:194–203.
70. Bowman SM, Free SJ. 2006. The structure and synthesis of the fungal cell wall. BioEssays 28:799–808.
71. Henrissat B, Bairoch A. 1993. New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem J 293:781–788.
72. Mohnen D. 2008. Pectin structure and biosynthesis. Curr Opin Plant Biol 11:266–277.
73. Marin-Rodriguez MC. 2002. Pectate lyases, cell wall degradation and fruit softening. J Exp Bot 53:2115–2119.
74. Matteo A Di, Giovane A. 2005. Structural basis for the interaction between pectin methylesterase and a specific inhibitor protein. Plant Cell … 17:849–858.
75. Fries M, Ihrig J, Brocklehurst K, Shevchik VE, Pickersgill RW. 2007. Molecular basis of the activity of the phytopathogen pectin methylesterase. EMBO J 26:3879–3887.
76. Vercauteren I, de Almeida Engler J, De Groodt R, Gheysen G. 2002. An Arabidopsis thaliana Pectin Acetylesterase Gene Is Upregulated in Nematode Feeding Sites Induced by Root-knot and Cyst Nematodes. Mol Plant-Microbe Interact 15:404–407.
77. Schäfer W. 1993. The role of cutinase in fungal pathogenicity. Trends Microbiol 1:69–71.
78. Yeats TH, Rose JKC. 2013. The formation and function of plant cuticles. Plant

Physiol 163:5–20.

79. Belbahri L, Calmin G, Mauch F, Andersson JO. 2008. Evolution of the cutinase gene family: Evidence for lateral gene transfer of a candidate Phytophthora virulence factor. Gene 408:1–8.

80. Li D, Ashby AM, Johnstone K. 2003. Molecular Evidence that the Extracellular Cutinase Pbc1 Is Required for Pathogenicity of Pyrenopeziza brassicae on Oilseed Rape. Mol Plant-Microbe Interact 16:545–552.

81. Kamoun S, Furzer O, Jones JDG, Judelson HS, Ali GS, Dalio RJD, Roy SG, Schena L, Zambounis A, Panabières F, Cahill D, Ruocco M, Figueiredo A, Chen X-R, Hulvey J, Stam R, Lamour K, Gijzen M, Tyler BM, Grünwald NJ, Mukhtar MS, Tomé DFA, Tör M, Van Den Ackerveken G, McDowell J, Daayf F, Fry WE, Lindqvist-Kreuze H, Meijer HJG, Petre B, Ristaino J, Yoshida K, Birch PRJ, Govers F. 2015. The Top 10 oomycete pathogens in molecular plant pathology. Mol Plant Pathol 16:413–434.

82. Orsomando G, Lorenzi M, Raffaelli N, Dalla Rizza M, Mezzetti B, Ruggieri S. 2001. Phytotoxic Protein PcF, Purification, Characterization, and cDNA Sequencing of a Novel Hydroxyproline-containing Factor Secreted by the Strawberry Pathogen Phytophthora cactorum. J Biol Chem 276:21578–21584.

83. Meijer HJG, Mancuso FM, Espadas G, Seidl MF, Chiva C, Govers F, Sabido E. 2014. Profiling the Secretome and Extracellular Proteome of the Potato Late Blight Pathogen Phytophthora infestans. Mol Cell Proteomics 13:2101–2113.

84. Stahl E a, Bishop JG. 2000. Plant–pathogen arms races at the molecular level. Curr Opin Plant Biol 3:299–304.

85. Win J, Morgan W, Bos J, Krasileva K V., Cano LM, Chaparro-garcia A, Ammar R, Staskawicz BJ, Kamoun S. 2007. Adaptive Evolution Has Targeted the C-Terminal Domain of the RXLR Effectors of Plant Pathogenic Oomycetes © American Society of Plant Biologists Adaptive Evolution Has Targeted the C-Terminal Domain of the RXLR Effectors of Plant Pathogenic Oomycetes. Plant Cell 19:2349–2369.

86. Boutemy LS, King SRFF, Win J, Hughes RK, Clarke TA, Blumenschein TMAA, Kamoun S, Banfield MJ. 2011. Structures of Phytophthora RXLR effector proteins: A conserved but adaptable fold underpins functional diversity. J Biol Chem 286:35834–35842.

87. Van West P, De Bruijn I, Minor KL, Phillips AJ, Robertson EJ, Wawra S, Bain J,

Anderson VL, Secombes CJ. 2010. The putative RxLR effector protein SpHtp1 from the fish pathogenic oomycete Saprolegnia parasitica is translocated into fish cells. FEMS Microbiol Lett 310:127–137.

88. Wawra S, Bain J, Durward E, de Bruijn I, Minor KL, Matena A, Lobach L, Whisson SC, Bayer P, Porter AJ, Birch PRJ, Secombes CJ, van West P. 2012. Host-targeting protein 1 (SpHtp1) from the oomycete Saprolegnia parasitica translocates specifically into fish cells in a tyrosine-O-sulphate-dependent manner. Proc Natl Acad Sci 109:2096–2101.
89. Kale SD, Gu B, Capelluto DGS, Dou D, Feldman E, Rumore A, Arredondo FD, Hanlon R, Fudal I, Rouxel T, Lawrence CB, Shan W, Tyler BM. 2010. External Lipid PI3P Mediates Entry of Eukaryotic Pathogen Effectors into Plant and Animal Host Cells. Cell 142:284–295.
90. Allen RL. 2004. Host-Parasite Coevolutionary Conflict Between Arabidopsis and Downy Mildew. Science (80-) 306:1957–1960.
91. Stanke M, Morgenstern B. 2005. AUGUSTUS: A web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res 33:465–467.
92. Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M. 2008. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. Genome Res 18:1979–1990.
93. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305:567–80.
94. Bendtsen JD, Nielsen H, Von Heijne G, Brunak S. 2004. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 340:783–795.
95. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674–3676.
96. Edgar RC. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797.
97. Eddy S. 1998. Profile hidden Markov models. Bioinformatics 14:755–763.
98. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402.

99. Bastian M, Heymann S, Jacomy M. 2009. Gephi: An Open Source Software for Exploring and Manipulating Networks. Third Int AAAI Conf Weblogs Soc Media 361–362.
100. Fruchterman TMJ, Reingold EM. 1991. Graph drawing by force-directed placement. Softw Pract Exp 21:1129–1164.
101. Keane TM, Creevey CJ, Pentony MM, Naughton TJM, McInerney JO. 2006. Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. BMC Evol Biol 6:29.
102. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New Algorithms and Mehtods to Estimate Maximum-Likelihood Phylogenies: Asessing the Performance of PhyML 2.0. Syst Biol 59:307–321.
103. Letunic I, Bork P. 2007. Interactive Tree Of Life (iTOL): An online tool for phylogenetic tree display and annotation. Bioinformatics 23:127–128.

Chapter 3
Comparative Analysis of Oomycete Genome Evolution using the Oomycete Gene Order Browser (OGOB)

This chapter has been published in the journal Genome Biology and Evolution.

McGowan, J., Byrne, K.P., Fitzpatrick, D.A., 2019. Comparative Analysis of Oomycete Genome Evolution Using the Oomycete Gene Order Browser (OGOB). Genome Biology and Evolution. 11, 189–206.

Abstract

The oomycetes are a class of microscopic, filamentous eukaryotes within the stramenopiles-alveolates-rhizaria (SAR) eukaryotic supergroup. They include some of the most destructive pathogens of animals and plants, such as *Phytophthora infestans*, the causative agent of late potato blight. Despite the threat they pose to worldwide food security and natural ecosystems, there is a lack of tools and databases available to study oomycete genetics and evolution. To this end, we have developed the Oomycete Gene Order Browser (OGOB), a curated database that facilitates comparative genomic and syntenic analyses of oomycete species. OGOB incorporates genomic data for 20 oomycete species including functional annotations and a number of bioinformatics tools. OGOB hosts a robust set of orthologous oomycete genes for evolutionary analyses. Here we present the structure and function of OGOB as well as a number of comparative genomic analyses we have performed to better understand oomycete genome evolution. We analyse the extent of oomycete gene duplication and identify tandem gene duplication as a driving force of the evolution and expansion of secreted oomycete genes. We use phylostratigraphy to compare the composition of oomycete genomes. We identify core genes that are present and microsyntenically conserved in oomycete lineages and identify the degree of microsynteny between each pair of 20 species housed in OGOB. Consistent with previous comparative synteny analyses between a small number of oomycete species, our results reveal an extensive degree of microsyntenic conservation amongst genes with housekeeping functions within the oomycetes. OGOB is available at https://ogob.ie.

Introduction

The oomycetes are a class of filamentous, eukaryotic microorganisms that include some of the most devastating plant and animal pathogens (Beakes et al. 2012). They represent one of the biggest threats to worldwide food security and natural ecosystems (Fisher et al. 2012). Oomycetes resemble fungi in terms of their morphology, filamentous growth, ecological niches and modes of nutrition (Richards et al. 2006). Despite their extensive similarities, the evolutionary relationship between oomycetes and fungi represent one of the most distantly related evolutionary groupings within the eukaryotes (Burki 2014). Oomycetes are members of the stramenopiles lineage of the Stramenopiles-Alveolata-Rhizaria eukaryotic supergroup, with close relationships to the diatoms and brown algae (Burki 2014). Within the oomycete class, there are a number of highly diverse orders, including the Peronosporales, Pythiales, Albuginales and Saprolegniales orders (**Figure 1**). There is significant diversity both between and within these orders in terms of lifestyle, pathogenicity and host range. The Peronosporales order is the most extensively studied order, consisting largely of phytopathogens, including the hemibiotrophic genus *Phytophthora* (the "plant destroyers") (**Figure 1**). The most notorious of which is *Phytophthora infestans*, the causative agent of late potato blight and causative agent of the Irish potato famine which resulted in the death of one million people in Ireland and the emigration of another million (Haas et al. 2009). *Phytophthora infestans* is reported to cause billions of euros' worth of worldwide potato crop loss annually (Haverkort et al. 2008). Other highly destructive *Phytophthora* species include *Ph. sojae* and *Ph. ramorum*. *Ph. sojae* has a narrow host range, infecting soybean and costs between one and two billion dollars in crop loss per year (Tyler 2006; Tyler 2007). *Phytophthora ramorum*, in contrast, has a very wide host range with more than 100 host species reported and is destroying forest ecosystems worldwide (Rizzo et al. 2005; Tyler 2006). Other

members of the Peronosporales order include the obligate biotrophic genera *Hyaloperonospora* and *Plasmopara* (**Figure 1**), which cause downy mildew in a number of economically important crops (Coates and Beynon 2010; Gascuel et al. 2015; Sharma et al. 2015). The Pythiales order includes the cosmopolitan genus *Pythium* (**Figure 1**), which are generalist necrotrophs with broad host ranges that cause root rot in many important crops and ornamental plants that mainly inhabit soils (Lévesque et al. 2010; Adhikari et al. 2013). *Pythium ultimum var. ultimum* (hereafter referred to as *Py. ultimum*) is one of the most pathogenic *Pythium* species, with a broad host range including corn, soybean, wheat and ornamental plants (Cheung et al. 2008). The Albuginales order is a more basal order (**Figure 1**) that includes the obligate biotrophic genus *Albugo* which causes "white blister rust" disease in various Brassicaceae species, including mustard and cabbage family plants (Kemen et al. 2011; Links et al. 2011). The Saprolegniales order (**Figure 1**) include animal and plant pathogens from the *Aphanomyces* genus (Diéguez-Uribeondo et al. 2009; Makkonen et al. 2016), and the *Saprolegnia* genus which causes severe infection of animals, in particular they cause "cotton mould" disease in many fish that are important in the global aquaculture industry (van den Berg et al. 2013; Jiang et al. 2013).

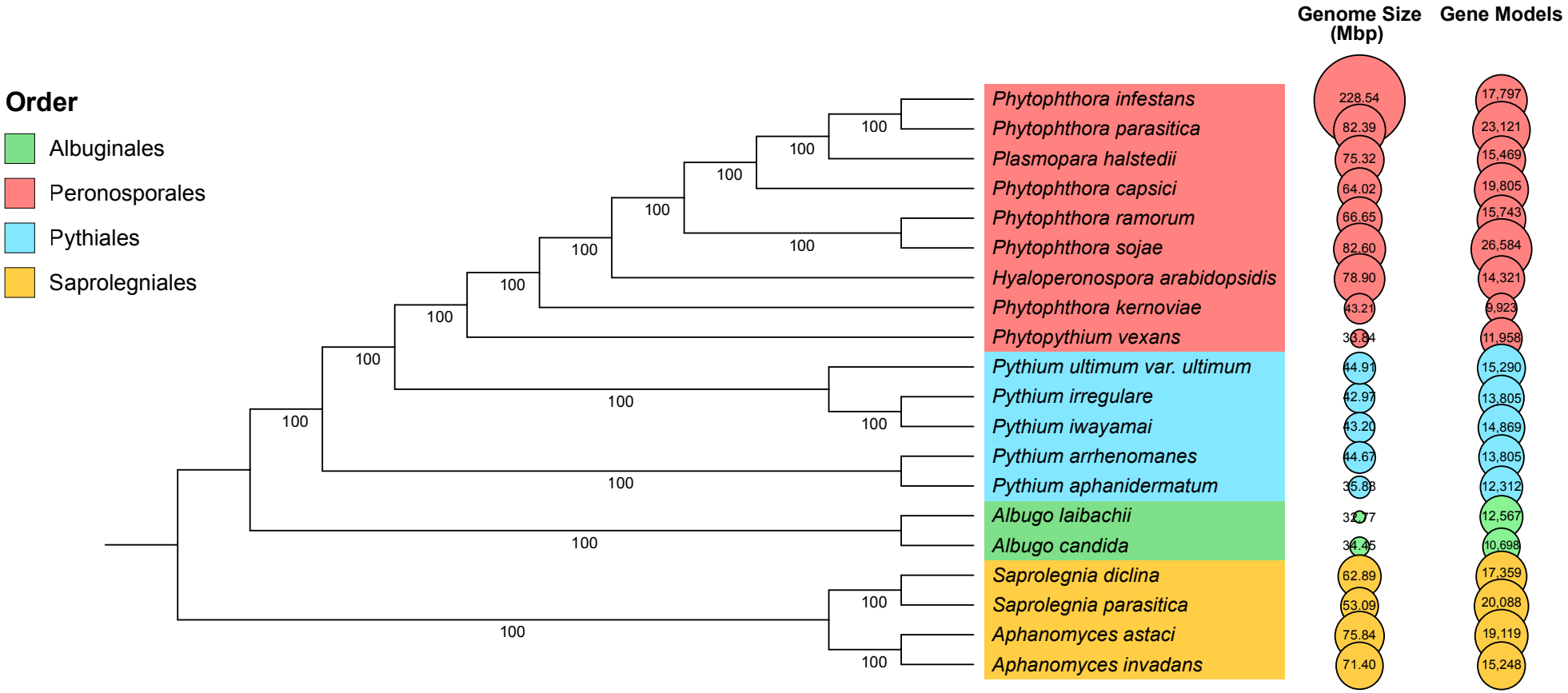


Figure 1. Matrix representation with parsimony (MRP) supertree of the 17,738 OGOB pillars that contain at least 4 genes. The supertree was generated in CLANN and is rooted at the Saprolegniales order. All nodes have 100% bootstrap support. Species are coloured by order as follows: green, Albuginales; red, Peronosporales; blue, Pythiales; orange, Saprolegniales.

The genomes of a number of oomycete species have been sequenced in recent years and have revealed substantial differences in terms of genome size, gene content and organisation. Assembly sizes of sequenced oomycetes range from 33 Mb for *Al. laibachii* (Kemen et al. 2011), to 229 Mb for *Ph. infestans* (Haas et al. 2009) (**Table 1)**. Differences in genome size are largely accounted for by proliferations of repetitive DNA and transposable elements as opposed to increases in the number of genes. For example, repetitive DNA accounts for 74% of the *Ph. infestans* genome (Haas et al. 2009). Where differences in gene content do occur, they may be due to expansions of large arsenals of secreted effector proteins that facilitate pathogenicity (Kamoun 2006; McGowan and Fitzpatrick 2017). Effector genes mediate infection by degrading host cell components, dampening host immune responses and inducing necrosis. Previous analyses have detected a high degree of synteny (conserved gene order) between *Phytophthora* species (Jiang et al. 2006; Haas et al. 2009; Ospina-Giraldo et al. 2010). Conservation of synteny also extends to *Hy. arabidopsidis* (Baxter et al. 2010), however when compared to more distantly related species such as *Pythium* (Lévesque et al. 2010) or *Albugo* (Kemen et al. 2011) species, a lesser degree of syntenic conservation is observed. Syntenically conserved regions of the genome are typically gene-dense and contain housekeeping genes whereas effector proteins are found at synteny breakpoints in gene-sparse, repeat-rich regions of the genome (Tyler 2006; Haas et al. 2009; Jiang and Tyler 2012).

Table 1. Genome statistics and references for oomycete species hosted on OGOB.

Species	Assembly Size (Mb)	Scaffolds	N50	L50	% GC	% Gaps	Genes	Orfans[a]	G50	G90	BUSCO[b]	Reference
Phytophthora infestans	228.5	4,921	1588.6	38	51.0	16.8	17,797	793	18	103	95.8%	Haas et al. 2009
Phytophthora parasitica	82.4	708	888.3	25	49.5	34.6	23,121	3,130	20	87	94.1%	Broad Institute
Plasmopara halstedii	75.3	3,162	1546.2	16	45.3	11.3	15,469	4,830	16	52	93.2%	Sharma et al. 2015
Phytophthora capsici	64.0	917	705.7	29	50.4	12.5	19,805	390	27	89	91.0%	Lamour et al. 2012
Hyaloperonospora arabidopsidis	78.9	3,044	332.4	70	47.2	10.2	14,321	4,132	59	242	91.8%	Baxter et al. 2010
Phytophthora sojae	82.6	83	7609.2	4	54.6	4.0	26,584	1,725	4	13	97.9%	Tyler et al. 2006
Phytophthora ramorum	66.7	2,576	308.0	63	53.9	18.3	15,743	150	47	548	97.4%	Tyler et al. 2006
Phytophthora kernoviae	43.2	1,805	73.0	162	49.7	0.4	9,923	492	163	603	96.2%	Sambles et al. 2015
Phytopythium vexans	33.8	3,685	29.2	340	58.9	0.8	11,958	251	338	1,365	93.2%	Adhikari et al. 2013
Pythium iwayamai	43.2	11,541	11.0	1107	55.1	3.6	14,869	336	1,086	4,351	80.3%	Adhikari et al. 2013
Pythium irregulare	43.0	5,887	23.2	494	53.8	0.7	13,805	167	473	2,343	94.0%	Adhikari et al. 2013

Table 1. Continued.

Species	Assembly Size (Mb)	Scaffolds	N50	L50	% GC	% Gaps	Genes	Orfans[a]	G50	G90	BUSCO[b]	Reference
Pythium ultimum	44.9	975	837.8	19	52.3	4.7	15,290	557	19	54	97.0%	Lévesque et al. 2010
Pythium arrhenomanes	44.7	10,972	9.8	1195	56.9	4.2	13,805	299	1,010	4,252	81.2%	Adhikari et al. 2013
Pythium aphanidermatum	35.9	1,774	37.4	270	53.8	4.5	12,312	189	244	873	90.6%	Adhikari et al. 2013
Albugo laibachii	32.8	3,827	69.4	130	44.3	0.0	12,567	2,238	106	391	80.3%	Kemen et al. 2011
Albugo candida	34.5	5,216	51.4	171	43.2	0.0	10,698	2,334	108	443	79.9%	Links et al. 2011
Saprolegnia diclina	62.9	390	602.6	34	58.6	35.7	17,359	361	26	95	93.2%	Broad Institute
Saprolegnia parasitica	53.1	1,442	280.9	46	58.5	9.3	20,088	603	41	371	87.2%	Jiang et al. 2013
Aphanomyces invadans	71.4	481	1130.2	19	54.2	41.9	15,248	986	14	56	90.6%	Broad Institute
Aphanomyces astaci	75.8	835	657.5	31	49.8	22.8	19,119	2,127	24	109	87.2%	Broad Institute

[a]Orphans were identified as species-specific in our phylostratigraphy analysis.

[b]Percentage completeness as determined by BUSCO v3 against the Alveolata/Stramenopile BUSCO data set.

Despite their economic impact and the threat that they pose to worldwide food security, there is a lack of tools and databases available to study oomycete genes and genomes. This is particularly striking when compared with other taxonomic groups such as fungi. Databases such as *Pythium* Genome Database and the Comprehensive Phytopathogen Genomics Resource (Hamilton et al. 2011) have been retired. FungiDB contains genome data for 16 oomycete species including information pertaining to orthology and synteny (Basenko et al. 2018), however the genome browser in FungiDB displays a to-scale representation of chromosomal regions making it unsuitable for the analysis of gene order and evolution. EumicrobeDB (Panda et al. 2018) was recently published and contains the genomes of several oomycete species and a large number of bioinformatics tools. EumicrobeDB has a tool for comparing syntenic regions between species, however, it is limited to comparing two species and displays a to-scale representation of chromosomal regions. Furthermore, it is not immediately obvious if an ortholog is absent in a species or if it is present in another area of the genome. These issues make EumicrobeDB unsuitable for detailed analysis of gene order and evolution across multiple species. To overcome this, we have developed the Oomycete Gene Order Browser (OGOB).

OGOB is a curated database that currently hosts genomic data for 20 oomycete species. Species included in OGOB were selected to include a broad range of representatives from the oomycete class and also based on the availability of gene sets. A recent review carried out a survey to rank the "top 10" oomycetes in terms of their economic and scientific importance (Kamoun et al. 2015). OGOB hosts eight of these species. OGOB also hosts a number of useful bioinformatics tools that allow users to carry out bioinformatic analyses in the web browser without installing local command line tools. This makes OGOB useful for comparative genomic, syntenic and evolutionary

analyses of oomycete genomes as well as for the analysis of individual genes and gene families. OGOB is based on the original engine developed for the Yeast Gene Order Browser (YGOB) (Byrne and Wolfe 2005; Byrne and Wolfe 2006), to which we have made a number of functional and visual upgrades.

Here we describe the structure and functionality of OGOB. We have also undertaken a number of comparative genomic analyses using the genome data housed in OGOB. These analyses yield insights into the evolution of oomycete genomes and the effect that gene duplication has had in shaping the gene repertoire of individual species. Using OGOB, we have investigated the overall conserved core of oomycete genes as well as individual orders. Furthermore, we have also completed a comprehensive analysis of the 190 possible pairwise synteny comparisons between the 20 species hosted in OGOB. OGOB is available at https://ogob.ie.

Results and Discussion

OGOB Structure and Functions

OGOB includes 20 oomycete species (**Table 1**) that were selected to include a broad range of representatives from the oomycete class and based on the availability of gene sets. The database hosts six *Phytophthora* species *(Ph. infestans*, *Ph. parasitica*, *Ph. capsici*, *Ph. sojae*, *Ph. ramorum* and *Ph. kernoviae*), two downy mildews (*Pl. halstedii* and *Hy. arabidopsidis*), five *Pythium* species (*Py. iwayamai*, *Py. irregulare*, *Py. ultimum var. ultimum*, *Py. arrhenomanes* and *Py. aphanidermatum*), *Phytopythium vexans*, two *Albugo* species (*Al. candida* and *Al. laibachii*), two *Saprolegnia* species (*Sa. diclina* and *Sa. parasitica*) and two *Aphanomyces* species (*Ap. invadans* and *Ap. astaci*). A total of 319,881 putative protein coding genes are hosted by OGOB.

Similar to YGOB (Byrne and Wolfe 2005), OGOB's visual browser contains two key structures: horizontal tracks and vertical orthology pillars (**Figure 2A**). The horizontal tracks represent chromosomal (scaffold) segments, with the species name shown to the right. The colours of gene boxes correspond to genes on the same scaffold. Vertical orthology pillars list orthologous genes across species. In OGOB, orthologs can be present and syntenically conserved, present and not syntenically conserved or absent. Gene boxes are coloured grey when there is no evidence of syntenic conservation. A pillar contains a vacant slot when an ortholog could not be found in that particular species.

Links below each pillar (**Figure 2A**) allow users to retrieve the corresponding amino acid and nucleotide sequences for that pillar. Clicking the "i" button on any gene launches an information page for that gene showing any functional annotations (**Figure 2B**). Clicking on any annotation will link the user to the relevant annotation database (e.g. Pfam and InterPro). The information page for each gene also has a BLAST facility that permits users to search the corresponding nucleotide/protein sequence against OGOB's

gene/protein data sets. Users can also search against the full genome sequences or intergenic regions only. This facility allows users to confirm that a gene is missing from an assembly for example and in part helps overcome any of the shortfalls associated with genes that may have been missed during gene calling. Links at the top of each pillar (**Figure 2A)** allow users to construct phylogenetic trees ("Tree"), perform multiple sequence alignments ("Align") and calculate evolutionary rates ("Rates"). Maximum likelihood phylogenies are generated using PhyML (Guindon et al. 2010) and displayed in an interactive interface that allows users to manipulate and root trees, implemented using phylotree.js (**Figure 2C**). Furthermore, users can also download phylogenies in Newick format for further processing. The rate of non-synonymous (d_N) and synonymous substitutions (d_S) are calculated using yn00 (Yang and Nielsen 2000). A d_N/d_s ratio > 1 is indicative of positive selection and is highlighted in OGOB (**Figure 2D**). Multiple sequence alignments are performed using MUSCLE (Edgar 2004) and displayed in an interactive interface implemented using MSAViewer (Yachdav et al. 2016), the consensus sequence for the pillar is also shown (**Figure 2E**). Clicking the "b" button on any gene in OGOB launches a BLAST search of that gene against the entire OGOB database. BLAST results are coloured to highlight orthologs, paralogs, tandem duplicates, singletons and syntenically conserved hits. Users can select hits from the BLAST search and perform the above functions. This allows users to quickly analyse BLAST hits without having to manually obtain their sequences. We have also integrated a search interface into OGOB that makes it easier to study particular genes or gene families without the need to know individual gene identifiers. For example, users can search for genes that contain specific Gene Ontology terms or Pfam domains and easily compare the presence or absence of gene families across species and investigate their syntenic context. In addition, we have incorporated BLAST search support allowing users to

search protein or nucleotide sequences against the OGOB database. The BLAST results page provides links to view hits in OGOB.

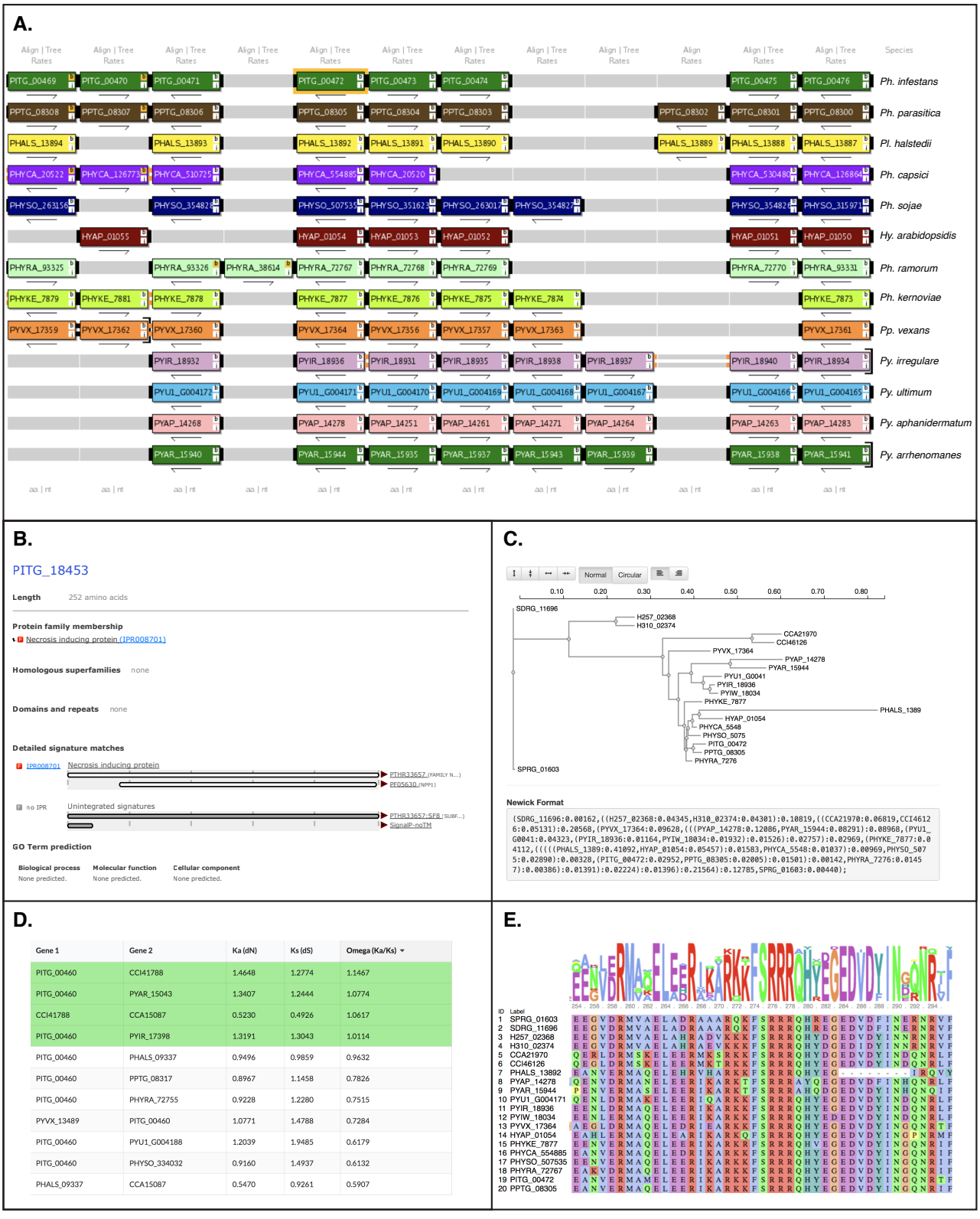


Figure 2. The Oomycete Gene Order Browser. **(A)** OGOB Screenshot. Each horizontal track represents a chromosomal segment from one species, with species labels on the right.

Figure 2 (continued). Each box represents a protein coding gene, with the gene ID shown. Genes that are in the same pillar are orthologous. Each colour represents a chromosome/scaffold. A change in colour represents a breakdown in synteny. Genes coloured grey indicate a non-syntenic ortholog. Arrows under gene boxes represent the relative transcriptional orientation. Adjacent genes are connected by a sloid black connector. Connector are coloured grey if there is a gap in that genome. Double small bars connect non-adjacent genes that are < 5 genes apart and single small bars connect genes that are < 20 genes apart. Connectors are coloured orange if there is an inversion. In this screenshot, the browser is focused on PITG_00472, an mRNA splicing factor SYF2 gene. This gene is in a genomic segment that is syntenically conserved in the Peronosporales and Pythiales, as shown by the coloured blocks. **(B)** Gene info page showing functional annotations carried out by InterProScan, accessed by clicking the “i" button on gene boxes in OGOB. **(C)** Interactive maximum likelihood phylogeny of the genes in the same pillar as PITG_00472, accessed by clicking the “Tree” button at the top of pillars. **(D)** Pairwise yn00 evolutionary rates of genes in OGOB accessed by clicking the “Rates” button at the top of pillars. Genes highlighted green show evidence of positive selection, i.e. a d_N/d_S ratio ≥ 1. **(E)** MUSCLE multiple sequence alignment of all genes in the same pillar as PITG_00472, accessed by clicking the “Align” button at the top of pillars. The consensus sequence of the pillar is also shown.

A recent evaluation of synteny analysis methods has highlighted the negative effect that poor assembly contiguity has, resulting in an underestimation of syntenic conservation and the authors have recommend that a minimum N50 score of 1 Mb is required for robust synteny analysis (Liu et al. 2018). Many of the oomycete genome assemblies are highly fragmented for example *Py. arrhenomanes* has an N50 score of

only 9.8 Kb and *Py. iwayamai* has an N50 score of 11.5 Kb (**Table 1**). We have decided to include such assemblies in OGOB and in our synteny analyses regardless of their fragmented nature, to ensure that a broad set of oomycete species are represented. Completeness of each genome assembly was assessed using BUSCO v3 (Waterhouse et al. 2018) based on the alveolata/stramenopiles set of common BUSCOs (benchmarking universal single-copy orthologs). This analysis revealed genome completeness ranging from 79.9% in *Al. candida* to 97.9% in *Ph. sojae*, with an overall average completeness of 90.6% (**Table 1; Figure S1**). This indicates that the genomes included in OGOB are of high completeness in terms of gene content despite their fragmented nature. Also, by anchoring OGOB on genomes with higher quality assemblies (e.g. *Ph. sojae*, *Ph. infestans* or *Ap. invadans*), we alleviate some of the negative effects that poor assembly quality has on synteny estimates. OGOB also uses a microsyntenic approach to determine syntenic conservation, focusing on local gene order and does not take into account intergenic distance, rather than whole genome alignments which succumb more severely to the effects of fragmented assemblies.

N50 scores are the most common score used to assess the quality of genome assemblies. However, it is well known that the metric suffers many problems. N50 scores can be artificially inflated in assemblies with large proportions of gaps or misassemblies. Furthermore, it can be difficult to determine how fragmented an assembly is based on N50 score alone without knowing the number of contigs/scaffolds or predicted genome size and they do not take into account gene models. Here we report an alternative metric which we call a "G50 score". An assembly's G50 score is the minimum number of scaffolds that contain at least 50% percent of genes. More generally, G*x* is the minimum number of contigs/scaffolds that contain at least *x* % of genes. G50 scores make it more immediately obvious how fragmented an assembly is and is better suited for synteny

analyses than N50 scores. G50 scores are roughly proportional to L50 scores (**Table 1**), except they do not take into account scaffolds that do not contain genes. For example, *Ph. sojae* has a G50 score of 4 (**Table 1**) indicating a very high quality, contiguous assembly whereby 4 scaffolds contain at least 50% of the total genes. In comparison, *Py. iwayami* has a G50 score of 1,086 (**Table 1**) indicating that this is a very fragmented assembly where 50% of the genes are distributed across 1,086 scaffolds. Without knowing any other metrics of the assembly we can tell that many scaffolds have only 1 gene therefore it is not possible to determine the syntenic context of these genes. Due to the fragmented nature, low N50 scores and poor G50 scores of some assemblies in OGOB, the levels of global synteny reported herein may be underestimates.

Orthology Curation and Syntenolog Search

Identification of orthologous genes is an important first step in many evolutionary and comparative genomic analyses and is essential for the functional annotation of newly sequenced genomes. Most orthology prediction methods rely on sequence similarity searches however events such as gene duplications, gene losses and rapid evolution can have a significant negative effect on the accuracy of orthology prediction. In OGOB, we use a combination of sequence similarity and syntenic conservation to identify and host a robust set of oomycete orthologs.

Genes were initially added to orthology pillars in OGOB using a reciprocal best BLAST hits strategy. Genes that are each other's best hits in a BLASTp search (E-value cut-off $1e^{-10}$) are considered orthologs and added to the same pillar. This strategy initially placed the 319,881 oomycete genes into 146,768 pillars. A large number of pillars were singleton pillars (pillars with only 1 gene) that had significant BLAST hits to genes in other pillars but not reciprocal best hits. Using a similar approach to Synteno-BLAST

which was used in CGOB (Maguire et al. 2013) and SearchDOGS (OhEigeartaigh et al. 2014), we have developed an automated strategy called "Syntenolog Search" that combines results from BLAST searches with synteny information to identify microsyntenically conserved orthologs that cannot be identified using reciprocal best BLAST hits searches alone. We use the term "syntenolog" to describe syntenically conserved orthologs. Syntenolog Search systematically examines all singleton pillars to check if they can be merged with another pillar based on homology and microsyntenic context. Compared to Synteno-BLAST, we use a stricter E-value cut-off and a more permissive definition of microsynteny. Briefly, each singleton is searched against the OGOB database using BLASTp (E-value cut-off $1e^{-10}$). Hits are then examined for microsyntenic conservation to determine if there exists a pair of neighbouring orthologs, within a distance of 20 genes that serve as anchor points, i.e. the query gene and hit gene are in a conserved genomic neighbourhood. If such a hit exists, the two genes are considered syntenologs and the pillars are merged. For example, consider the two genes PITG_00248 and PPTG_10928, both of which are Papain family cysteine proteases (PF00112). These genes do not have reciprocal best hits in a BLASTp search, likely because they are both members of large paralogous families. However, they were identified as syntenologs by Syntenolog Search and as a result both genes were moved to the same pillar (**Figure S2A**). Upon manual inspection in OGOB, they are obvious orthologs. They share significant sequence similarity ($1e^{-125}$) (**Figure S2B, Figure S2C**) to each other and are syntenically conserved, co-occurring at the same loci (i.e. distance = 0) (**Figure S2A**). This highlights the power of Syntenolog Search in identifying reliable orthologous relationships that cannot be identified using BLAST alone. Syntenolog Search inferred orthologous relationships for a further 22,708 oomycete genes, resulting in a final pillar count of 124,060. Thus, on average each pillar in OGOB has 2.58 genes.

Tandem Gene Duplications

Gene duplication is a very common occurrence in eukaryotic species and is one of the main mechanisms by which species acquire new genes and potentially new functions (Kaessmann 2010). Tandem gene duplication occurs when duplicated genes are located adjacent to each other in the genome. Genes that arose via tandem duplication can subsequently undergo chromosomal rearrangement and become dispersed throughout the genome. Such occurrences are more difficult to identify. We set out to identify clusters of tandemly duplicated genes in each oomycete genome. We defined a tandem cluster as two or more adjacent genes that hit each other in a BLASTp search with an E-value cut-off of $1e^{-10}$ and a highest scoring pair (HSP) length greater than half the length of the shortest sequence.

In total 12,541 tandem clusters, corresponding to 29,717 genes, were identified across the 20 oomycete species (**Table 2, Supplementary Table 1**). The overall average number of genes per tandem cluster is 2.3 (**Table 2, Supplementary Table 1**). *Ph. sojae* has the highest number of tandem clusters with 1,389 tandemly duplicated clusters, which corresponds to 3,411 genes or 12.83% of the total genes (**Table 2**). The obligate biotrophic species *Pl. halstedii*, *Al. laibachii* and *Al. candida* have the smallest number of tandem clusters (149, 135, 195 clusters respectively) and also have the smallest proportions of their proteome belonging to tandem clusters (2.02%, 2.19% and 3.91% respectively) (**Table 2**). Tandemly duplicated genes in *Sa. diclina* represent the highest proportion of the proteome (15.35% corresponding to 2,664 genes) (**Table 2**) compared to the other species.

Table 2. Oomycete tandem duplication analysis

Species	Total Genes	Tandem Clusters	Genes in Tandem Clusters	% Total Genes	Avg. # of Genes per Cluster
Ph. infestans	17,797	802	2,002	11.25%	2.50
Ph. parasitica	23,121	925	2,294	9.92%	2.48
Pl. halstedii	15,469	149	312	2.02%	2.09
Ph. capsici	19,805	852	2,095	10.58%	2.46
Hy. arabidopsidis	14,321	876	1,781	12.44%	2.03
Ph. sojae	26,584	1,389	3,411	12.83%	2.46
Ph. ramorum	15,743	885	2,210	14.04%	2.50
Ph. kernoviae	9,923	267	600	6.05%	2.25
Pp. vexans	11,958	369	814	6.81%	2.21
Py. iwayamai	14,869	284	624	4.20%	2.20
Py. irregulare	13,805	363	829	6.01%	2.28
Py. ultimum	15,290	735	1,871	12.24%	2.55
Py. arrhenomanes	13,805	332	716	5.19%	2.16
Py. aphanidermatum	12,312	443	1,003	8.15%	2.26
Al. laibachii	12,567	135	275	2.19%	2.04
Al. candida	10,698	195	418	3.91%	2.14
Sa. diclina	17,359	1,080	2,664	15.35%	2.47
Sa. parasitica	20,088	1,103	2,717	13.53%	2.46
Ap. invadans	15,248	589	1,318	8.64%	2.24
Ap. astaci	19,119	768	1,763	9.22%	2.30

We set out to identify biological functions that are enriched or under-represented in tandemly duplicated clusters for each species. This was achieved by comparing the frequency of Gene Ontology (GO) Slim terms in tandem clusters relative to the non-tandemly duplicated proportion of the proteome using the Fisher exact test, corrected for false discovery rate (FDR) using the Benjamini-Hochberg procedure. Here we report corrected P-values < 0.05 as significant. Our results show enrichment in tandem clusters

for a number of GO Slim terms in each species (**Supplementary Table 2**), except for the two *Albugo* species, where no GO term was detected as being enriched or purified. We detected enrichment for terms related to transport, including establishment of localization (GO:0051234; 12 species), transmembrane transport (GO:0051234; 16 species) and transmembrane transporter activity (GO:0022857; 14 species) (**Supplementary Table 2**). As with previous analyses (Martens and Van de Peer 2010), we also detected enrichment for terms that are potentially involved in pathogenicity such as extracellular region (GO:0005576; 15 species), hydrolase activity, acting on glycosyl bonds (GO:0016798; 13 species), carbohydrate metabolic process (GO:0005975; 9 species), catalytic activity, acting on a protein (GO:0140096; 8 species), hydrolase activity (GO:0016787; 8 species) and peptidase activity (GO:0008233; 8 species) (**Supplementary Table 2**). Significantly, we also detected enrichment of proteins that contain signal peptides in tandem clusters for all species (**Supplementary Table 2**) suggesting that tandem duplication events may be a major driving force for the evolution and expansion of secreted oomycete effectors.

Our analysis detected a number of terms related to housekeeping functions that are significantly under-represented in tandem clusters (**Supplementary Table 2**), including intracellular part (GO:0044424; 17 species), nucleic acid metabolic process (GO:0090304; 16 species), intracellular organelle (GO:0043229; 15 species), nucleic acid binding (GO:0003676; 15 species), biosynthetic process (GO:0009058; 15 species), translation (GO:0006412; 14 species), ribosome (GO:0005840; 14 species), DNA metabolic process (GO:0006259; 13 species), RNA metabolic process (GO:0016070; 12 species), tRNA metabolic process (GO:0006399; 10 species), ncRNA metabolic process (GO:0034660; 10 species) and RNA binding (GO:0003723; 10 species). The majority of these terms describe cellular “housekeeping” genes that are usually members of large

protein interaction networks. In yeast, these categories of genes have been shown to be recalcitrant to gene duplication as they interfere with highly constrained cellular systems and the dosage-balance hypothesis predicts that selection will remove these from populations (Papp et al. 2003; He and Zhang 2006; Li et al. 2006). Similarly in angiosperms single-copy genes are often involved in essential housekeeping functions that are highly conserved across all eukaryotes and are also resistant to duplication (De Smet et al. 2013).

Sequence similarity searches are not sufficient to identify highly divergent tandem duplicates. Using synteny information hosted by OGOB, it is possible to use slowly evolving tandem duplicates in one species to identify rapidly evolving tandems in other species that are so divergent that they cannot be identified by BLAST homology searches. For example, consider the tandem cluster of 4 *Ph. infestans* genes (PITG_01020, PITG_01022, PITG_01023 and PITG_01024) which have the elicitin Pfam domain (PF00964). This tandem cluster is conserved in *Ph. parasitica* and *Ph. sojae* but in *Ph. capsici* only two members were identified as tandemly duplicated (**Figure S4A**). Upon manual inspection in OGOB we see orthologs in *Ph. capsici* to the two remaining tandem duplicates. These were not defined as tandemly duplicated as they did not meet our initial BLAST criteria, both genes are considerably longer that the orthologs in the other three species so violated the HSP cut-off. Furthermore, by comparing tandem duplicates between species, we can use OGOB to identify genes that arose via tandem duplication and later dispersed elsewhere in the genome. For example, *Ph. infestans* contains a cluster of 5 tandemly duplicated sugar efflux transporters (PITG_04998 – PITG_05002). This cluster is syntenically conserved in *Ph. parasitica*, *Ph. capsici*, *Ph. ramorum* and *Ph. sojae* (**Figure S4B**). However, 2 members of the *Ph. sojae* tandem cluster have relocated to another area on the same scaffold (**Figure S4B**).

It should be noted that it is well known that short read genome assemblers are prone to collapse tandemly repeat regions of the genome, therefore the assembler incorrectly joins reads from distinct chromosomal regions into a single unit (Phillippy et al. 2008). This in turn may result in an underestimation of the number of tandemly repeated genes. Long read sequencing technologies have the potential to overcome these issues and can produce gold-standard de novo genome assemblies. Until these gold-standard oomycete genomes become available the numbers presented above should be viewed as a conservative estimate.

The Oomycete Paranome

We also identified the paranome for each oomycete species, that is the set of all paralogous multigene families. *Saprolegnia parasitica* has the highest number of multigene families (3,010), whereas *Ph. kernoviae* (757) has the lowest number (**Table 3**). *Phytophthora ramorum* has the lowest number (5,827) of genes that do not belong to multigene families, whereas *Ph. parasitica* has the highest (11,922) (**Table 3**). The proportion of genes that belong to multigene families varies greatly between oomycete species. In *Pl. halstedii*, only 28% of genes belong to a multigene family, whereas 63% of *Ph. sojae* genes belong to multigene families (**Table 3**). On average, approximately 45.5% of oomycete genes belong to a multigene family. The average number of genes in each family ranges from 3.3 to 6.2, with an overall average of 4.6 genes per family (**Table 3**).

We also carried out a GO enrichment analysis to determine if any GO terms are over or under-represented in the paranome of each species. This identified enrichment of terms including ion binding (GO:0043167; 20 species), ATPase activity (GO:0016887; 20 species), cellular proteins modification process (GO:0006464; 20 species), and

regulation of cellular process (GO:0050794; 9 species) (**Supplementary Table 2**). Similar to our tandem duplication analysis, we see enrichment of terms related to transport (GO:0006810; 17 species) and establishment of localization (GO:0051234; 17 species) (**Supplementary Table 2**). We also see enrichment of terms potentially involved in pathogenicity including carbohydrate metabolic process (GO:0005975; 20 species) and extracellular region (GO:0005576; 6 species) (**Supplementary Table 2**). In terms of underrepresented GO terms in the oomycete paranome, our results largely match that of tandem clusters. We see GO terms related to housekeeping functions underrepresented in the paranome of most species, including cytoplasmic part (GO:0044444; 20 species), RNA processing (GO:0006396; 20 species), cellular component organisation (GO:0016043; 20 species), nucleus (GO:0016043; 20 species), RNA binding (GO:0003723; 20 species), translation (GO:0006412; 19 species), ribosome (GO:0005840; 19 species), ribosome biogenesis (GO:0042254; 18 species), nuclease activity (GO:0004518; 17 species) and cell cycle (GO:0007049; 14 species) (**Supplementary Table 2**). Our results above are in line with previous analyses of *Phytophthora* and *Pythium* species that have shown that pathogenicity related genes are typically expanded relative to genes not directly linked to pathogenicity (Tyler 2006; Haas et al. 2009; Lévesque et al. 2010).

Table 3. Oomycete paranome analysis

								Multigene Family Size		
Species	**Genes**	**Single Copy Genes**	**% Single Copy Genes**	**Genes in Multigene Families**	**% Genes in Multigene Families**	**Multigene Families**	**Avg # Genes Per Family**	**2 members**	**3 members**	**4 members**	**≥5 members**
Ph. infestans	17,797	8,097	45.50%	9,700	54.50%	2,167	4.48	12.16%	6.39%	4.81%	31.15%
Ph. parasitica	23,121	11,922	51.56%	11,199	48.44%	2,162	5.18	9.42%	4.90%	2.99%	31.12%
Pl. halstedii	15,469	11,139	72.01%	4,330	27.99%	1,054	4.11	7.01%	4.09%	2.33%	14.56%
Ph. capsici	19,805	8,310	41.96%	11,495	58.04%	2,007	5.73	8.95%	5.12%	4.20%	39.77%
Hy. arabidopsidis	14,321	8,743	61.05%	5,578	38.95%	1,645	3.39	15.57%	5.36%	2.32%	15.70%
Ph. sojae	26,584	9,776	36.77%	16,808	63.23%	2,723	6.17	9.25%	4.83%	3.64%	45.50%
Ph. ramorum	15,743	5,827	37.01%	9,916	62.99%	1,755	5.65	11.13%	5.68%	3.96%	42.22%
Ph. kernoviae	9,923	6,730	67.82%	3,193	32.18%	757	4.22	8.00%	3.60%	2.82%	17.76%
Pp. vexans	11,958	7,025	58.75%	4,933	41.25%	1,163	4.24	9.82%	5.75%	3.38%	22.31%
Py. iwayamai	14,869	8,876	59.69%	5,993	40.31%	1,318	4.55	9.24%	4.50%	2.74%	23.82%

Table 3. Continued.

								Multigene Family Size		
Species	**Genes**	**Single Copy Genes**	**% Single Copy Genes**	**Genes in Multigene Families**	**% Genes in Multigene Families**	**Multigene Families**	**Avg # Genes Per Family**	**2 members**	**3 members**	**4 members**	**≥5 members**
Py. irregulare	13,805	7,966	57.70%	5,839	42.30%	1,240	4.71	8.68%	4.89%	3.13%	25.60%
Py. ultimum	15,290	8,761	57.30%	6,529	42.70%	1,313	4.97	7.89%	4.53%	3.22%	27.06%
Py. arrhenomanes	13,805	8,231	59.62%	5,574	40.38%	1,260	4.42	9.03%	4.67%	3.53%	23.14%
Py. aphanidermatum	12,312	7,360	59.78%	4,952	40.22%	1,173	4.22	9.36%	5.48%	3.02%	22.36%
Al. laibachii	12,567	6,842	54.44%	5,725	45.56%	1,536	3.73	16.95%	4.92%	2.96%	20.73%
Al. candida	10,698	7,203	67.33%	3,495	32.67%	1,049	3.33	12.68%	4.43%	2.99%	12.57%
Sa. diclina	17,359	8,711	50.18%	8,648	49.82%	1,700	5.09	9.76%	4.96%	3.39%	31.71%
Sa. parasitica	20,088	7,966	39.66%	12,122	60.34%	3,010	4.03	18.66%	6.54%	3.94%	31.20%
Ap. invadans	15,248	8,955	58.73%	6,293	41.27%	1,389	4.53	9.80%	4.19%	3.46%	23.82%
Ap. astaci	19,119	9,875	51.65%	9,244	48.35%	1,700	5.44	8.65%	4.53%	3.54%	31.63%

Phylostratigraphy Analysis

To further elucidate oomycete genome evolution we carried out a phylostratigraphic analysis of each oomycete species housed in OGOB. Phylostratigraphy is a statistical approach for reconstructing macroevolutionary transitions by identifying the evolutionary emergence of founder genes across the tree of life (Domazet-Lošo et al. 2007; Tautz and Domazet-Lošo 2011; Sestak and Domazet-Loso 2015). Here we apply phylostratigraphy to compare the composition of the 20 oomycete genomes hosted on OGOB. We estimate the age and emergence of all 319,881 oomycete genes by identifying their founder genes in a database containing species across the tree of life. We generated our database by merging all sequences hosted on OGOB with those from a previous analysis with broad phyletic distribution (Drost et al. 2015), resulting in final database of 17,826,795 amino acid sequences. Each oomycete gene was searched against the database using BLASTp (E-value cut-off $1e^{-5}$) and genes were assigned to a phylostratum based on their most ancient hit.

In total for the 20 species, 104,662 genes (32.7%) were placed at the origin of cellular organisms (i.e. homologs were identified in bacteria or archaea), 92,218 genes (28.8%) were eukaryotic in origin, 5,355 genes (1.7%) arose in the stramenopiles and 65,015 (20.3%) arose in the oomycetes (**Figure 3A – D**). The remaining genes were assigned as unique to particular oomycete lineages, 26,090 (8.2%) of which were determined to be unique to individual species (orphan genes). By comparing phylostratigraphic maps for each species we can identify macroevolutionary trends. The overall trends identified for each oomycete order are largely similar. With few exceptions, genes of ancient Prokaryotic origin represent the largest proportion of each oomycete genome (**Figure 3A – D**). A slightly smaller proportion arose in the evolution of Eukaryotes (**Figure 3A – D)**. *Ph. capsici* and *Al. laibachii* are exceptions, whereby more

genes were identified as Eukaryotic in origin than Prokaryotic (**Figure 3A, Figure 3C**). Very few genes (between 159 and 382 genes in each species) were detected to have arisen during the evolution of stramenopiles (**Figure 3A – D**). Our results suggest that either very few genes were gained during the evolution of stramenopiles or else a large number of genes of stramenopiles origin were later lost. Following this, in each species we see a large burst of oomycete genes being formed (**Figure 3A – D**). On average, genes of oomycete origin correspond to 21.34% of Peronosporales genomes (**Figure 3A**), 26.24% of *Pythium* genomes (**Figure 3B**), 14.68% of *Albugo* genomes (**Figure 3C**) and only 12.53% of Saprolegniales genomes (**Figure 3D)**. *Ph. sojae* has more genes of oomycete origin than any other species, making up the largest proportion of its total genes (27.62%), suggesting large scale duplication of genes of oomycete origin (**Figure 3A**). For species in the Peronosporales order (**Figure 3A**) we see that genes of Peronosporales origin correspond to very few genes (between 163 genes in *Py. vexans* to 1,052 genes in *Ph. sojae*). Additionally, genes of *Phytophthora* origin represents only a small proportion of *Phytophthora* genomes, corresponding to an average of only 6.15% of genes (**Figure 3A**). We see bursts of emerging orphans in the downy-mildews (*Hy. arabidopsidis* and *Pl. halstedii*) and also in some *Phytophthora* species (*Ph. parasitica* and *Ph. sojae*) (**Figure 3A**). 793 genes (4.46%) in *Ph. infestans* were identified as unique to *Ph. infestans*. The same is true for *Pythium* species as very few genes were assigned to the *Pythium* phylostratum (between 124 and 490 genes) (**Figure 3B**). Furthermore, we see very few orphan genes in each *Pythium* species (between 167 and 557 genes) (**Figure 3B)**, suggesting there is not a great deal of gene content diversity. This is in contrast to the Albuginales and Saprolegniales species were larger numbers of genus specific and species specific genes were detected (**Figure 3C, Figure 3D)**.

Perhaps most interesting are orphan genes as these do not have homologs in any other species (at least within our dataset) and may represent evolutionary novelty. *Ph. ramorum* has the lowest number of orphans (150 genes; < 1% total genes), whereas *Pl. halstedii* has the most (4,830 genes; 31% total genes) (**Figure 3A**). *Hy. arabidopsidis* also has a very high number of orphans (4,132 genes; 29% total genes). On average, each oomycete genome in our dataset has 1,305 orphan genes. In general, the Peronosporales and the Albuginales tend to have more orphan genes (typically more than 1,700 orphans) (**Figure 3A, Figure 3D**) which may correspond to greater functional diversity in terms of gene content and in turn, greater diversity between species in terms of pathogenicity and host range.

Our phylostratigraphy approach allows us to account for differences in the number of genes each species has. For example, *Al. laibachii* has more genes than its closest relative *Al. candida* (12,567 vs 10,698 genes). *Albugo laibachii* has 4,077 genes that were assigned to the eukaryotic node, these genes are distributed across 2,549 gene families. Similarly, *Al. candida* has 2,923 eukaryotic genes located across 2,378 families. Furthermore, at the oomycete node, *Al. laibachii* has 2,000 genes that are grouped into 1,400 families while *Al. candida* has 1,440 grouped into 1,250 families. The number of orphan genes found in both species is very similar (2,334 vs 2,238). Therefore, *Al. laibachii* has more genes than *Al. candida* not because of the de novo formation of orphan genes but, rather, it has more copies of genes that can be mapped back to the eukaryotic and oomycete nodes (**Figure 3C**). These differences are due to retention and expansion of gene families from these nodes in *Al. laibachii*.

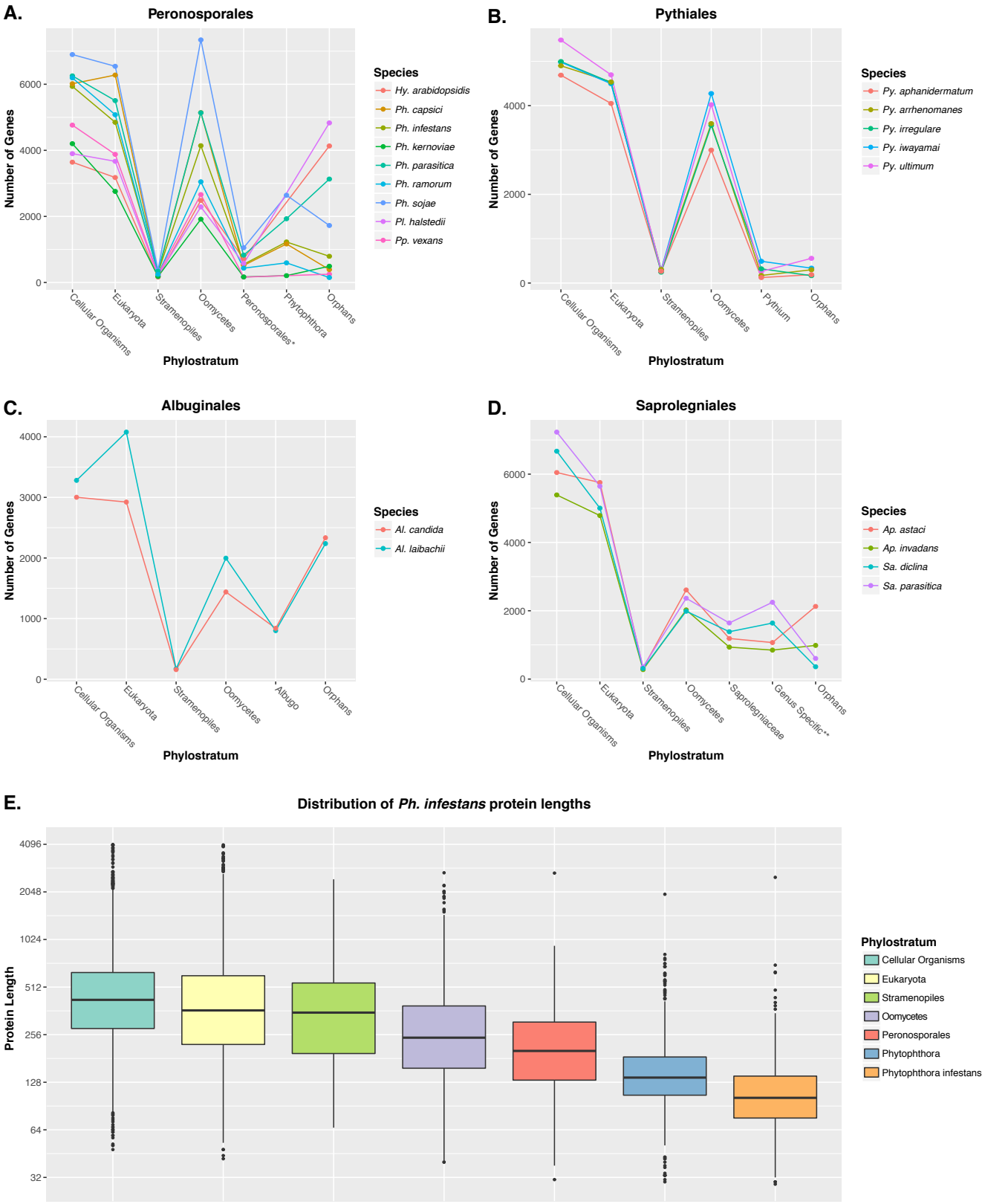


Figure 3. Phylostratigraphic analysis of 20 oomycete species to determine the number of founder genes that have arisen in each phylostratum in Peronosporales (**A**), Pythiales (**B**), Albuginales (**C**) and Saprolegniales (**D**).

Figure 3 (continued). * 20 *Hy. arabidopsidis* and 18 *Pl. halstedii* genes were identified as unique to the Peronosporaceae family, for visualisation purposes we have moved these hits to the Peronosporales phylostratum. ** genes that were identified as specific to either the *Aphanomyces* genus or the *Saprolegnia* genus. (**E**) Distribution of protein lengths in *Ph. infestans* across phylostrata shows a continuous increase in length with evolutionary age.

A similar trend can be seen when we compare the gene content of *Ph. sojae* and *Ph. kernoviae* (26,548 vs 9,923 genes). *Ph. sojae* has 6,543 genes of eukaryotic origin distributed across 3,752 families, while *Ph. kernoviae* only has 2,757 genes of eukaryotic origin belonging to 2,331 families. Furthermore, *Ph. sojae* has 7,342 genes of oomycete origin distributed across 3,416 families whereas *Ph. kernoviae* only has 1,915 genes of oomycete origin distributed across 1,664 families. It should be noted however that *Ph. sojae* has an additional 1,233 orphan genes relative to *Ph. kernoviae* (1,725 vs 492).

Comparisons of average protein length across phylostrata in *Ph. infestans* reveals that the length of proteins increases with evolutionary age (**Figure 3E)**. Proteins found at the youngest phylostratum (i.e. orphan genes) are the shortest, and protein length increases across each older phylostrata, with the longest proteins being found at the oldest phylostratum (Cellular Organisms) (**Figure 3E**). A similar trend was identified in all species in our dataset (**Figure S3**). This has also been observed in other species such as yeast (Carvunis et al. 2012), *Arabidopsis thaliana* (Guo 2013) and metazoa (Neme and Tautz 2013), suggesting that similar evolutionary pressures are influencing genome and molecular evolution across distantly related eukaryotic species.

The Core Oomycete Ortholog Gene Set

We used OGOB's orthology pillars to identify core oomycete genes. We define core genes as the set of orthologs that are present in all species (i.e. pillars with 20 genes). We also define syntenologs as core orthologs that are microsyntenically conserved in all species (i.e. pillars with 20 core orthologs whereby each gene is microsyntenically conserved with every other gene). Our analysis revealed 1,835 core oomycete pillars. Thus, on average 12% of all oomycete genes have an ortholog in every other species (**Supplementary Table 3**). Only 37 syntenolog pillars (2% or core pillars) were identified (**Supplementary Table 3**). Oomycete syntenologs correspond to an average of only 0.25% of total genes in a genome. However, this is a very strict approach as each ortholog must be microsyntenically conserved with every other gene. Furthermore, the fragmented nature of some of the assemblies in OGOB may contribute to this low number. We repeated this analysis to identify core orthologs and syntenologs individually in the Peronosporales order, the Saprolegniales order, the *Pythium* genus and the *Albugo* genus.

Overall we identified 4,063 core pillars in the 8 Peronosporales species, of which 2,279 (56%) belong to the syntenolog category, corresponding to between 8.57% and 22.97% of total genes in Peronosporales species (**Supplementary Table 3**). Core pillars (6,483) were identified for the five species in the *Pythium* genus, of which 2,863 (44.16%) belong to the syntenolog category (**Supplementary Table 3**). This corresponds to between 18.72% and 23.25% of total genes in *Pythium* species. Analysis of the four species in the Saprolegniales order revealed an even more extensive degree of syntenic conservation, 8,910 core pillars were identified, of which 7,718 (86.62%) belong to the syntenolog category (**Supplementary Table 3**). This corresponds to between 38.42% and 50.62% of total genes in Saprolegniales genomes being both ubiquitous and microsyntenically conserved. The highest degree of synteny was detected in the *Albugo*

genus where 6,719 core pillars and 6,313 syntenolog pillars were detected (**Supplementary Table 3**). This means that 93.96% of orthologs within *Albugo* are microsyntenically conserved. This corresponds to between 50.23% and 59% of total *Albugo* genes. This result may be biased, however, as there are only two closely related *Albugo* genomes in OGOB.

For each group of species (Peronosporales, Saprolegniales, *Pythium* and *Albugo*) we carried out an enrichment analysis of syntenically conserved core orthologs by comparing the frequency of GO terms associated with genes found in syntenolog pillars relative to non-syntenolog pillars. As expected, our results show that syntenically conserved genes are enriched for housekeeping functions including GO terms such as ribosome (GO:0005840), translation (GO:0006412), cellular macromolecule biosynthetic process (GO:0034645), amide biosynthetic process (GO:0043604), RNA binding (GO:0003723), nucleus (GO:0005634), nucleolus (GO:0005730) (**Supplementary Table 2**). These findings are consistent with the hypothesis that oomycete genomes contain "gated communities" where conserved and housekeeping genes reside (Bhowmick and Tripathy 2014). Terms underrepresented in syntenic orthologs include establishment of localization (GO:0051234), carbohydrate metabolic process (GO:0005975), transmembrane transport (GO:0055085), extracellular region (GO:0005576) and ATPase activity (GO:0016887) (**Supplementary Table 2**).

To fine tune our analysis even further, we also investigated the degree of microsynteny in each possible pair of the 20 oomycete species. For each pair of species, we identify orthologs and quantify the proportion that are microsyntenically conserved. We consider a pair of orthologs to be microsyntenically conserved if there exists another pair of orthologs within a distance of 20 genes. Unsurprisingly, our results reveal very high levels of microsynteny between closely related species within oomycete orders, and

a breakdown in synteny between more distantly related species across orders (**Supplementary Table 4**). We use the proportion of microsyntenically conserved genes to generate a distance matrix and use this to cluster species based on microsyntenic conservation (**Figure 4**). As expected, more closely related organisms share a higher degree of microsynteny and are clustered together into their orders and genera (**Figure 4**). When comparing any two oomycete species the proportion of orthologs that are microsyntenically conserved is between 27.57% and 96.39% (**Supplementary Table 4**). *Sa. diclina* and *Sa. parasitica* share the highest degree of microsynteny (82.41% of genes or 96.35% of orthologs), followed by *Ap. astaci* and *Ap. invadans* (75.55% of genes of 96.39% of orthologs) (**Figure 4, Supplementary Table 4**). The two species showing the lowest level of microsynteny are *Ap. astaci* and *Py. arrhenomanes* (10.34% of genes or 28.01% of orthologs) (**Figure 4, Supplementary Table 4**). On average 33.90% of total genes or 64.91% of orthologs are microsyntenic when comparing any two oomycete species in OGOB (**Supplementary Table 4**). It is worth noting that we also performed the above analysis with more restrictive window sizes (i.e., a window size of 5 instead of 20) and results were largely congruent (not shown).

Our results are largely in agreement with previous analyses. For example, a previous study determined that over 75% of exons in *Ph. ramorum* and *Ph. sojae* aligned in a whole genome alignment (Tyler 2006). Here, we find that 11,070 orthologs are shared between *Ph. ramorum* and *Ph. sojae*, of which 10,158 (91.76%) were detected to be microsyntenic. This corresponds to approximately 65% of genes in *Ph ramorum*. The *Ph. infestans* genome paper reports that 90% of orthologs shared by *Ph. infestans*, *Ph. ramorum* and *Ph. sojae* are found in blocks of conserved gene order (Haas et al. 2009). Our results are in agreement with this. We identified 9,219 orthologs that are present in all 3 species, of which 8,322 are microsyntenically conserved (90%). Another analysis

reported extensive synteny between *Py. ultimum* and *Phytophthora* species (Lévesque et al. 2010). We detect that up to 94% of orthologs between *Py. ultimum* and any *Phytophthora* species are syntenologs (**Supplementary Table 4**). Surprisingly, our results suggest that there is a greater degree of microsyntenic conservation between *Phytophthora* species and *Py. ultimum* than between *Phytophthora* species and *Hy. arabidopsidis* or *Pl. halstedii* (**Figure 4**). This may be due to gene loss events or extensive genome rearrangements in the evolution of obligate biotrophy in these downy mildew species (Baxter et al. 2010).

When comparing microsynteny between orders, the Saprolegniales species are most divergent from other species. On average only 40.70% of orthologs or 16.05% of total genes are microsyntenically conserved when compared to species outside of the order (**Figure 4, Supplementary Table 4**). This is not surprising as the Saprolegniales are thought to have diverged from other oomycetes approximately 200 million years ago (Matari and Blair 2014). However, species within the Saprolegniales order have the highest degree of microsyntenic conservation, on average 91.68% of their orthologs are microsyntenically conserved or 61% of total genes. When comparing any two Peronosporales species, between 73.79% and 94.9% of orthologs are syntenically conserved. This corresponds to between 34.8% and 68.03% of total genes (**Supplementary Table 4**). In *Pythium* species this range is 54.72% to 91.23% of orthologs or 28.82 to 68.63% of total genes (**Supplementary Table 4**).

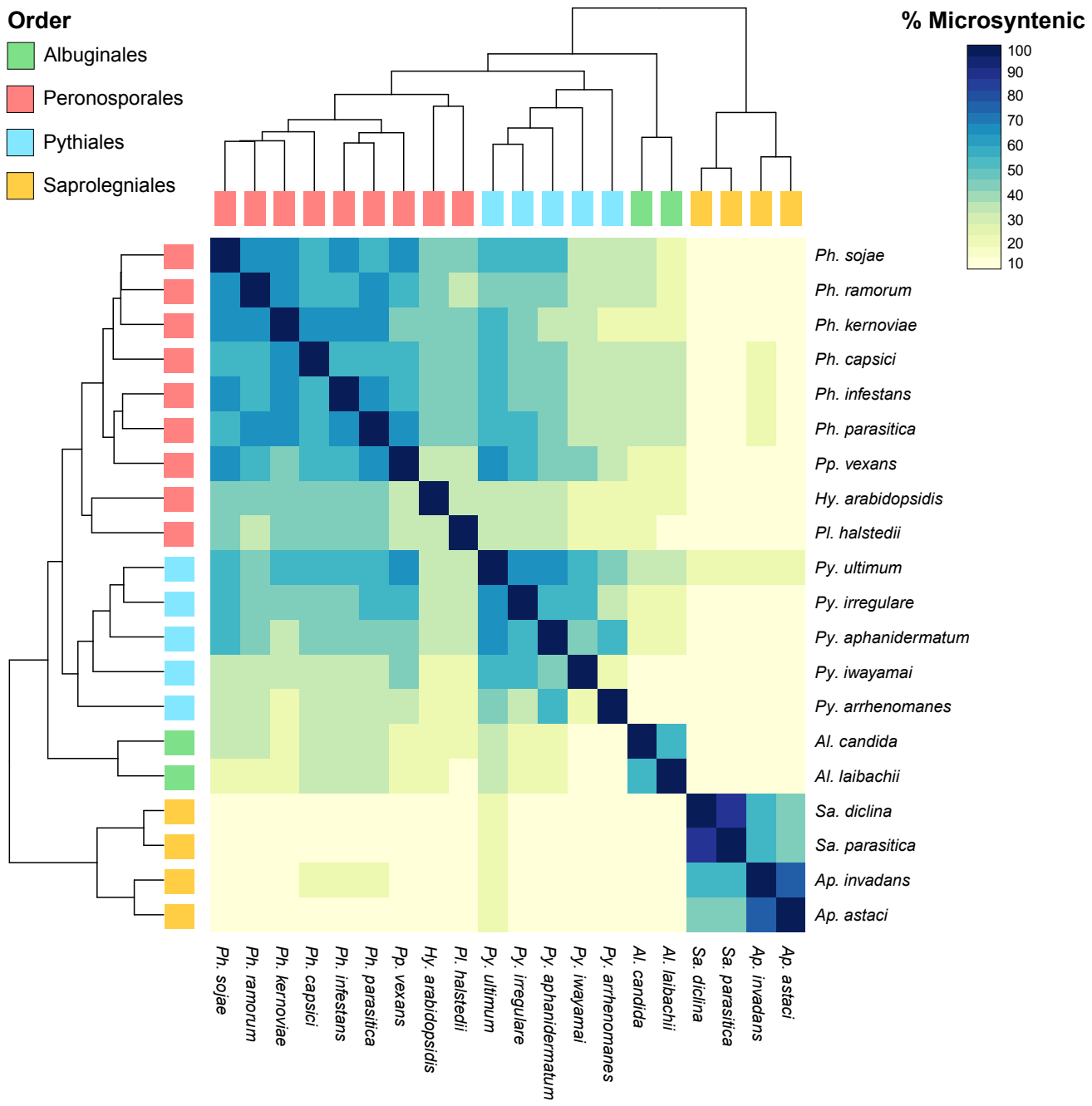


Figure 4. Pairwise microsyntenic analysis of oomycete species. Heatmap values represent the total proportion of genes in the smallest genome that were identified as syntenologs. These values are used to cluster the species based on microsynteny.

Using OGOB to Visualise Expansions of Necrosis-Inducing Proteins

Necrosis-inducing proteins (NLPs) are apoplastic effectors found in bacteria, fungi, and oomycetes (Feng et al. 2014). NLPs are known to induce necrosis, elicit immune responses, and trigger ethylene accumulation in dicotyledons (Oome and Van den Ackerveken 2014). Proteins containing the NLP PFAM domain (PF05630) are significantly overrepresented in numerous *Phytophthora* and Pythiales species but are completely absent from *Albugo*, *Aphanomyces*, and *Saprolegnia* species (McGowan and Fitzpatrick 2017). NLPs are highly expanded in *Phytophthora* species. In particular, *Ph. capsici*, *Ph. ramorum, Ph. parasitica*, and *Ph. sojae* have 65, 69, 74 and 80 putative proteins with NLP domains (McGowan and Fitzpatrick 2017).

Using OGOB it is possible to visualise the mechanisms that are partly responsible for the expansions of NLPs in these *Phytophthora* species (**Figure 5**). There are numerous genomic loci where tandem duplications have given rise to clusters of NLP paralogs in selected *Phytophthora* species. For example, *Ph. parasitica* has five NLP paralogs (PPTG_07660, PPTG_07661, PPTG_07664, PPTG_07667, and PPTG_07668) clustered together on scaffold 12 in a window of 10 genes (**Figure 5A**). Closer examination shows that PPTG_07660 & PPTG_07661 and PPTG_07667 & PPTG_07668 are tandem duplicates as all have an orange coloured BLAST ("b") button associated with them. Orthologs for these five genes are present in *Ph. sojae* and *Ph. ramorum* and high levels of synteny are observed (**Figure 5A**). Orthologs are absent in all Pythiales species (not shown). Similarly, *Ph. parasitica* contains a tandem array of five NLP paralogs (PPTG_15230, PPTG_15231, PPTG_15233, PPTG_15234, and PPTG_15235) on scaffold 48 (**Figure 5B**). A number of orthologs are present in other *Phytophthora* species and synteny around this array is generally well conserved (**Figure 5B**). Orthologs for PPTG_15230 and PPTG_15231 are observed in the majority of Pythiales species but levels of synteny are generally low (not shown).

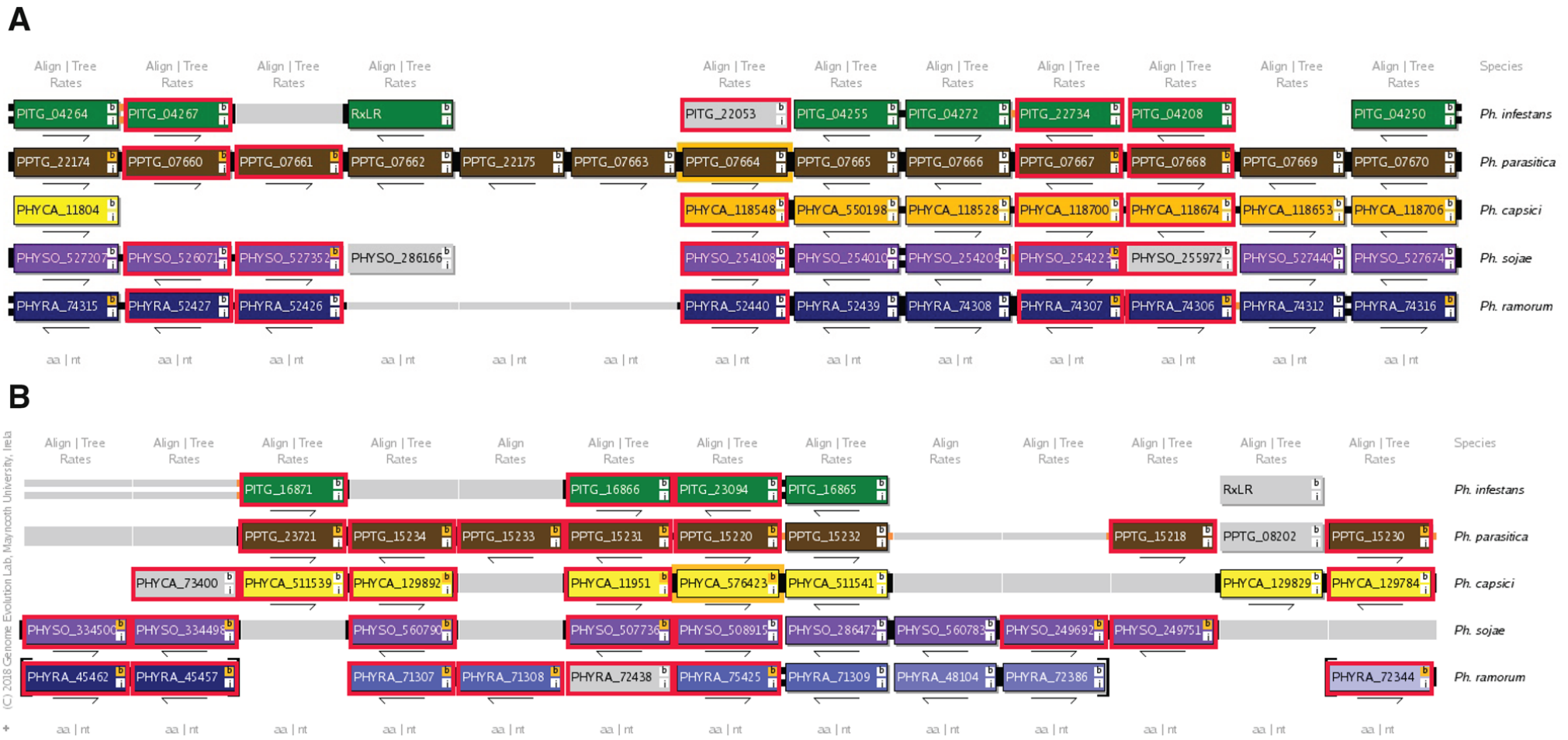


Figure 5. Two loci with multiple NLPs. Orthologs that contain a PFAM NLP domain (PF05630) are indicated with a red box. For display purposes only *Ph. infestans*, *Ph. parasitica*, *Ph. capsici*, *Ph. sojae*, and *Ph. ramorum* are shown. *Albugo*, *Aphanomyces*, and *Saprolegnia* species lack proteins with this domain. (**A**) Screenshot from OGOB, browser centred around *Ph. parasitica* NLP domain containing ortholog (PPTG_07664). PPTG_07660 and PPTG_07661 are tandem duplicates (signified by yellow "b" button) as are PPTG_07667 and PPTG_07668. Orthologs for PPTG_07667 and PPTG_07668 are observed in all species displayed and synteny is relatively conserved except for the ortholog of PPTG_07668 in *Ph. sojae* (PHYSO_255972). *Phytophthora capsici* is missing PPTG_07660 and PPTG_07661 orthologs while *Ph. infestans* is missing the PPTG_07661 ortholog. Orthologs are missing in all Pythiales species (not shown). (**B**) Screenshot from OGOB, browser centred around a *Ph. parasitica* NLP domain containing ortholog (PHYCA_576423). *Phytophthora parasitica* contains a tandem array of five NLP paralogs (PPTG_15230, PPTG_15231, PPTG_15233, PPTG_15234, and PPTG_15235). Orthologs of PPTG_15230 and PPTG_15231 are present in the majority of Pythiales species (not shown).

Conclusion

We report here the development of OGOB, a database and tool for performing comparative genomic and synteny analyses of oomycete species. We highlight the usefulness of synteny information in identifying orthologs and use synteny to identify orthologous relationships for 22,708 genes that could not be identified using BLAST searches alone. Phylostratigraphy was used to determine the composition of 20 oomycete genomes and estimate the evolutionary age and emergence of 319,881 oomycete genes. The extent of gene duplication was determined and tandem duplication events were identified as a driving force for the expansion of secreted effector arsenals. Core conserved genes for each oomycete order were identified. Synteny analysis of the 20 oomycete species hosted by OGOB revealed a high degree of syntenic conservation. Our results suggest that conserved genes with housekeeping functions are more likely to be syntenically conserved. Going forward, it is our goal to include additional gold standard genomes from diverse clades in OGOB. For example, currently of the ten recognized *Phytophthora* clades, only data for five clades (clades 1, 2, 7, 8, and 10) are represented. Furthermore, we will also investigate the possibility of implementing robust automated pipelines to locate putative genes that may have been missed at the gene calling stage of annotating genomes. OGOB is a valuable, central resource that will be of interest to plant pathologists and the oomycete community. OGOB is available at https://ogob.ie.

Materials and Methods

OGOB Database Construction

Genomic data for the 20 oomycete species were retrieved from the sources listed in **Table 1**, including genome assemblies and gene sets. Gene sets were manually inspected and dubious gene calls were removed. For genes with alternative transcripts, the longest transcript was retained. The final dataset contains 319,881 protein coding genes. BUSCO v3 (Waterhouse et al. 2018) was used to assess the gene space completeness of each assembly with the alveolata/stramenopiles dataset of BUSCOs. InterProScan 5 (Jones et al. 2014) was run on all 319,881 oomycete proteins in the OGOB database. Proteins were annotated for functional domains using the InterPro (Finn et al. 2017), Pfam (Finn et al. 2016) and PANTHER (Mi et al. 2017) databases, as well as for Gene Ontology terms (Ashburner et al. 2000). Signal peptides were predicted using SignalP (Bendtsen et al. 2004) and transmembrane domains were predicted with TMHMM (Krogh et al. 2001). Functional annotations are displayed on OGOB gene information pages and link back to the original annotation databases. Metabolic pathways were also annotated using the KEGG (Ogata et al. 1999), MetaCyc (Caspi et al. 2018) and Reactome (Fabregat et al. 2018) databases. All annotations can be downloaded from the OGOB data page (https://ogob.ie/gob/data.html).

Phylogenetic Analysis

A maximum-parsimony supertree approach was carried out to generate the oomycete species phylogeny (**Figure 1**). All pillars containing at least 4 genes (17,738 pillars) were retrieved and individually aligned using MUSCLE (Edgar 2004). Individual phylogenies for each of the 17,738 pillars were generated using FastTree v2.1.9 (Price et al. 2010). A supertree was constructed using the Matrix Representation with Parsimony (MRP)

method implemented in Clann (Creevey et al. 2004; Creevey and McInerney 2005) with 100 bootstrap replicates. The phylogeny was visualised and annotated using the Interactive Tree of Life (iTOL)(Letunic and Bork 2007).

Phylostratigraphy Analysis

Individual phylostratigraphic maps for each of the 20 oomycetes were constructed following previously published methods (Quint et al. 2012; Drost et al. 2015). The data set used by Drost et al. (2015) was retrieved. This data set contains amino acid sequences for 4,557 species including 1,787 eukaryotes (883 animals, 364 plants, 344 fungi and 196 other eukaryotes) and 2,770 prokaryotes (2,511 bacteria and 259 archaea). We added all sequences hosted by OGOB to this database, resulting in a final database of 17,826,795 amino acid sequences. Each oomycete protein was searched against this database using BLASTp (Altschul et al. 1997). Each protein is assigned to the oldest phylostrata that contains at least one BLAST hit with an E-value cut-off of $1e^{-5}$. A gene is assigned to the youngest phylostratum (i.e. species-specific orphan) if it does not have any such BLAST hit.

Gene Duplications

Tandemly duplicated genes were identified using BLAST (Altschul et al. 1997). In each genome, every gene was aligned to its adjacent genes. Alignments with an E-value below $1e^{-10}$ and a highest scoring pair (HSP) length greater than half the length of the shortest sequence were considered tandemly duplicated.

Multigene families were identified for each species by performing all-versus-all BLASTp searches (Altschul et al. 1997) of each gene against every other gene in its

genome with an E-value cut-off of $1e^{-30}$, followed by Markov clustering using MCL (Enright et al. 2002) with an inflation value of 1.5.

Gene Enrichment Analyses

Gene enrichment analysis were performed using Fisher's exact test. SignalP v3 (Bendtsen et al. 2004) was used to predict signal peptides for enrichment analyses of secreted proteins. We used SignalP v3 instead of later versions of the software as previous studies have found v3 to be the most sensitive in identifying oomycete signal peptides (Sperschneider et al. 2015). Transmembrane domains were predicted using TMHMM (Krogh et al. 2001). Proteins were considered secreted if they had an HMM S probability value ≥ 0.9, an NN Y_{max} score of ≥ 0.5, and an NN D score of ≥ 0.5 with predicted localization "Secreted" and no transmembrane domain after the signal peptide cleavage site. Enrichment tests of GO Slim terms were carried out using goatools (Klopfenstein et al. 2018) with Benjamini-Hochberg correction for FDR. Corrected P-values < 0.05 were considered significant.

Acknowledgements

We acknowledge the DJEI/DES/SFI/HEA Irish Centre for High-End Computing (ICHEC) for the provision of computational facilities and support. J.M. is funded by a postgraduate scholarship from the Irish Research Council, Government of Ireland (grant number GOIPG/2016/1112).

Supplementary Data

Supplementary data for this chapter is available from the publisher website: https://academic.oup.com/gbe/article/11/1/189/5238079

Supplementary data is also available at the following GitHub repository: https://github.com/jamiemcg/ThesisSupplementaryMaterial

Supplementary Figures

Supplementary Figure 1. Genome completeness assessment of oomycete genomes using BUSCO with the Alveolata-Stramenopile set (234 BUSCOs in total).

Supplementary Figure 2. PITG_00248 and PPTG_10928 (labelled with asterisks) were identified as orthologous using Syntenolog Search. **(A)** Screenshot of OGOB showing that PITG_00248 and PPTG_10928 are syntenically conserved. (**B**) BLAST search results of PPTG_10928 against the OGOB databases. PPTG_10928 has a significant ($1e^{-125}$) but not reciprocal best hit to PITG_00248. **(C)** MUSCLE multiple sequence alignment of PITG_00248 and PPTG_10928.

Supplementary Figure 3. Distribution of oomycete protein lengths across phylostrata.

Supplementary Figure 4. Synteny information hosted by OGOB can be used to identify rapidly evolving tandem duplications and tandem duplicates that have undergone chromosomal rearrangement. (**A**) A cluster of 4 tandemly duplicated elicitin proteins that are syntenically conserved in *Ph. infestans*, *Ph. parasitica*, *Ph. capsici* and *Ph. sojae*. 2 of the *Ph. capsici* proteins (PHYCA_529852 and PHYCA_509798) did not meet our initial BLAST criteria, however, they are obvious tandem duplicates when syntenic conservation is considered. (**B**) A cluster of 5 tandemly duplicated sugar efflux transporters that are syntenically conserved in *Ph. infestans*, *Ph. parasitica*, *Ph. capsici*,

Ph. sojae and *Ph. ramorum*. However, 2 members of the *Ph. sojae* tandem cluster have relocated to another area on the same scaffold.

Supplementary Tables

Supplementary Table 1. Clusters of tandemly duplicated genes in oomycete species.

Supplementary Table 2. Enrichment analysis of secreted proteins and GO terms in oomycete tandem clusters, the oomycete paranome and core oomycete pillars.

Supplementary Table 3. Analysis of the number of core and syntenolog oomycete, Peronosporales, *Pythium*, *Albugo* and Saprolegniales pillars. For the Peronosporales analysis we excluded *Py. vexans*, despite it being a member of the Peronosporales, it is thought to be an intermediate between *Phytophthora* and *Pythium* species in terms of its gene content and genome organisation.

Supplementary Table 4. Pairwise microsyntenic analysis of oomycete species. Counts of the number of orthologs and syntenologs shared by each possible pair of 20 oomycete species is shown. Also shown is the proportion of orthologs that were identified as syntenologs as well as the proportion of total genes identified as syntenologs.

References

Adhikari BN, Hamilton JP, Zerillo MM, Tisserat N, Lévesque CA, Buell CR. 2013. Comparative Genomics Reveals Insight into Virulence Strategies of Plant Pathogenic Oomycetes. PLoS One 8.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389–3402.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. 2000. Gene Ontology: tool for the unification of biology. Nat. Genet. 25:25–29.

Basenko E, Pulman J, Shanmugasundram A, Harb O, Crouch K, Starns D, Warrenfeltz S, Aurrecoechea C, Stoeckert C, Kissinger J, et al. 2018. FungiDB: An Integrated Bioinformatic Resource for Fungi and Oomycetes. J. Fungi. 4:39.

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 (80-.). 330:1549–1551.

Beakes GW, Glockling SL, Sekimoto S. 2012. The evolutionary phylogeny of the oomycete “fungi.” Protoplasma 249:3–19.

Bendtsen JD, Nielsen H, Von Heijne G, Brunak S. 2004. Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. 340:783–795.

van den Berg AH, McLaggan D, Diéguez-Uribeondo J, van West P. 2013. The impact of the water moulds Saprolegnia diclina and Saprolegnia parasitica on natural ecosystems and the aquaculture industry. Fungal Biol. Rev. 27:33–42.

Bhowmick S, Tripathy S. 2014. A Tale of Effectors; Their Secretory Mechanisms and Computational Discovery in Pathogenic, Non-Pathogenic and Commensal Microbes. Mol. Biol. 03:1–14.

Burki F. 2014. The eukaryotic tree of life from a global phylogenomic perspective. Cold Spring Harb. Perspect. Biol. 6:1–17.

Byrne KP, Wolfe KH. 2005. The Yeast Gene Order Browser: Combining curated homology and syntenic context reveals gene fate in polyploid species. Genome Res. 15:1456–1461.

Byrne KP, Wolfe KH. 2006. Visualizing syntenic relationships among the hemiascomycetes with the Yeast Gene Order Browser. Nucleic Acids Res. 34:D452-5.

Carvunis A-R, Rolland T, Wapinski I, Calderwood MA, Yildirim MA, Simonis N, Charloteaux B, Hidalgo CA, Barbette J, Santhanam B, et al. 2012. Proto-genes and de novo gene birth. Nature 487:370–374.

Caspi R, Billington R, Fulcher CA, Keseler IM, Kothari A, Krummenacker M, Latendresse M, Midford PE, Ong Q, Ong WK, et al. 2018. The MetaCyc database of metabolic pathways and enzymes. Nucleic Acids Res. 46:D633–D639.

Cheung F, Win J, Lang JM, Hamilton J, Vuong H, Leach JE, Kamoun S, André Lévesque C, Tisserat N, Buell CR. 2008. Analysis of the Pythium ultimum transcriptome using Sanger and Pyrosequencing approaches. BMC Genomics 9:542.

Coates ME, Beynon JL. 2010. Hyaloperonospora Arabidopsidis as a pathogen model. Annu. Rev. Phytopathol. 48:329–345.

Creevey CJ, Fitzpatrick DA, Philip GK, Kinsella RJ, O'Connell MJ, Pentony MM, Travers SA, Wilkinson M, McInerney JO. 2004. Does a tree-like phylogeny only exist at the tips in the prokaryotes? Proc. R. Soc. B Biol. Sci. 271:2551–2558.

Creevey CJ, McInerney JO. 2005. Clann: Investigating phylogenetic information through supertree analyses. Bioinformatics 21:390–392.

Diéguez-Uribeondo J, García MA, Cerenius L, Kozubíková E, Ballesteros I, Windels C, Weiland J, Kator H, Söderhäll K, Martín MP. 2009. Phylogenetic relationships among plant and animal parasites, and saprotrophs in Aphanomyces (Oomycetes). Fungal Genet. Biol. 46:365–376.

Domazet-Lošo T, Brajković J, Tautz D. 2007. A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. Trends Genet. 23:533–539.

Drost H-G, Gabel A, Grosse I, Quint M. 2015. Evidence for Active Maintenance of Phylotranscriptomic Hourglass Patterns in Animal and Plant Embryogenesis. Mol. Biol. Evol. 32:1221–1231.

Edgar RC. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.

Enright AJ, Van Dongen S, Ouzounis CA. 2002. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. 30:1575–1584.

Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, Haw R, Jassal B, Korninger F, May B, et al. 2018. The Reactome Pathway Knowledgebase. Nucleic Acids Res. 46:D649–D655.

Feng BZ, Zhu XP, Fu L, Lv RF, Storey D, Tooley P, Zhang XG. 2014. Characterization of necrosis-inducing NLP proteins in Phytophthora capsici. BMC Plant Biol. 14:126.

Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, Chang H-Y, Dosztányi Z, El-Gebali S, Fraser M, et al. 2017. InterPro in 2017—beyond protein family and domain annotations. Nucleic Acids Res. 45:D190–D199.

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, et al. 2016. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 44:D279–D285.

Fisher MC, Henk DA, Briggs CJ, Brownstein JS, Madoff LC, McCraw SL, Gurr SJ. 2012. Emerging fungal threats to animal, plant and ecosystem health. Nature 484:186–194.

Gascuel Q, Martinez Y, Boniface M-C, Vear F, Pichon M, Godiard L. 2015. The sunflower downy mildew pathogen Plasmopara halstedii. Mol. Plant Pathol. 16:109–122.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New Algorithms and Mehtods to Estimate Maximum-Likelihood Phylogenies: Asessing the Performance of PhyML 2.0. Syst. Biol. 59:307–321.

Guo Y-L. 2013. Gene family evolution in green plants with emphasis on the origination and evolution of Arabidopsis thaliana genes. Plant J. 73:941–951.

Haas BJ, Kamoun S, Zody MC, Jiang RHY, Handsaker RE, Cano LM, Grabherr M, Kodira CD, Raffaele S, Torto-Alalibo T, et al. 2009. Genome sequence and analysis of the Irish potato famine pathogen Phytophthora infestans. Nature 461:393–398.

Hamilton JP, Neeno-Eckwall EC, Adhikari BN, Perna NT, Tisserat N, Leach JE, Levesque CA, Buell CR. 2011. The Comprehensive Phytopathogen Genomics Resource: a web-based resource for data-mining plant pathogen genomes. Database 2011:bar053–bar053.

Haverkort AJ, Boonekamp PM, Hutten R, Jacobsen E, Lotz LAP, Kessel GJT, Visser RGF, van der Vossen EAG. 2008. Societal Costs of Late Blight in Potato and Prospects of Durable Resistance Through Cisgenic Modification. Potato Res. 51:47–57.

He X, Zhang J. 2006. Higher Duplicability of Less Important Genes in Yeast Genomes. Mol. Biol. Evol. 23:144–151.

Jiang RHY, de Bruijn I, Haas BJ, Belmonte R, Löbach L, Christie J, van den Ackerveken G, Bottin A, Bulone V, Díaz-Moreno SM, et al. 2013. Distinctive Expansion of Potential Virulence Genes in the Genome of the Oomycete Fish Pathogen Saprolegnia parasitica.McDowell JM, editor. PLoS Genet. 9:e1003272.

Jiang RHY, Tyler BM. 2012. Mechanisms and Evolution of Virulence in Oomycetes. Annu. Rev. Phytopathol. 50:295–318.

Jiang RHY, Tyler BM, Govers F. 2006. 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.

Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236–1240.

Kaessmann H. 2010. Origins, evolution, and phenotypic impact of new genes. Genome Res. 20:1313–1326.

Kamoun S. 2006. A catalogue of the effector secretome of plant pathogenic oomycetes. Annu. Rev. Phytopathol. 44:41–60.

Kamoun S, Furzer O, Jones JDG, Judelson HS, Ali GS, Dalio RJD, Roy SG, Schena L, Zambounis A, Panabières F, et al. 2015. The Top 10 oomycete pathogens in molecular plant pathology. Mol. Plant Pathol. 16:413–434.

Kemen E, Gardiner A, Schultz-Larsen T, Kemen AC, Balmuth AL, Robert-Seilaniantz A, Bailey K, Holub E, Studholme DJ, 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.

Klopfenstein D V., Zhang L, Pedersen BS, Ramírez F, Vesztrocy AW, Naldi A, Mungall CJ, Yunes JM, Botvinnik O, Weigel M, et al. 2018. GOATOOLS: A Python library for Gene Ontology analyses. Sci. Rep. 8:1–17.

Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305:567–580.

Letunic I, Bork P. 2007. Interactive Tree Of Life (iTOL): An online tool for phylogenetic tree display and annotation. Bioinformatics 23:127–128.

Lévesque CA, Brouwer H, Cano L, Hamilton JP, Holt C, Huitema E, Raffaele S, Robideau GP, 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 L, Huang Y, Xia X, Sun Z. 2006. Preferential Duplication in the Sparse Part of Yeast Protein Interaction Network. Mol. Biol. Evol. 23:2467–2473.

Links MG, Holub E, Jiang RH, Sharpe AG, Hegedus D, Beynon E, Sillito D, Clarke WE, Uzuhashi S, Borhan MH. 2011. De novo sequence assembly of Albugo candida reveals a small genome relative to other biotrophic oomycetes. BMC Genomics 12:503.

Liu D, Hunt M, Tsai IJ. 2018. Inferring synteny between genome assemblies: a systematic evaluation. BMC Bioinformatics 19:26.

Maguire SL, ÓhÉigeartaigh SS, Byrne KP, Schröder MS, O'Gaora P, Wolfe KH, Butler G. 2013. Comparative genome analysis and gene finding in Candida species using CGOB. Mol. Biol. Evol. 30:1281–1291.

Makkonen J, Vesterbacka A, Martin F, Jussila J, Diéguez-Uribeondo J, Kortet R, Kokko H. 2016. Mitochondrial genomes and comparative genomics of Aphanomyces astaci and Aphanomyces invadans. Sci. Rep. 6:36089.

Martens C, Van de Peer Y. 2010. The hidden duplication past of the plant pathogen Phytophthora and its consequences for infection. BMC Genomics. 11:353.

Matari NH, Blair JE. 2014. A multilocus timescale for oomycete evolution estimated under three distinct molecular clock models. BMC Evol. Biol. 14:101.

McGowan J, Fitzpatrick DA. 2017. Genomic, Network, and Phylogenetic Analysis of the Oomycete Effector Arsenal. mSphere 2:e00408-17.

Mi H, Huang X, Muruganujan A, Tang H, Mills C, Kang D, Thomas PD. 2017. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Res. 45:D183–D189.

Neme R, Tautz D. 2013. Phylogenetic patterns of emergence of new genes support a model of frequent de novo evolution. BMC Genomics. 14:117.

Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. 1999. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 27:29–34.

OhEigeartaigh SS, Armisen D, Byrne KP, Wolfe KH. 2014. SearchDOGS Bacteria, Software That Provides Automated Identification of Potentially Missed Genes in Annotated Bacterial Genomes. J. Bacteriol. 196:2030–2042.

Oome S, Van den Ackerveken G. 2014. Comparative and functional analysis of the widely occurring family of Nep1-like proteins. Mol. Plant. Microbe. Interact. 27:1–51.

Ospina-Giraldo MD, Griffith JG, Laird EW, Mingora C. 2010. The CAZyome of Phytophthora spp.: A comprehensive analysis of the gene complement coding for carbohydrate-active enzymes in species of the genus Phytophthora. BMC Genomics 11:525.

Panda A, Sen D, Ghosh A, Gupta A, C. MM, Prakash Mishra G, Singh D, Ye W, Tyler BM, Tripathy S. 2018. EumicrobeDBLite: a lightweight genomic resource and analytic platform for draft oomycete genomes. Mol. Plant Pathol. 19:227–237.

Papp B, Pál C, Hurst LD. 2003. Dosage sensitivity and the evolution of gene families in yeast. Nature. 424:194–197.

Phillippy AM, Schatz MC, Pop M. 2008. Genome assembly forensics: finding the elusive mis-assembly. Genome Biol. 9:R55.

Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 - Approximately maximum-likelihood trees for large alignments. PLoS One 5.

Quint M, Drost H-G, Gabel A, Ullrich KK, Bönn M, Grosse I. 2012. A transcriptomic hourglass in plant embryogenesis. Nature 490:98–101.

Richards TA, Dacks JB, Jenkinson JM, Thornton CR, Talbot NJ. 2006. Evolution of Filamentous Plant Pathogens: Gene Exchange across Eukaryotic Kingdoms. Curr. Biol. 16:1857–1864.

Rizzo DM, Garbelotto M, Hansen EM. 2005. Phytophthora ramorum: Integrative Research and Management of an Emerging Pathogen in California and Oregon Forests. Annu. Rev. Phytopathol. 43:309–335.

Sambles C, Schlenzig A, O'Neill P, Grant M, Studholme DJ. 2015. Draft genome sequences of Phytophthora kernoviae and Phytophthora ramorum lineage EU2 from Scotland. Genomics Data. 6:193–194.

Sestak MS, Domazet-Loso T. 2015. Phylostratigraphic profiles in zebrafish uncover chordate origins of the vertebrate brain. Mol. Biol. Evol. 32:299–312.

Sharma R, Xia X, Cano LM, Evangelisti E, Kemen E, Judelson H, Oome S, Sambles C, van den Hoogen DJ, Kitner M, et al. 2015. Genome analyses of the sunflower pathogen Plasmopara halstedii provide insights into effector evolution in downy mildews and Phytophthora. BMC Genomics 16:741.

De Smet R, Adams KL, Vandepoele K, Van Montagu MCE, Maere S, Van de Peer Y. 2013. Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. Proc. Natl. Acad. Sci. 110:2898–2903.

Sperschneider J, Williams AH, Hane JK, Singh KB, Taylor JM. 2015. Evaluation of Secretion Prediction Highlights Differing Approaches Needed for Oomycete and Fungal Effectors. Front. Plant Sci. 6:1–14.

Tautz D, Domazet-Lošo T. 2011. The evolutionary origin of orphan genes. Nat. Rev.

Genet. 12:692–702.

Tyler BM. 2006. Phytophthora Genome Sequences Uncover Evolutionary Origins and Mechanisms of Pathogenesis. Science (80-.). 313:1261–1266.

Tyler BM. 2007. Phytophthora sojae: root rot pathogen of soybean and model oomycete. Mol. Plant Pathol. 8:1–8.

Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva E V, Zdobnov EM. 2018. BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. Mol. Biol. Evol. 35:543–548.

Yachdav G, Wilzbach S, Rauscher B, Sheridan R, Sillitoe I, Procter J, Lewis SE, Rost B, Goldberg T. 2016. MSAViewer: interactive JavaScript visualization of multiple sequence alignments. Bioinformatics 32:btw474.

Yang Z, Nielsen R. 2000. Estimating Synonymous and Nonsynonymous Substitution Rates Under Realistic Evolutionary Models. Mol. Biol. Evol. 17:32–43.

Chapter 4

Comparative Genomic and Proteomic Analyses of Three Widespread *Phytophthora* Species: *Phytophthora chlamydospora*, *Phytophthora gonapodyides* and *Phytophthora pseudosyringae*

This chapter has been published in the journal Microorganisms.

McGowan, J., O'Hanlon, R., Owens, R.A., Fitzpatrick, D.A., 2020. Comparative Genomic and Proteomic Analyses of Three Widespread *Phytophthora* Species: *Phytophthora chlamydospora*, *Phytophthora gonapodyides* and *Phytophthora pseudosyringae*. Microorganisms. 8, 653.

Abstract

The *Phytophthora* genus includes some of the most devastating plant pathogens. Here we report draft genome sequences for three ubiquitous *Phytophthora* species—*Phytophthora chlamydospora*, *Phytophthora gonapodyides*, and *Phytophthora pseudosyringae*. *Phytophthora pseudosyringae* is an important forest pathogen that is abundant in Europe and North America. *Phytophthora chlamydospora* and *Ph. gonapodyides* are globally widespread species often associated with aquatic habitats. They are both regarded as opportunistic plant pathogens. The three sequenced genomes range in size from 45 Mb to 61 Mb. Similar to other oomycete species, tandem gene duplication appears to have played an important role in the expansion of effector arsenals. Comparative analysis of carbohydrate-active enzymes (CAZymes) across 44 oomycete genomes indicates that oomycete lifestyles may be linked to CAZyme repertoires. The mitochondrial genome sequence of each species was also determined, and their gene content and genome structure were compared. Using mass spectrometry, we characterised the extracellular proteome of each species and identified large numbers of proteins putatively involved in pathogenicity and osmotrophy. The mycelial proteome of each species was also characterised using mass spectrometry. In total, the expression of approximately 3000 genes per species was validated at the protein level. These genome resources will be valuable for future studies to understand the behaviour of these three widespread *Phytophthora* species.

Introduction

Phytophthora are filamentous, osmotrophic eukaryotes that morphologically resemble fungi but belong to the Oomycota class within the Stramenopila [1]. *Phytophthora* species include some of the most destructive plant pathogens, including devastating pathogens of important crops, ornamental plants, and forests. The number of identified *Phytophthora* species is rapidly increasing and currently includes more than 180 provisionally named species [2]. In recent years, the genomes of several *Phytophthora* species have been sequenced which has increased our understanding of *Phytophthora* evolution and pathology [3–9].

As osmotrophs, *Phytophthora* species secrete numerous classes of hydrolytic enzymes including carbohydrate-active enzymes (CAZymes) and proteases that digest complex extracellular substrates into simpler subunits for nutritional sources [10]. *Phytophthora* genomes also encode large arsenals of effector proteins that facilitate infection [11,12]. Effectors are divided into two broad classes based on where they localise. Apoplastic effectors are secreted by the pathogen and act in the host's extracellular environment [13]. Apoplastic effector proteins encompass a broad range of functions, including hydrolytic enzymes that degrade plant cell walls, facilitating hyphal penetration. Other apoplastic effector families include elicitins, necrosis-inducing proteins (NLPs), and Pcf phytotoxins that play roles in inducing host cell death [14]. *Phytophthora* also secrete cytoplasmic effectors that are translocated into host cells, namely the RxLR and Crinkler (CRN) families of effectors [13].

To date, most large-scale analyses of oomycete secretomes or effector arsenals have used in silico bioinformatics-based approaches applied to genome sequences to predict protein sequences bearing signal peptides. However, mass spectrometry is a powerful technique that can be used to characterise the extracellular proteomes of filamentous plant pathogens. It can also be used to validate the secretion of proteins that

have been computationally predicted to be secreted as well as identifying extracellular proteins that lack conventional signal peptides, which would otherwise be overlooked. To date, proteomic studies of oomycete extracellular proteins are limited. Liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis of the potato late blight pathogen *Ph. infestans* led to the identification of 283 extracellular proteins, many of which are members of known effector families such as RxLRs, CRNs, elicitins and NLPs [15]. Similarly, LC-MS/MS analysis of *Ph. plurivora*, an important forest pathogen, led to the detection of 272 secreted proteins, including important effectors and cell-wall degrading enzymes [16]. Proteomic studies of oomycete hyphae have also revealed important insights into metabolism, pathogenicity, and species-specific proteins [17–19].

Here, we report draft genome sequences for three globally widespread *Phytophthora* species—*Phytophthora chlamydospora*, *Phytophthora gonapodyides* and *Phytophthora pseudosyringae*. We selected these species because they are (i) widespread *Phytophthora* species in western Europe, (ii) have broadly different lifestyles, and (iii) are not regulated organisms under plant health legislation in Europe. Furthermore, all three species belong to clades that are generally underrepresented in terms of the amount of genomic data available for those clades. *Phytophthora pseudosyringae* belongs to clade 3 and is an important forest pathogen with a broad host range that is widespread across Europe and America [20–22]. *Phytophthora pseudosyringae* has been reported to cause collar, root or stem rot in a wide range of host species including European beech (*Fagus sylvatica*), southern beech (*Nothofagus* spp.), oak (*Quercus* spp.), alder (*Alnus* spp.), Japanese larch (*Larix kaempferi*), sweet chestnut (*Castanea sativa*), horse chestnut (*Aesculus hippocastanum*), and shrubs (*Vaccinium myrtillus*) [23–27]. In a recent metabarcoding study, *Ph. pseudosyringae* was the most abundant *Phytophthora* species detected in Britain [28].

Phytophthora chlamydospora and *Ph. gonapodyides* both belong to clade 6, which contains species that are primarily associated with streams and riparian ecosystems, and have evolved to tolerate higher temperatures, facilitating survival throughout seasonal changes [29,30]. *Phytophthora gonapodyides* is reported to be the most widely distributed *Phytophthora* species in natural habitats worldwide, followed by *Ph. chlamydospora* [23,31]. *Phytophthora gonapodyides* was the third most abundant *Phytophthora* species detected via metabarcoding in Britain [28]. *Phytophthora chlamydospora* was also frequently detected. *Phytophthora chlamydospora*, previously informally known as *Phytophthora* taxon Pgchlamydo, was formally designated as a species in 2015 [31]. Similar to other clade 6 *Phytophthora* species, both *Ph. chlamydospora* and *Ph. gonapodyides* are thought to exhibit primarily saprophytic lifestyles, maintaining their populations in natural ecosystems by colonising plant litter or through pathogenesis of fine-roots [30]. However, both species are known to act as opportunistic pathogens, with reports of *Ph. chlamydospora* causing disease on *Rhododendron*, *Viburnum tinus,* sour cherry (*Prunus cerasus*), almond (*Prunus dulcis*), walnut (*Juglans regia*), and evergreen nursery stock [22,32–36]. Similarly, *Ph. gonapodyides* has been detected causing disease on European beech (*F. sylvatica*), *Rhododendron*, oak (*Quercus* spp.), apple (*Malus* spp.), walnut (*J. regia*) and tanoak (*Notholithocarpus* spp.) [23,37–40].

Using a combination of bioinformatics, comparative genomics and mass spectrometry-based approaches, we provide a comprehensive characterisation of the nuclear genomes, mitochondrial genomes, putative effector arsenals, extracellular proteomes, and mycelial proteomes of *Ph. chlamydospora*, *Ph. gonapodyides,* and *Ph. pseudosyringae.* The genomic and proteomic data reported herein represent valuable new resources to study the pathogenicity mechanisms of these widespread phytopathogens and further elucidate *Phytophthora* genome evolution.

Materials and Methods

Isolation and Culturing

The details of the isolates used in this study are provided in **Table 1**. Methods for the isolation of the isolates and tentative morphological identification have been reported previously [41,42]. Briefly, this involved plating the samples onto P_5ARP[H] agar (Cornmeal agar with antibiotics [43]) and incubating at room temperature for two weeks. These agar plates were checked at least every three days for *Phytophthora* like structures (e.g., sporangia, *coenocytic* hyphae, oospores). Mycelia from the edge of *Phytophthora* like cultures was then transferred to Carrot Piece Agar (CPA) [44] plates and incubated at room temperature, studied morphologically and compared to *Phytophthora* taxonomy reference text [45,46] (www.phytophthoradb.org). Identification of the isolates based on BLAST [47] searches of the ITS region was carried out by O'Hanlon et al. (2016). Isolates were stored on CPA agar slopes under sterile mineral oil at room temperature between 2016 and 2018. From 2018 onwards, isolates were routinely cultured on 10% clarified V8 juice (cV8) agar in the dark at 20 °C for *Ph. pseudosyringae* and 25 °C for *Ph. chlamydospora* and *Ph. gonapodyides*. Species identities were confirmed based on a combination of β-tub, COX1, COX2, ITS, and Rps10 markers retrieved from the whole genome sequences and compared to reference sequences from PhytophthoraDB (http://www.phytophthoradb.org) and NCBI GenBank using BLASTn [47].

Table 1. Details of the isolates used in this study.

Species	Isolate Number	Source	Date Collected	Location (Latitude, Longitude)	BioProject Accession
Phytophthora chlamydospora	P17-99	*Rhododendron ponticum* leaf baiting of a stream	08/08/2017	Tollymore forest, Co. Down, Northern Ireland, UK (54°13'16.9"N, 5°55'49.3"W)	PRJNA599565
Phytophthora gonapodyides	P17-128	*Rhododendron ponticum* leaf baiting of a stream	28/08/2017	Rostrevor forest, Co. Louth, Ireland (54°07'43.4"N, 6°09'24.2"W)	PRJNA599567
Phytophthora pseudosyringae	PR13-731	Bleeding bark canker of *Fagus sylvatica*	31/07/2015	Mullaghreelan forest, Co. Kildare, Ireland (52°56'00.4"N, 6°52'41.0"W)	PRJNA599564

DNA Extraction and Sequencing

To prepare mycelium for DNA extraction, *Phytophthora* cultures were grown in 10% cV8 liquid medium for 5 days. Mycelia were harvested using Miracloth, washed with sterile distilled water, flash-frozen in liquid nitrogen, lyophilized and stored at −80 °C until used for DNA extraction. Lyophilized mycelium was ground to a fine powder with a mortar and pestle under liquid nitrogen. DNA was extracted by transferring 20–40 mg of ground mycelia to a tube containing 800 µL of extraction buffer (0.2 M Tris-HCl, 0.25 M NaCl, 25 mM EDTA and 0.5% SDS) and 2 µL Proteinase K (20 mg/mL; Qiagen,

Redwood City, CA, USA). Samples were incubated at 55 °C for 30 min. Samples were then treated with 3 µl RNase A (10 mg/mL; Thermo Fisher Scientific, Waltham, MA, USA) and incubated at 37 °C for 30 min. 800 µL of 24:1 chloroform:isoamyl alcohol was added to samples, mixed by inversion and centrifuged for 10 min at 13,000× *g*. The upper phase was transferred to a new tube and the chloroform step was repeated. DNA was precipitated by the addition of ½ volume of 5 M ammonium acetate and 2 volumes of 100% ethanol, followed by overnight incubation at −20 °C. Precipitated DNA was pelleted by centrifugation for 15 min at 13,000× *g*. The DNA pellet was washed twice, first with 70% ethanol and then with 100% ethanol. DNA was air-dried and resuspended in 70 µL TE buffer (10 mM Tris-HCl, 0.5 mM EDTA). DNA purity was assessed using a Nanodrop spectrophotometer (Thermo Fisher Scientific) based on the 260/280 and 260/230 absorbance ratios. DNA concentration was determined using a Qubit fluorometer with the dsDNA BR kit (Invitrogen, Carlsbad, CA). DNA quality was assessed via agarose gel electrophoresis on a 1% agarose gel. DNA library construction and paired-end sequencing were carried out by BGI Tech Solutions Co., Ltd. (Hong Kong, China) on the Illumina HiSeq X Ten platform. Sequenced reads were deposited on the NCBI Sequence Read Archive (accessions: SRR10849951 for *Ph. chlamydospora*, SRR10849950 for *Ph. gonapodyides* and SRR10849937 for *Ph. pseudosyringae*).

Genome Assembly

Genome sizes and heterozygosity levels were estimated by generating *k*-mer count histograms of sequence reads with Jellyfish [48], which were used as input for GenomeScope [49]. De novo genome assembly was performed using SPAdes (v3.13.1) [50]. Assemblies were further scaffolded using SSPACE (v3.0) [51] and gaps were closed with GapFiller (v1.10) [52]. Scaffolds with low coverage (less than 30×) were manually

assessed for contamination by BLAST searches against the NCBI database and removed if they had a top hit outside the Oomycota. Assembly metrics were calculated using Quast (v5.0.2) [53]. BUSCO (v3) [54] was used to assess the gene space completeness of assemblies with the alveolata–stramenopiles dataset of BUSCOs. De novo repeat family identification and repeat masking were performed using RepeatModeler (v2.0) [55] and RepeatMasker (v4.1.0). Mitochondrial genomes were de novo assembled separately using NOVOPlasty (v3.7) [56] and visualised using OGDRAW (v1.3.1) [57].

Gene Annotation

Gene models were predicted using BRAKER2 [58] with the ProtHint pipeline [59]. In brief, initial gene sets were predicted using GeneMark-ES (v4.46) [60]. Homologs of the initial gene predictions were identified using Diamond [61] searches against a database of 14 Peronosporales proteomes containing 267,298 proteins (**Supplementary Table S1**). Intron hints were generated by performing spliced alignments using Spaln2 [62] with the ProtHint pipeline [59]. GeneMark-EP [59] was trained with the intron hints and used to generate another gene set. The GeneMark-EP predictions, along with the intron hints, were then used to train Augustus [63] to generate the final gene sets. Completeness of gene sets were assessed using BUSCO v3.

Genes were functionally annotated using InterProScan (v5.39-77.0) [64] and eggNOG-mapper (v2) [65]. Secreted proteins and transmembrane proteins were predicted using SignalP v3 [66] and TMHMM (v2.0) [67], respectively. SignalP v3 was implemented instead of earlier or later versions of the software as previous studies have found v3 to be more sensitive in predicting oomycete signal peptides [68]. For a protein to be considered secreted, it had to have a positive prediction from SignalP, an HMM S probability value ≥0.9, an NN Y_{max} score of ≥0.5, an NN D score of ≥ 0.5, and no

transmembrane domains downstream of the predicted signal peptide cleavage site. These criteria permitted comparisons to previous studies [12,68,69]. Proteins predicted to be secreted were submitted to ApoplastP [70] to predict if they are localised to the plant apoplast. CAZymes were annotated using dbCAN2 [71]. Homologs of experimentally verified effector proteins were identified by performing BLASTp searches [47] against the pathogen-host interaction database (PHI-Base Release 4.8) [72] with an *E* value cut-off of $1e^{-20}$.

Gene families were identified by performing all-versus-all BLASTp [47] searches with an *E* value cut-off of $1e^{-10}$, followed by Markov clustering using MCL [73] with an inflation value of 1.5. Tandemly duplicated genes were identified using BLASTp [47]. Tandem clusters were defined as two or more adjacent genes that hit each other in a BLASTp search with an *E* value cut-off of $1e^{-10}$ and highest-scoring pair (HSP) length greater than half the length of the shortest sequence. Enrichment analyses were preformed using Fisher's exact test. Gene Ontology enrichment analyses were performed using Fisher's exact test with Benjamini–Hochberg correction for multiple testing using GOATOOLS [74]. Corrected *p*-values < 0.05 were considered significant.

Identification of Cytoplasmic Effectors

RxLRs were classified using four methods as in McGowan and Fitzpatrick (2017) [12]. **(i) The Win method -** proteins must contain a signal peptide with a predicted cleavage site within the first 30 amino acids and an RxLR motif within residues 30–60 [75]. **(ii) HMM method** - hidden Markov model (HMM) searches were performed with HMMER (v3.2.1) [76] against all proteins predicted to be secreted using the "cropped.hmm" HMM profile constructed by Whisson et al. (2007) [77]. Hits with a bit score > 0 were retained. **(iii) Regex method** - proteins must contain a signal peptide between residues 10 to 40

and an RxLR motif within the following 100 residues followed by the EER motif within 40 residues downstream of the RxLR motif, allowing for replacements of E to D and R to K [77]. This search was performed using the regular expression "^.{10,40}.{1,96}R.LR.{1,40}[ED][ED][KR]". **(iv) Homology method** - proteins with a positive SignalP prediction were searched using BLASTp against a set of 1207 putative *Phytophthora* RxLRs from *Ph. infestans*, *Ph. ramorum* and *Ph. sojae* [3]. An *E* value cut-off of $1e^{-20}$ was applied. Secreted proteins that met at least one of these four criteria were considered to be putative RxLRs. Additionally, an HMM search was performed on all putative RxLRs to determine if they have WY-domains, using the HMM described by Boutemy et al. (2011) [78].

CRNs were identified using the regular expression "^.{30,70}L[FY]LA[RK]". Proteins with a positive hit from the regular expression search were aligned and an HMM model was constructed. The CRN HMM was then searched against the predicted proteomes using HMMER [76] and all proteins with a bit score >0 were considered the final set of putative CRNs.

Phylogenomics

A dataset of 33 Peronosporales genomes (**Supplementary Table S1**) was used for phylogenomic analysis. We also included *Pythium ultimum* as an outgroup. BUSCO analysis revealed 208 BUSCO families that are present and single copy in at least 90% of the species (i.e., at least 31 of the 34 species). Each BUSCO family was individually aligned with MUSCLE (v3.8.31) [79] and trimmed using trimAl (v1.4) [80] with the parameter "-automated1" to remove poorly aligned regions. Trimmed alignments were concatenated together resulting in a final supermatrix alignment of 106,315 amino acids. Maximum-likelihood (ML) phylogenetic reconstruction was performed using IQ-TREE

(v1.6.12) [81] with the JTT+F+R5 model, which was the best fit model according to ModelFinder [82], and 100 bootstrap replicates were undertaken to infer branch support values. Bayesian analysis was also performed using PhyloBayes MPI (v1.8) [83] with the CAT model. Two independent chains were run for 10,000 cycles and convergence was assessed using bpcomp and tracecomp. A consensus Bayesian phylogeny was generated with a burn-in of 10%. The phylogeny was visualised and annotated using the Interactive Tree of Life (iTOL) [84].

Phylostratigraphy

Phylostratigraphic maps were determined for each species following previously published methods [85,86]. The database constructed by Drost et al. (2015) was retrieved which contains amino acid sequences from 4557 species, including 1787 eukaryotes (883 animals, 364 plants, 344 fungi and 193 other eukaryotes) and 2770 prokaryotes (2511 bacteria and 259 archaea) [85]. Protein sequences from the three *Phytophthora* species sequenced in this study and any publicly available oomycete proteomes were added to the database. The final database comprised 18,084,866 proteins, including 578,493 proteins from 38 oomycete genomes. Each *Phytophthora* protein was then searched against this database using BLASTp [47]. Proteins were assigned to the oldest phylostratum that contained at least one BLAST hit with an *E* value cut-off < $1e^{-5}$. Genes that did not have a BLAST hit to any other species were considered species-specific (orphans).

Culturing Conditions and Extraction of Phytophthora Extracellular Proteins

Petri dishes containing 15 mL of liquid medium (either 10% V8 broth or 10% cV8 broth) were inoculated with a 10 mm agar plug of *Phytophthora* mycelium cut from the edge of a growing *Phytophthora* colony. Cultures were incubated in the dark, non-shaking for 10 days at their optimum temperatures (20 °C for *Ph. pseudosyringae* and 25 °C for *Ph. chlamydospora* and *Ph. gonapodyides*). Spent growth medium was harvested using a syringe without disturbing the mycelium. Supernatant from four petri dishes were pooled to make up one replicate. Collected supernatant was passed through a 0.2 µm syringe filter, frozen overnight at −20 °C and lyophilized. Lyophilized supernatant was resuspended in minimal volumes of PBS, desalted and concentrated using Amicon Ultra centrifugal filters (Millipore, Billerica, MA, USA) with a 3 kDa cut-off. Samples were clarified by centrifugation at 12,000× *g* for 5 min and brought to 15% (*v/v*) trichloroacetic acid (TCA) using 100% TCA. Precipitated proteins were washed twice with ice-cold acetone. Dried protein pellets were resuspended in 6 M Urea, 2 M Thiourea and 0.1 M Tris-HCl pH 8.0. Protein concentration was determined using a Qubit fluorometer (Invitrogen).

Culturing Conditions and Extraction of Phytophthora Mycelial Proteins

Phytophthora mycelium was cultured under three growth conditions. **(i) Normal** - cultures were grown for ten days at their optimum temperatures. **(ii) Heat** - cultures were grown for 7 days at their optimum temperatures, followed by incubation at 30 °C for three days. **(iii) Oxidative stress** - cultures were grown for ten days at their optimum temperatures, then exposed to 1 mM H_2O_2 for three hours. All cultures were grown in 50 mL of 10% cV8, non-shaking. Mycelia were harvested using Miracloth, washed with

sterile distilled water, flash-frozen in liquid nitrogen and stored at −80 °C until used for protein extraction.

To extract proteins, mycelium was ground to a fine powder with a mortar and pestle under liquid nitrogen. 200–300 mg of ground mycelium was resuspended in 400 µL of lysis buffer followed by sonication (Bandelin Sonopuls HD2200 sonicator, Cycle 6, Berlin, Germany, 3 × 10 s, Power 20%). Protein concentration was determined using a Qubit fluorometer (Invitrogen). Protein lysates (0.25 mg/mL) were incubated at 95 °C for 5 min.

Protein Digestion and LC-MS/MS Identification of Phytophthora Proteins

Three independent biological replicates were analysed for each condition. Proteins were reduced and alkylated prior to overnight trypsin digestion as described previously [87,88]. Digestion was terminated by the addition of 1 µL of 100% trifluoroacetic acid (TFA). Sample clean-up was performed using C18 ZipTips® (Millipore), following the manufacturer's instructions. Shotgun proteomics was performed using an Ultimate 3000 RSLC from Dionex, coupled to a Thermo Scientific Q-Exactive mass spectrometer. Peptide mixtures were separated on a 50 cm EASY-Spray PepMap C18 column with 75 µm diameter (2 µm particle size) using a 10–40 % B gradient (A: 0.1% (*v/v*) formic acid, 3% (*v/v*) acetonitrile; B: 0.1% (*v/v*) formic acid, 80% (*v/v*) acetonitrile). Data were acquired for 105 min, at 70,000 resolution for MS and a Top 15 method for MS2 collection.

Protein identification from the data was performed using the Andromeda search engine [89] in MaxQuant [90] with the predicted proteomes for each species as a search database. To account for possible protein contamination from V8 juice medium, we appended the tomato proteome to the default MaxQuant contaminants database. Search

parameters were as described in Delgado et al. (2019) [91]. Identified protein groups were filtered using Perseus [92], to remove protein groups that were identified only by site, or had hits to either the contaminants database or the reverse database. Proteins were considered present in a condition if they were identified by 2 or more peptides and detected in at least 2 out of 3 replicates. Proteins were considered unique to a condition if they were not detected in any replicate of any other condition.

Data Deposition

This project has been deposited at DDBJ/EMBL/GenBank as individual WGS BioProjects, with the following BioProject accession numbers: PRJNA599565 (*Ph. chlamydospora*), PRJNA599567 (*Ph. gonapodyides*) and PRJNA599564 (*Ph. pseudosyringae*).

Results and Discussion

Genome Sequencing and Assembly

Paired-end Illumina sequencing generated approximately 5.2 Gb of sequencing data for each of the three species. Genome sizes and heterozygosity levels were estimated based on K-mer analysis of sequence reads using Jellyfish [48] and GenomeScope [49], which estimated genome sizes of 51.1 Mb, 65.2 Mb and 51.0 Mb (**Table 2**) for *Ph. chlamydospora*, *Ph. gonapodyides*, and *Ph. pseudosyringae*, respectively. The heterozygosity of *Ph. gonapodyides* was estimated to be 1.88% (**Table 2**), which was much higher than *Ph. chlamydospora* (0.68%) and *Ph. pseudosyringae* (0.15%) (Table 2), and high compared to other oomycetes which typically have heterozygosity levels less than 1% [93]. De novo genome assembly using SPAdes [50] generated draft genome assemblies with assembly sizes of 45.3 Mb for *Ph. chlamydospora*, 61.1 Mb for *Ph. gonapodyides* and 47.9 Mb for *Ph. pseudosyringae* (**Table 2**), which compares favourably to the genome sizes estimated by GenomeScope. The *Ph. gonapodyides* assembly was much more fragmented (16,449 scaffolds) than *Ph. chlamydospora* (4077 scaffolds) and *Ph. pseudosyringae* (3627 scaffolds) (**Table 2**). We expect that this is due to higher levels of heterozygosity found in the *Ph. gonapodyides* genome assembly and due to expansions of repetitive elements (**Supplementary Table S2**). BUSCO analysis [54] was performed using the Alveolata-Stramenopiles dataset which contains 234 target BUSCO proteins that are expected to be present and single-copy. BUSCO results suggests that the assemblies are of high gene space completeness with BUSCO completeness values of 97.8% for *Ph. chlamydospora*, 87.2% for *Ph. gonapodyides*, and 94.1% for *Ph. pseudosyringae*) (**Table 2 and Figure 1**). Furthermore, the low number of duplicated BUSCOs suggest that haplotypes were correctly collapsed (**Figure 1**). De novo repeat

annotation using RepeatModeler2 [55] and RepeatMasker led to the identification of 4.1 Mb (9.0%) of repetitive elements in *Ph. chlamydospora*, 9.8 Mb (16.1%) in *Ph. gonapodyides* and 6.4 Mb (13.3%) in *Ph. pseudosyringae* (**Table 2**). The majority of identified repeats were unclassified or were classified as long terminal repeat (LTR) retroelements (**Supplementary Table S2**). Overall, the proportions of repetitive elements identified are similar to that of *Ph. parasitica* (8%), *Ph. plurivora* (15%), *Ph. cactorum* (18%) and *Ph. capsici* (21%) [5,6] but less than that of *Ph. sojae* (31%), *Ph. ramorum* (54%) and *Ph. infestans* (74%) [3,94,95].

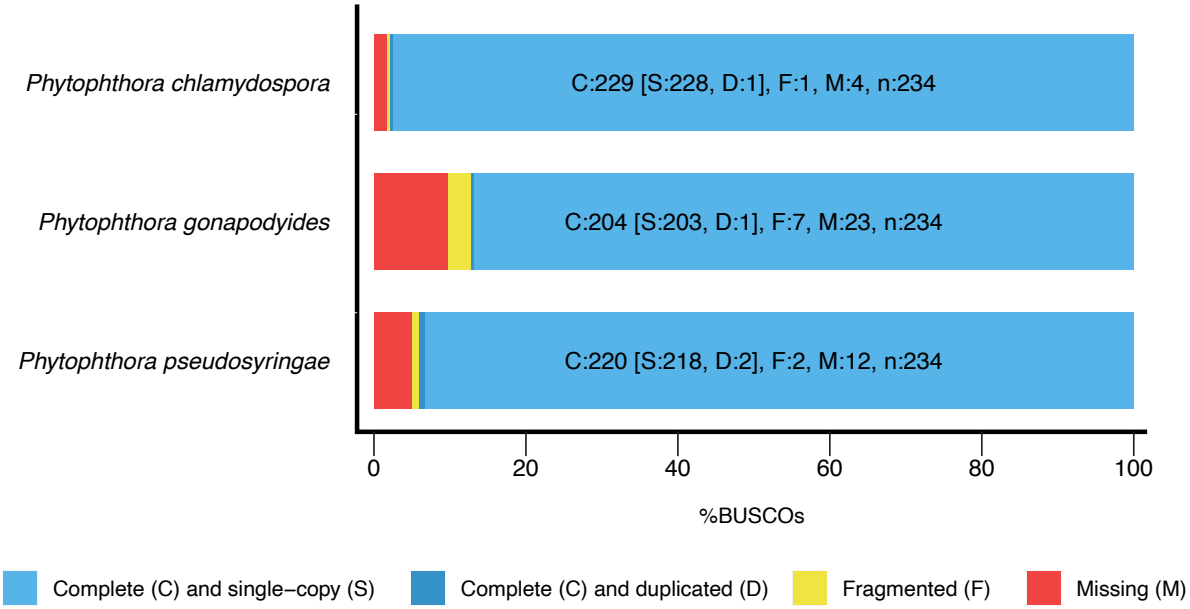


Figure 1. BUSCO analysis of *Ph. chlamydospora*, *Ph. gonapodyides*, and *Ph. pseudosyringae* using the Alveolata-Stramenopiles dataset. BUSCO completeness in all three species is high (97.8%, 87.2% and 94.1%) indicating that the genome assemblies are of high gene space completeness. The lower level of BUSCO completeness in *Ph. gonapodyides* is most likely the result of a fragmented assembly due to high levels of heterozygosity.

Table 2. Genome assembly statistics.

Genome Assembly	***Phytophthora chlamydospora***	***Phytophthora gonapodyides***	***Phytophthora pseudosyringae***
Estimated Genome Size (bp)	51,100,498	65,211,327	51,026,880
Assembly Size (bp)	45,264,984	61,088,431	47,882,184
Number of Scaffolds	4077	16,449	3627
N50 (bp)	26,559	5455	26,492
L50	466	2927	526
GC Content	55.7%	55.7%	54.8%
Sequencing Coverage	98x	76x	102x
Repeat Masked	9.0%	16.1%	13.3%
Estimated Heterozygosity	0.68%	1.88%	0.15%
Gene Models	17,872	23,348	17,439
CDS density	56.3%	43.3%	49.2%
Genome BUSCO Completeness	97.8%	87.2%	94.1%
Gene Set BUSCO Completeness	97.5%	88.1%	94.4%
Proteins with Pfam domains	10,759 (60.2%)	12,181 (52.2%)	10,130 (58.1%)

Gene prediction led to the annotation of 17,872 gene models for *Ph. chlamydospora*, 23,348 for *Ph. gonapodyides* and 17,439 for *Ph. pseudosyringae* (**Table 2**), with BUSCO completeness scores of 97.5%, 87.2% and 94.4% (**Table 2**), suggesting high quality gene model annotation. Between 52.5% and 60.2% of proteins were annotated with one or more Pfam domains (**Table 2 and Supplementary Table S3**). The percentage of the genome assemblies covered by coding sequences were 56.3%, 43.3% and 49.2 % for *Ph. chlamydospora*, *Ph. gonapodyides* and *Ph. pseudosyringae*, respectively, which are similar to *Ph. cinnamomi* (43.7%), *Ph. parasitica* (43.0%) and *Ph. sojae* (45.6%) but higher than what is observed in *Ph. ramorum* (33.7%), *Ph. capsici* (38.8%) and *Ph. infestans* (12.36%) [93].

Phylogenomics Analysis

A phylogenomic analysis was carried out to determine the phylogenetic relationships of the three *Phytophthora* species using the genome sequences of 33 Peronosporales species and *Py. ultimum* as an outgroup (**Supplementary Table S1**). A supermatrix alignment was constructed from 208 highly conserved BUSCO families. Maximum Likelihood (ML) and Bayesian phylogenetic reconstruction was undertaken on the supermatrix. Both ML and Bayesian methods resulted in phylogenies with identical topologies and most nodes had maximum Bootstrap Support (BP) or maximum Bayesian Posterior Probabilities (BPP) (**Figure 2**). All species were placed into their expected clades, and overall the placement of each species is in broad agreement with previous studies [93,96,97]. We also recovered the polyphyly of the downy mildews (**Figure 2**). Interestingly, the branch lengths of the Downy Mildew species are longer relative to the other Peronosporales species presented indicating higher levels of genetic divergence (**Figure 2**). *Ph. pseudosyringae* was placed as sister to *Ph. pluvialis*, which is also a clade

3 species, with maximum support from both ML and Bayesian methods (**Figure 2**). *Phytophthora gonapodyides* was placed as sister to *Ph. pinifolia*, to the exclusion of *Ph. chlamydospora,* with 87% BP from the ML phylogeny and maximum support from the Bayesian phylogeny (**Figure 2**). We note that this disagrees with previous phylogenies based on ITS sequences [31] and four concatenated mitochondrial loci [98], both of which group *Ph. chlamydospora* and *Ph. pinifolia* as being more closely related. Some markers, including the ITS sequence, are known to be identical or nearly identical between members of *Phytophthora* clade 6b, which *Ph. chlamydospora*, *Ph. gonapodyides*, and *Ph. pinifolia* belong to [99]. Due to the highly conserved nature of these markers, they may not reflect the true phylogenetic relationships between species. Furthermore, phylogenomic approaches are generally considered to be more informative than single gene phylogenies or phylogenies derived from small numbers of genes, as they utilise substantially greater amounts of phylogenetically informative genomic data [100].

Our phylogeny groups *Ph. cinnamomi* with *Ph. sojae* and *Ph. pisi*, to the exclusion of *Ph. fragariae* and *Ph. rubi*, with 92% BP from the ML phylogeny and maximum support in the Bayesian phylogeny (**Figure 2**). This is in disagreement with a phylogeny based on seven nuclear genetic markers which groups *Ph. sojae*, *Ph. pisi*, *Ph. fragariae*, and *Ph. rubi* together to the exclusion of *Ph. cinnamomi* [2]. However, our phylogeny is in agreement with two separate studies based on seven nuclear loci which group *Ph. cinnamomi* more closely related to *Ph. sojae* [97,98]. We anticipate that differences in topology are due to the inclusion or exclusion of different species in datasets.

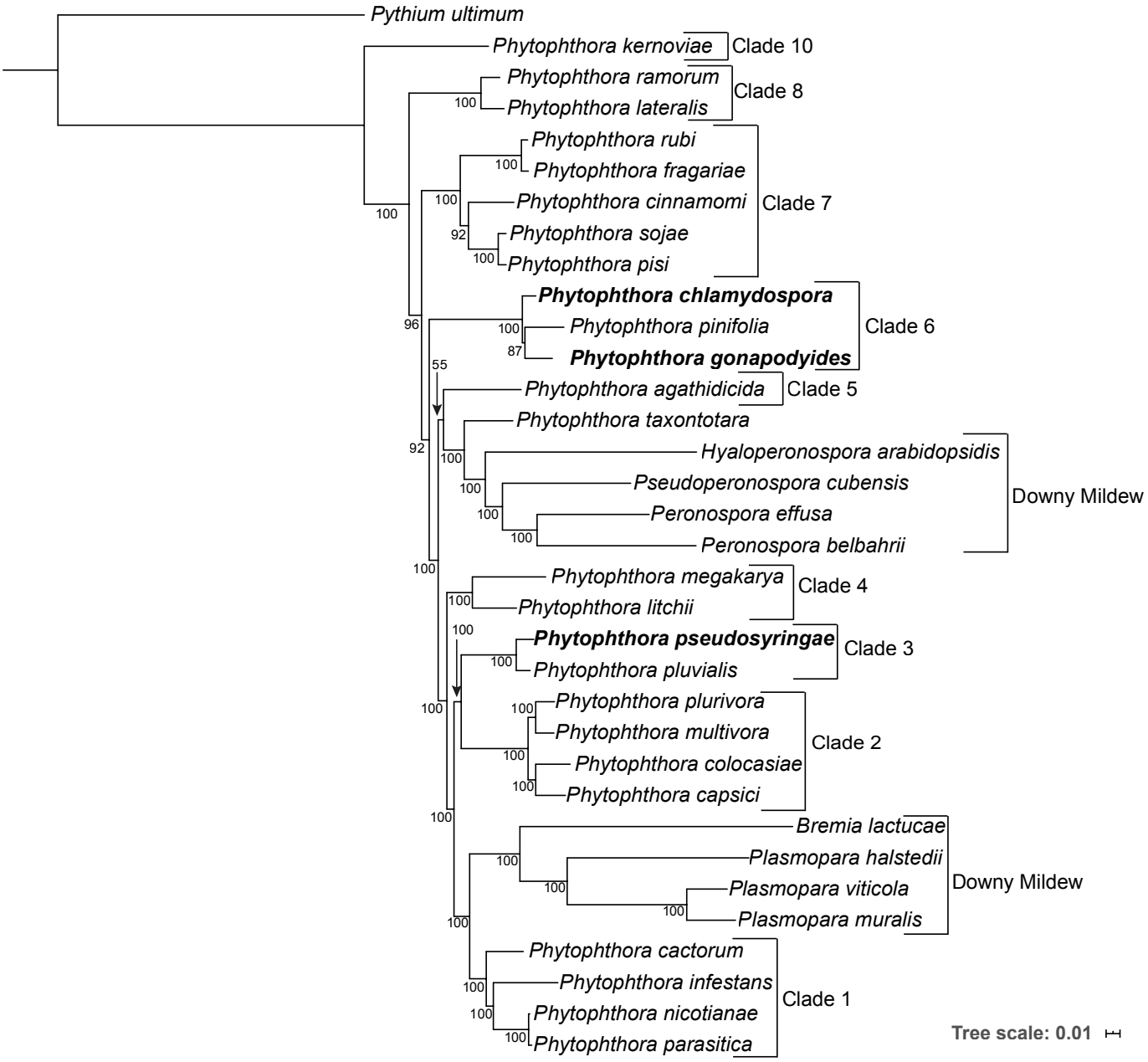


Figure 2. Supermatrix phylogeny of 33 Peronosporales species (208 BUSCO families, 106,315 amino acids). *Py. ultimum* is included as an outgroup. Phylogenomic analyses were performed using both maximum likelihood (IQ-TREE with JTT+F+R5 model) and Bayesian inference (PhyloBayes MPI with the CAT model). Both methods inferred phylogenies with identical topologies. Branch lengths are shown. Maximum likelihood bootstrap supports are indicated at all nodes. Bayesian posterior probabilities were 100 for all nodes and are not shown.

Phytophthora Mitochondrial Genomes

Mitochondrial genomes were assembled and circularised using NOVOPlasty, resulting in mitochondrial genome assemblies sizes of 38.33 Kb for *Ph. chlamydospora*, 43.97 Kb for *Ph. gonapodyides* and 39.14 Kb for *Ph. pseudosyringae* in length (**Figure 3**), which are similar in size to previously sequenced *Phytophthora* mitochondrial genomes [101]. The overall mitochondrial GC content is also highly similar to other *Phytophthora* species, with 22.5% for *Ph. chlamydospora*, 23.7% for *Ph. gonapodyides* and 22.0% for *Ph. pseudosyringae*. We did not detect any inverted repeats. The gene content of each mitochondrion is similar to that of other *Phytophthora* mitochondrial genomes, including 35 known protein-coding genes (18 respiratory chain proteins, 16 ribosomal proteins, and the import protein secY), two ribosomal RNA genes (rns and rnl) and 25 (*Ph. chlamydospora* and *Ph. pseudosyringae*) or 26 (*Ph. gonapodyides*) transfer RNA genes that specify 19 amino acids (**Figure 3 and Supplementary Table S4**). As with other oomycetes the tRNA gene for threonine was not located in the mitochondrial genomes of the three species presented here. Unlike animals and fungi, oomycete mitochondria use the standard genetic code [102]. The majority of mitochondrial genes had the TAA stop codon except for nad11 which has the TGA stop codon in all three species. Other exceptions include ORF24 (TAG) in *Ph. chlamydospora,* ORFS 13, 23, 40 (TGA) and ORF4 (TAG) in *Ph. gonapodyides* as well as ORFS8&25 in *Ph. pseudosyringae*. Coding regions account for 87.55%, 81.95% and 87.55% of the genomes of *Ph. chlamydospora, Ph. gonapodyides* and *Ph. pseudosyringae* respectively.

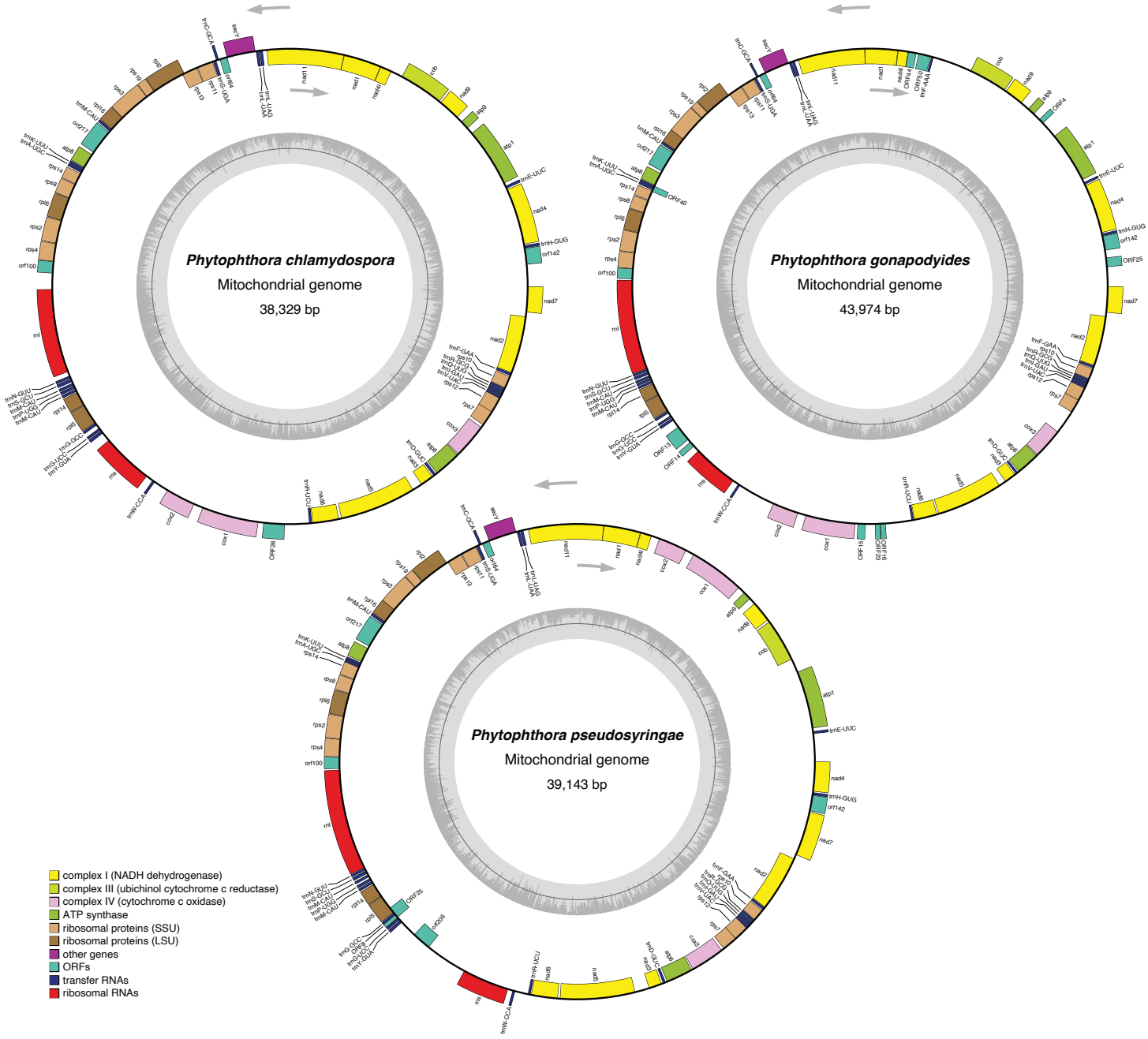


Figure 3. Circular maps of the mitochondrial genomes of *Ph. chlamydospora*, *Ph. gonapodyides* and *Ph. pseudosyringae*. The inner ring shows % GC content. Arrows indicate relative transcriptional orientation. The outer ring shows the predicted genes which are encoded on both strands. All three species are missing the tRNA gene for threonine.

Nucleotide alignment of the mitochondrial assemblies revealed that the *Ph. chlamydospora* and *Ph. gonapodyides* mitochondria are collinear (**Figure 3 and Supplementary Figure S1**). Two inversions are present in the *Ph. pseudosyringae* mitochondrial genome relative to *Ph. chlamydospora* and *Ph. gonapodyides* (**Figure 3 and Supplementary Figure S1**). We identified a number of open reading frames (ORFs) that are conserved between all three mitochondrial genomes and other *Phytophthora* mitochondrial genomes, including orf64, orf100, orf142 and orf217 (**Figure 3 and Supplementary Table S4**). The functions of these ORFs are unknown. *Phytophthora pseudosyringae* also shares an additional ORF that is homologous with *Ph. sojae* orf206 [103] (**Figure 3**). A large number of unique unannotated ORFs were identified in *Ph. gonapodyides* (ORF4, ORF13, ORF14, ORF15, ORF16, ORF23, ORF25, ORF40, ORF44, and ORF50), compared to two in *Ph. pseudosyringae* (ORF8 and ORF25) and only one in *Ph. chlamydospora* (ORF26) (**Figure 3**).

Bioinformatic Characterisation of *Phytophthora* Effector Arsenals

Bioinformatic annotation of *Phytophthora* secretomes was performed using SignalP. This analysis predicted 1140, 1291 and 1131 secreted proteins for *Ph. chlamydospora*, *Ph. gonapodyides,* and *Ph. pseudosyringae* respectively (**Table 3**), accounting for 6.38%, 5.53% and 6.49% of their total genome complement, similar to the number of secreted proteins reported for other *Phytophthora* genomes [12]. ApoplastP predicted that 47.1% to 48.8% of putative secreted proteins localise to the plant apoplast (**Table 3**). Approximately 20% of all putatively secreted proteins are homologous to experimentally verified effectors in PHI-Base (**Table 3**). InterProScan [64] was used to annotate putative effector proteins based on conserved Pfam domains known to be implicated in plant pathogenicity. Some of these effectors are discussed below.

Table 3. Counts of putative effector proteins.

	Phytophthora chlamydospora	***Phytophthora gonapodyides***	***Phytophthora pseudosyringae***
Secreted Proteins	1140 (6.38%)	1291 (5.53%)	1131 (6.49%)
ApoplastP Hits	554 (48.5%)	630 (48.8%)	533 (47.1%)
PHI-Base homologs	249 (21.8%)	234 (18.1%)	243 (21.5%)
Apoplastic effectors			
Berberine-like proteins	1 (1)	2 (1)	3 (1)
Cysteine-rich secretory proteins(CAP)	34 (22)	34 (25)	31 (22)
Elicitins	57 (45)	59 (47)	45 (34)
Necrosis-inducing proteins	25 (19)	33 (22)	22 (19)
PAN/Apple domain	25 (21)	32 (20)	20 (15)
PcF phytotoxins	1 (0)	1 (1)	1 (1)
Transglutaminase elicitors	14 (11)	16 (11)	17 (11)
Proteases and protease inhibitors			
Aspartyl proteases	59 (4)	53 (3)	26 (2)
Papain family cysteine proteases	19 (8)	22 (6)	20 (8)
Serine proteases	13 (6)	12 (4)	18 (9)
Kazal-type protease inhibitors	13 (12)	15 (11)	16 (10)
Cathepsin propeptide inhibitors	4 (2)	4 (1)	4 (3)
Cytoplasmic effectors			
Crinklers	77 (28)	80 (18)	90 (37)
RxLRs	132 (132)	132 (132)	186 (186)
Polysaccharide modifying enzymes			
Cellulases	30 (7)	35 (4)	22 (5)
Lytic polysaccharide mono-oxygenases	6 (5)	4 (2)	5 (5)
Cutinases	3 (2)	4 (2)	5 (5)
Chitinases	1 (1)	2 (0)	2 (1)
Fungal cellulose binding domains	8 (6)	10 (7)	8 (8)
Pectate lyases	34 (23)	25 (14)	31 (23)
Pectin acetylesterases	6 (4)	6 (3)	6 (5)
Pectinesterases	7 (6)	7 (5)	15 (9)

Putative effectors were annotated based on Pfam domains or manually curated (CRNs and RxLRs). Numbers in brackets represent proteins belonging to the predicted secretome.

Elicitins are secreted proteins that bind sterols and lipids, allowing *Phytophthora* spp. to overcome their inability to synthesise sterols by sequestering sterols from their hosts or environments [104]. Elicitins also act as microbe-associated molecular patterns (MAMPs), triggering host cell death upon recognition by the host plant. Elicitin proteins are usually members of large multi-copy gene families in oomycete genomes [105]. Here, we have identified between 45 and 59 proteins with an elicitin domain (PF00964) for each *Phytophthora* species, of which approximately 78% are predicted to be secreted (**Table 3**). In contrast to elicitins, necrosis-inducing proteins (NLPs) have a broad taxonomic distribution having been identified in bacteria, fungi and oomycetes [106]. NLPs are known to induce ethylene accumulation and trigger necrosis in dicots [106]. Here, we identified 25, 33 and 22 proteins containing the NLP domain (PF05630) in *Ph. chlamydospora*, *Ph. gonapodyides* and *Ph. pseudosyringae*, respectively, of which 19 (76%), 22 (67%) and 19 (86%) were predicted to be secreted (**Table 3**). Multiple sequence alignment (not shown) of identified NLPs confirm they are all type 1 NLPs, characterised by the presence of two conserved cysteine residues. Other effectors of interest include the PcF phytotoxins, these are small cysteine-rich proteins that induce plant cell necrosis [107]. PcF phytotoxins appear to be unique to Peronosporales species based on available genomic data. We identified only one protein with the PcF phytotoxin domain (PF09461) in each of the three genomes, however, only the *Ph. gonapodyides* and *Ph. pseudosyringae* copies were predicted to be secreted (**Table 3**). Transglutaminases are proteins that strengthen structures such as cell walls by facilitating cross-linking between glutamine and lysine residues, conferring resistance to proteolysis [108]. Transglutaminases, such as *Ph. sojae* GP42, can elicit a host immune response upon recognition [109]. We identified 14, 16, and 17 proteins containing the transglutaminase elicitor domain (PF16683) in *Ph. chlamydospora*, *Ph. gonapodyides*

and *Ph. pseudosyringae*, respectively (**Table 3**). Each of the three genomes encode 11 transglutaminase elicitors that are predicted to be secreted (**Table 3**), which is similar to the number predicted to be secreted by *Ph. cactorum* (15) [6].

The PAN/Apple domain (PF00024, PF14295) is enriched in the secretomes of most oomycete species [12]. This domain is associated with carbohydrate-binding modules, for example, cellulose-binding elicitor lectins (CBEL). Knockdown of a *Ph. parasitica* CBEL with two PAN/Apple domains affected its ability to adhere to cellulosic substrates, such as plant cell walls [110]. We identified 25 proteins with PAN/Apple domains in *Ph. chlamydospora*, 32 in *Ph. gonapodyides* and 20 in *Ph. pseudosyringae* (**Table 3**), of which 21 (84%), 20 (62.5%) and 15 (75%) were predicted to be secreted (**Table 3**). 47% of all identified PAN/Apple domain-containing proteins had two or more PAN/Apple domains. Proteins belonging to the cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 proteins (CAP) family (PF00188) are also enriched in most oomycete secretomes [12]. *Saccharomyces cerevisiae* CAP family proteins function in sterol binding and export, and are linked to fungal virulence [111]. However, little is known about their involvement in oomycete infection. We identified 34 CAP proteins in *Ph. chlamydospora*, and *Ph. gonapodyides* individually and 31 in *Ph. pseudosyringae* (Table 3), of which 22 (64.7%), 25 (73.5%) and 22 (71.0%) were predicted to be secreted.

In total, we annotated 18, 13 and 19 proteases bearing signal peptides (aspartyl proteases, papain family cysteine proteases or serine proteases) in *Ph. chlamydospora*, *Ph. gonapodyides* and *Ph. pseudosyringae*, respectively (**Table 3**). While these proteins are annotated as proteases, they may not be proteolytically active. For example, two *Ph. sojae* proteins GIP1 and GIP2 share significant similarity with trypsin but are proteolytically non-functional and instead inhibit an endo-β-1,3 glucanase from soybean [11]. In a classic example of a co-evolutionary arms race, plant hosts secrete apoplastic

proteases to degrade pathogen effectors. To counteract these plant defenses, pathogens secrete protease inhibitors to inhibit the host proteases. A total of 13, 15, and 16 proteins were annotated with Kazal-type serine protease inhibitors domains (PF00050) in *Ph. chlamydospora, Ph. gonapodyides,* and *Ph. pseudosyringae*, of which 12 (92.3%), 11 (73.3%) and 10 (62.5%) were predicted to be secreted (**Table 3**). We also identified 4 Cathepsin propeptide inhibitors (PF08246) in each genome (**Table 3**), 2 were predicted to be secreted in *Ph. chlamydospora*, 1 in *Ph. gonapodyides* and 3 in *Ph. pseudosyringae* (**Table 3**).

We also annotated a large number of proteins putatively involved in the breakdown or binding of exogenous carbohydrates, such as plant cell walls, including cellulases, cellulose-binding proteins, cutinases, lytic polysaccharide mono-oxygenases, and pectin modifying enzymes (**Table 3**). Pectin modifying enzymes were the most numerous and include pectate lyases, pectinesterases and pectin acetylesterases. Pectate lyases cleave pectin, a major component of plant cell walls. Pectinesterases catalyse the de-esterification of pectin, while pectin acetylesterases deacetylate pectin, making the pectin backbone more accessible to pectate lyases [112,113]. In total, we identified 47 pectin modifying enzymes in *Ph. chlamydospora*, 38 in *Ph. gonapodyides*, and 52 in *Ph. pseudosyringae* (**Table 3**), of which 33 (70.2%), 22 (57.9%) and 37 (71.2%) are predicted to be secreted, suggesting a putative role in the breakdown of plant cells.

RxLR effectors are named due to the highly conserved RxLR motif found in their N-terminus which act as a trafficking motif, signaling the effectors to be delivered into plant cells [77]. The RxLR motif is followed by an EER motif in many RxLR effectors [77]. RxLR C-terminal domains are typically highly divergent although many contain one or more “WY” domains [114]. Many RxLRs are expressed early in infection and play roles in the suppression of host immune responses [115]. However, the function of most

RxLRs is unknown and many have been shown to localise to diverse subcellular locations within plant host cells [116]. RxLRs were identified using a combination of four independent criteria (see methods). For *Ph. chlamydospora*, 93 proteins had a hit according to the Win method, 68 with the Regex method, 64 with the HMM method and 81 with the homology method (**Supplementary Table S5**). In total, across the four methods, 132 unique putative RxLRs were identified for *Ph. chlamydospora* (**Table 3**), of which 34 had hits to the WY-fold HMM (**Supplementary Table S5**). For *Ph. gonapodyides*, 96 putative RxLRs were identified according to the Win method, 63 with the Regex method, 68 with the HMM method and 74 with the homology method (**Supplementary Table S5**). In total, 132 unique putative RxLRs were identified using the four methods (**Table 3**), 39 of which had hits to the WY-fold HMM (**Supplementary Table S5**). For *Ph. pseudosyringae*, 125 proteins were designated as putative RxLR effectors based on the Win method, 99 with the Regex method, 101 with the HMM method and 124 with the homology method (**Supplementary Table S5**). In total, across the four methods, 186 unique proteins were annotated as putative RxLRs (**Table 3**), of which 61 had hits from the WY-fold HMM (**Supplementary Table S5**). The number of putative RxLRs identified in *Ph. pseudosyringae* is similar to its clade 3 relative *Ph. pluvialis* (181) [8,12].

CRNs are modular proteins that contain highly conserved N-terminal domains containing a signal peptide and an "LxLFLAK" motif that mediates translocation into host cells [117]. CRNs are named after their crinkling and necrosis-inducing activity in leaves [118]. CRNs were identified using a combination of regular expression searches and HMM searches. The number of CRNs identified for each species is similar. In total, 77 putative CRNs were identified in *Ph. chlamydospora*, 80 in *Ph. gonapodyides* and 90 in *Ph. pseudosyringae* (**Table 3**). Similar to what has been observed for other oomycetes

[5,6,12], only a small proportion of identified CRNs have a positive SignalP prediction, with 28 (36.4%) in *Ph. chlamydospora*, 18 (22.5%) in *Ph. gonapodyides* and 37 (41.1%) in *Ph. pseudosyringae* (**Table 3**).

Carbohydrate Active Enzymes

Supplemental to the InterProScan analysis above, a more detailed analysis of CAZymes was performed using dbCAN2 [71]. This led to the identification of 483 putative CAZymes in *Ph. chlamydospora*, 487 in *Ph. gonapodyides* and 453 in *Ph. pseudosyringae* (**Table 4 and Supplementary Table S6**), of which, 213 (44.1%), 179 (36.8%) and 194 (42.8%) are predicted to be secreted (**Table 4 and Supplementary Table S6**). Of the identified CAZymes, glycoside hydrolases are the most numerous. Each species encodes between 210 and 243 glycoside hydrolases, of which 51 to 62% are predicted to be secreted (**Table 4**). Identified glycoside hydrolases belonged to 44 families/subfamilies (**Supplementary Table S6**). In total 33 to 38 polysaccharide lyases were identified for each species (**Table 4**), belonging to three families, PL1_4 (pectate lyase), PL3_2 (pectate lyase) and PL4_1 (rhamnogalacturonan endolyase) (Supplementary Table S6). Approximately 73% of all identified polysaccharide lyases are predicted to be secreted (**Table 4**).

Table 4. Counts of carbohydrate-active enzymes (CAZymes) in *Phytophthora* genomes

	Phytophthora chlamydospora	***Phytophthora gonapodyides***	***Phytophthora pseudosyringae***
Glycoside Hydrolases	234 (144)	243 (123)	210 (121)
Glycosyl Transferases	125 (11)	118 (10)	120 (12)
Polysaccharide Lyases	38 (27)	33 (22)	33 (27)
Carbohydrate Esterases	46 (16)	46 (12)	53 (22)
Auxiliary Activities	33 (14)	41 (11)	29 (11)
Carbohydrate-Binding Modules	10 (2)	12 (4)	10 (1)
Total CAZymes	**483 (213)**	**487 (179)**	**453 (194)**

Numbers in brackets represent proteins belonging to the predicted secretome. Full list of annotations per gene is available in **Supplementary Table S6**.

We extended this analysis by comparing the CAZyme repertoires of 44 oomycete species with different host ranges and broad lifestyles (**Figure 4 and Supplementary Table S1**). In agreement with previous studies [119], our results show that *Phytophthora* species tend to have larger numbers of CAZymes compared to other oomycete taxa (**Figure 4**). Specifically examining 24 *Phytophthora* species, the average number of CAZymes is 450. On average, each *Phytophthora* genome encodes 215 glycoside hydrolases, 111 glycoside transferases, 38 polysaccharide lyases, 34 proteins involved in auxiliary activities, 46 carbohydrate esterases and 9 proteins with carbohydrate-binding modules. The number of CAZymes identified in *Ph. pseudosyringae* is close to the average *Phytophthora* (**Figure 4**).

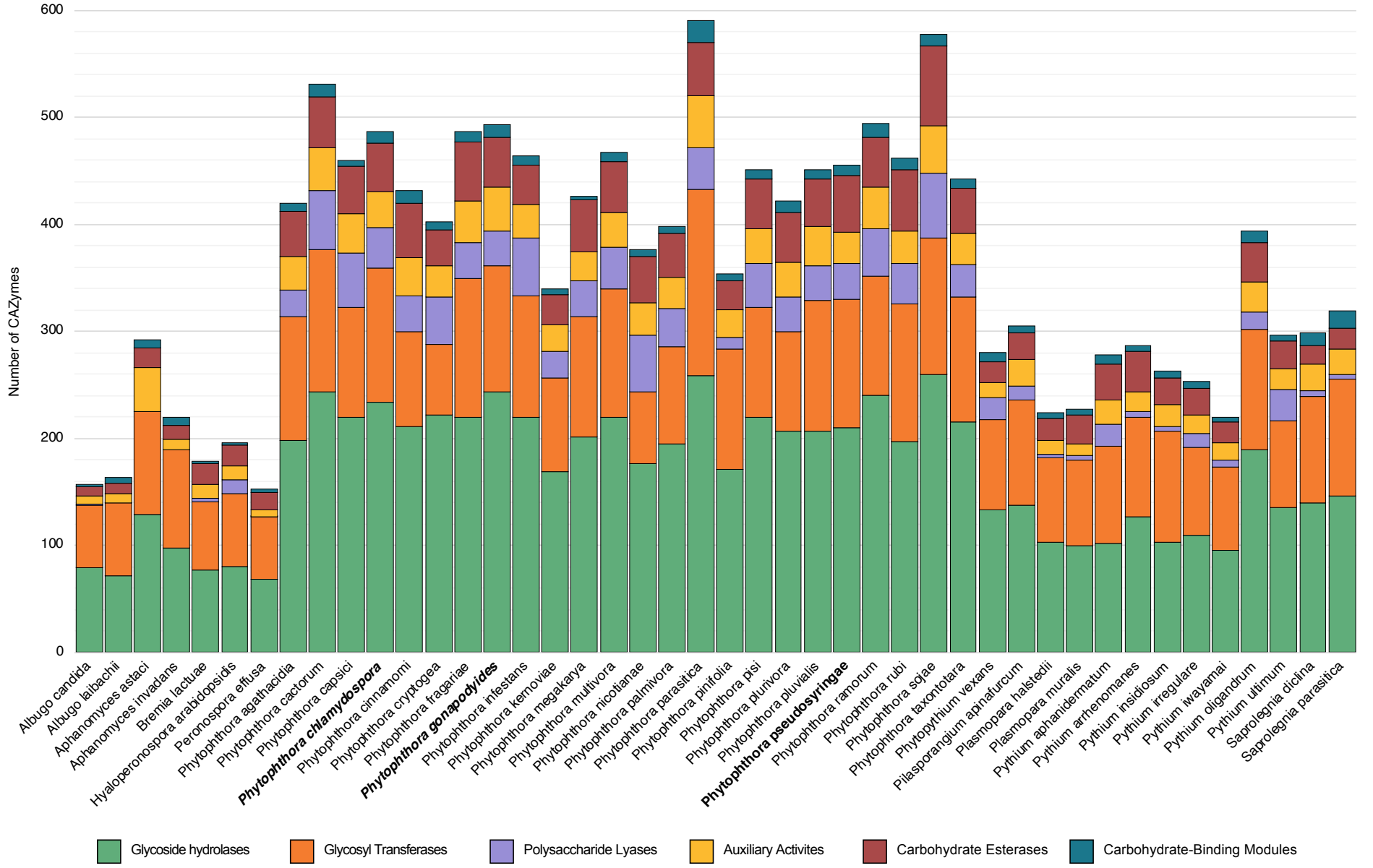


Figure 4. Distribution of CAZymes across 44 oomycete genomes. The arsenal of carbohydrate active enzymes in *Phytophthora* species is larger than what is observed in other oomycete species. In particular the difference in glycoside hydrolase repertoire is noticeable.

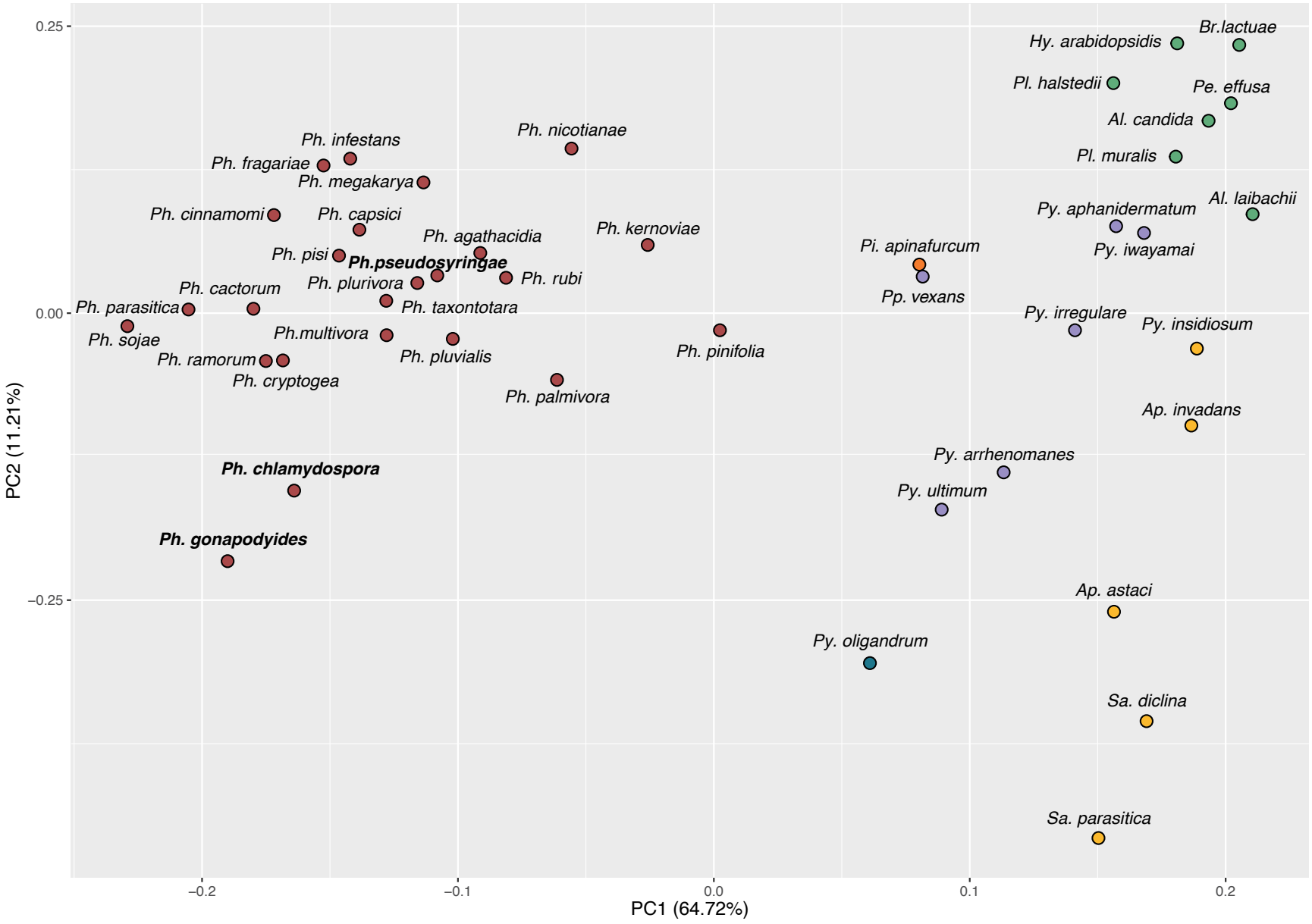

0.25
0.00
−0.25
PC2 (11.21%)
−0.2
−0.1
0.0
0.1
0.2
PC1 (64.72%)
Br.lactuae
Hy. arabidopsidis
Pl. halstedii
Pe. effusa
Al. candida
Pl. muralis
Al. laibachii
Py. aphanidermatum
Py. iwayamai
Pi. apinafurcum
Pp. vexans
Py. irregulare
Py. insidiosum
Ap. invadans
Py. arrhenomanes
Py. ultimum
Ap. astaci
Py. oligandrum
Sa. diclina
Sa. parasitica
Ph. nicotianae
Ph. infestans
Ph. fragariae
Ph. megakarya
Ph. cinnamomi
Ph. capsici
Ph. agathacidia
Ph. kernoviae
Ph. pisi
Ph.pseudosyringae
Ph. rubi
Ph. plurivora
Ph. cactorum
Ph. taxontotara
Ph. parasitica
Ph. sojae
Ph.multivora
Ph. pluvialis
Ph. ramorum
Ph. cryptogea
Ph. pinifolia
Ph. palmivora
Ph. chlamydospora
Ph. gonapodyides
Hemibiotrophic plant pathogen
Biotrophic plant pathogen
Necrotrophic plant pathogen
Mycoparasite
Animal Pathogen
Saprophyte

Figure 5. PCA clustering of oomycete species based on copy numbers of glycoside hydrolase families. Genus abbreviations are as follows: *Albugo* (*Al.*), *Aphanomyces* (*Ap.*), *Bremia* (*Br.*), *Hyaloperonospora* (*Hy.*), *Peronospora* (*Pe.*), *Phytophthora* (*Ph.*), *Phytopythium* (*Pp.*), *Pilasporangium* (*Pi.*), *Plasmopara* (*Pl.*), *Pythium* (*Py.*), *Saprolegnia* (*Sp.*). Colored dots relate to lifestyle. Species with similar lifestyles are clustered together suggesting oomycete lifestyles may be linked to their glycoside hydrolase repertoires.

The two clade 6 aquatic *Phytophthora* species have an expanded repertoire of CAZymes relative to the average *Phytophthora*, with 483 CAZymes in *Ph. chlamydospora* and 487 CAZymes for *Ph. gonapodyides* (**Figure 4**). In particular, they have a higher than average number of glycoside hydrolases, with 234 in *Ph. chlamydospora* and 243 in *Ph. gonapodyides* (**Figure 4 and Table 4**). All three genomes have a higher than average number of glycoside transferases, with 125 in *Ph. chlamydospora*, 118 in *Ph. gonapodyides* and 120 in *Ph. pseudosyringae* (**Figure 4 and Table 4**). *Ph. gonapodyides* also has a higher than average number of proteins involved in auxiliary activities with 41 proteins (**Figure 4 and Table 4**).

Interestingly, principal component analysis (PCA) of glycoside hydrolase family copy numbers clusters species with similar lifestyles together (**Figure 5**). All downy mildew species (*Albugo*, *Bremia*, *Hyaloperonospora*, *Peronospora*, and *Plasmopara*) cluster tightly together (**Figure 5**), despite evolving obligate biotrophism independently (**Figure 2**) [96,120]. Plant pathogenic *Pythium* species are also more distantly clustered (**Figure 5**). The mycoparasite *Pythium oligandrum* is placed distantly to all other *Pythium* species (**Figure 5**), suggesting it may have a specialised repertoire of glycoside hydrolases involved in infection of fungi and oomycetes [121]. The animal pathogens *Aphanomyces astaci*, *Saprolegnia diclina* and *Saprolegnia parasitica* are clustered together (**Figure 5**). *Aphanomyces invadans* is clustered with the mammalian pathogen *Pythium insidiosum* (**Figure 5**). The intermediate genera, *Pilasporangium* and *Phytopythium*, are clustered together between *Phytophthora* and *Pythium* (**Figure 5**). All *Phytophthora* species are clustered together, over a wide area (**Figure 5**). *Phytophthora chlamydospora* and *Ph. gonapodyides* are clustered together but relatively distant to all other *Phytophthora* species (**Figure 5**). This suggests that these opportunistic aquatic *Phytophthora* species may have distinctive glycoside hydrolase arsenals. Furthermore,

they are placed distantly from their closest relative in the dataset *Ph. pinifolia* (**Figure 2 and Figure 5**). Examining individual glycoside hydrolase families, both *Ph. chlamydospora* and *Ph. gonapodyides* have expansions of glycoside hydrolase families 1, 3, 5, 10, 13 and 43 (**Supplementary Figure S2**). These results suggest that oomycete lifestyles may be linked to their CAZyme repertoires, in particular glycoside hydrolase families. These findings are similar to recent analyses showing clustering of oomycete species with similar lifestyles based on metabolic networks [122,123].

Tandemly Duplicated Genes

Tandemly duplicated genes are duplicated genes that are located adjacent to each other in the genome. Analysis of tandemly duplicated genes using BLASTp led to the identification of 2513 (14.1%) tandemly duplicated genes in *Ph. chlamydospora*, 1863 (8.0%) in *Ph gonapodyides* and 2225 (12.8%) in *Ph. pseudosyringae* (**Table 5**). The tandemly duplicated genes are located in 833 to 979 tandem clusters (**Table 5 and Supplementary Table S7**). On average, each cluster has between 2.24 and 2.57 tandemly duplicated genes (**Table 5**). Fewer tandemly duplicated genes were identified in *Ph. gonapodyides* compared to *Ph. chlamydospora* and *Ph. pseudosyringae* (**Table 5**), however, this analysis may have been limited by the poor assembly contiguity of *Ph. gonapodyides*. Overall, counts of tandemly duplicated genes are similar to those observed in other *Phytophthora* genomes [124]. Proteins predicted to be secreted are significantly overrepresented in tandemly duplicated clusters ($p < 0.05$ (Fisher's exact test)), with 354 *Ph. chlamydospora* secreted proteins found in tandem clusters, 265 from *Ph. gonapodyides* and 328 from *Ph. pseudosyringae* (**Table 5**). This adds further evidence that tandem gene duplication has played a role in the expansion of oomycete secretomes [124]. Our results show that putative effector proteins are numerous in tandem clusters.

For example, in *Ph. chlamydospora*, 33 elicitins out of a total of 57 (58%) are located in 12 tandem clusters (**Supplementary Table S7**). Clusters of elicitins genes have also been reported in other *Phytophthora* species [6,105]. Similarly, 20 out of a total 34 CAP proteins (59%) are found in eight tandem clusters (**Supplementary Table S7**). Tandem gene duplication has also played a role in the expansion of *Phytophthora* CAZyme arsenals. For example, in *Ph. chlamydospora* 175 out of a total of 483 proteins (36%) annotated as putative CAZymes (**Table 4)** are found in tandem clusters. We observed similar trends in all three genome assemblies.

Table 5. Tandem gene duplication.

	Phytophthora chlamydospora	***Phytophthora gonapodyides***	***Phytophthora pseudosyringae***
Tandem Clusters	979	833	874
Genes in Tandem Clusters	2513 (14.1%)	1863 (8.0%)	2225 (12.8%)
Average Number of Genes Per Tandem Cluster	2.57	2.24	2.55
Secreted Proteins in Tandem Clusters	354 (31.1%)	265 (20.5%)	328 (29.0%)

LC-MS/MS Characterisation of *Phytophthora* Extracellular Proteomes

Here we used a mass spectrometry-based approach to characterise the in vivo secretomes and extracellular proteomes of *Ph. chlamydospora*, *Ph. gonapodyides* and *Ph. pseudosyringae*. Each species was cultured under two conditions with different media types –10% V8 juice or 10% cV8 juice. Extracellular medium was harvested 10 days after inoculation. To minimise the possibility of hyphal lysis, extracellular medium was

carefully harvested using a syringe without disrupting the *Phytophthora* hyphae. Proteins were extracted from extracellular medium and subjected to LC-MS/MS to identify extracellular proteins. Proteins were identified by searching spectra against the predicted proteome of each species. Protein groups (a group of indistinguishable proteins based on identified peptides) were considered present in a condition if they were identified based on at least two peptides and present in at least two out of three independent biological replicates. Protein groups were considered unique to a condition if they were not detected in any replicate of other conditions.

Examining *Ph. chlamydospora*, 302 protein groups (327 proteins) were identified in the cV8 samples and 251 protein groups (274 proteins) were identified in the V8 samples (**Supplementary Table S8**). 20 protein groups (20 proteins) were unique to the cV8 samples and 14 protein groups (17 proteins) were unique to the V8 samples (**Supplementary Table S8**). In total across the two conditions, 321 protein groups (351 proteins) were identified (**Table 6 and Supplementary Table S8**), of which 149 proteins (42%) overlap with the predicted secretome of *Ph. chlamydospora*. Reducing the strictness of the SignalP 3 analysis by considering all positive HMM predictions without applying additional constraints, 196 proteins (56%) are predicted to be secreted. This compares favourably to *Ph. plurivora*, where 60% of extracellular proteins identified by LC-MS/MS contained predicted N-terminal signal peptides [16]. The proteins lacking signal peptides may be present in the extracellular medium due to contamination of intracellular proteins caused by hyphal lysis during protein extraction. Additionally, these proteins may have legitimate signal peptides that cannot be detected due to inaccurate gene annotation, i.e. gene models with a truncated N-terminus will lack N-terminal signal peptides [15]. It is also possible that they are leaderless secretory proteins (LSPs), lacking signal peptides that enter non-classical secretory pathways. Extracellular proteins that

lack signal peptides were submitted to SecretomeP 2.0, an ab initio predictor of non-classically secreted proteins [125]. A total 78 proteins (22.2%) had a SecretomeP NN-score greater than 0.5, suggesting that they may be non-classically secreted proteins. It is important to note that SecretomeP has only been trained on mammalian LSPs, therefore its accuracy at predicting oomycete LSPs is unclear.

Table 6. Extracellular *Phytophthora* proteins identified by LC-MS/MS that are putatively involved in osmotrophy or virulence.

	Phytophthora chlamydospora	***Phytophthora gonapodyides***	***Phytophthora pseudosyringae***
Total protein groups identified	321	246	313
Total proteins identified	351	283	331
PAN/Apple domain	10	8	5
Transglutaminase elicitor	6	6	8
Elicitin	5	8	4
Cysteine-rich secretory protein family (CAP)	4	4	4
Necrosis inducing protein	2	4	2
PcF phytotoxin	0	0	1
Ribonuclease	1	1	3
Berberine-like protein	1	1	0
Glycoside hydrolases	60	68	61
Polysaccharide lyases	7	9	5
Carbohydrate esterases	4	4	2
Auxiliary activities	8	5	4
Carbohydrate-binding modules	3	4	3
PHI-Base homologs	140	96	109
Apoplastic proteins	99	90	92
Single transmembrane proteins	33	25	29

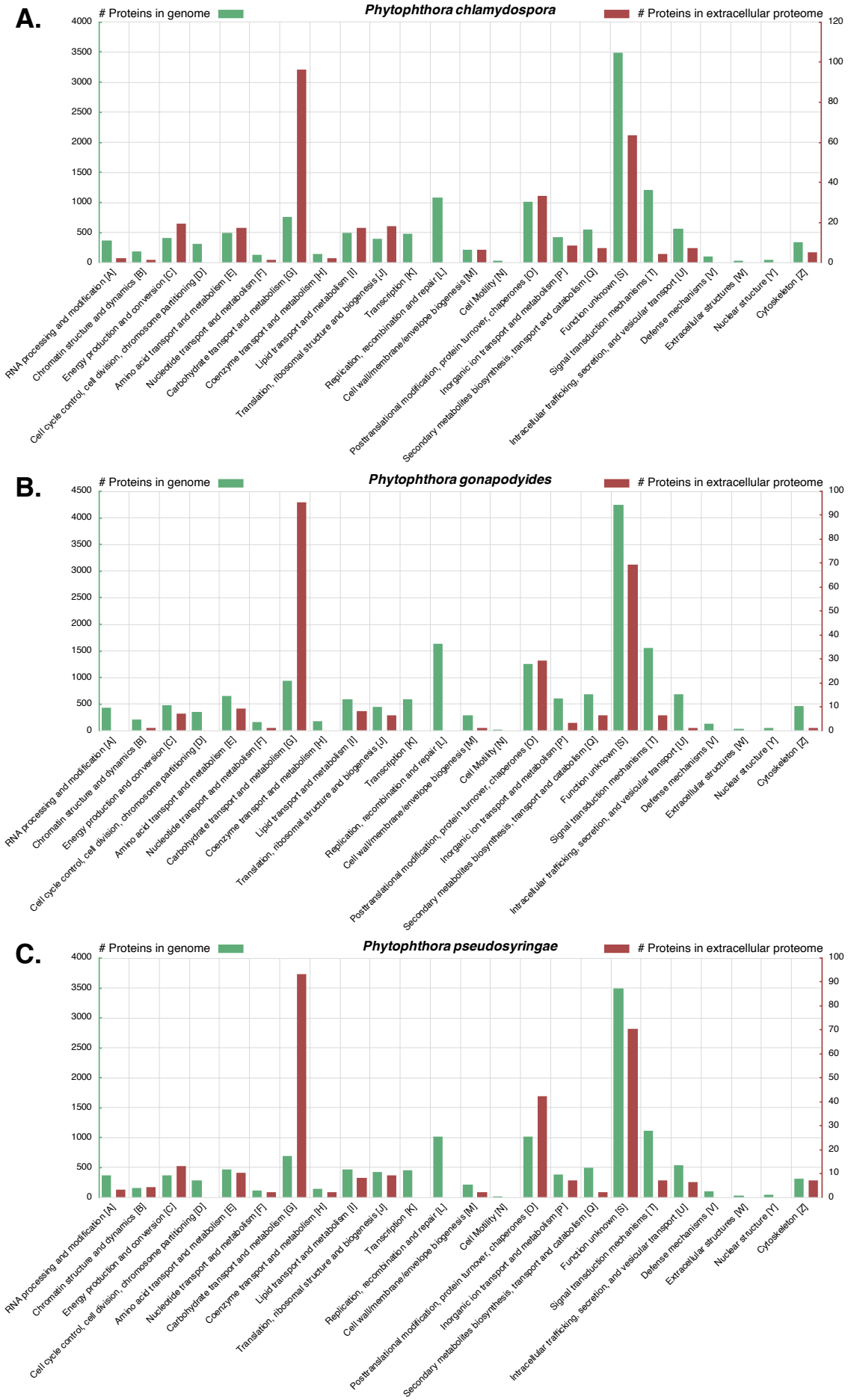


Figure 6. eggNOG functional annotation of the genome and LC-MS/MS extracellular proteome of **(A)** *Phytophthora chlamydospora*, **(B)** *Phytophthora gonapodyides* and **(C)** *Phytophthora pseudosyringae*. Note that counts for extracellular proteins are shown on a secondary axis (right) with a different scale.

eggNOG assigned 298 extracellular proteins (84.9%) to one or more functional COG groups. The most numerous functional categories were carbohydrate transport and metabolism (96 proteins, 32.2%), function unknown (63 proteins, 21.1%), posttranslational modification, protein turnover and chaperones (33 proteins, 11.1%), energy production and conversion (19 proteins, 6.4%), translation, ribosomal structure and biogenesis (18 proteins, 6.0%), amino acid transport and metabolism (17 proteins, 5.7%), and lipid transport and metabolism (17 proteins, 5.7%) (**Figure 6A**). The high number of extracellular proteins involved in transport and metabolism is to be expected for osmotrophs which obtain their nutrients externally from the environment or from their hosts. The proportion of extracellular proteins classified as carbohydrate transport and metabolism (96 proteins; 32.2%) is particularly enriched relative to the total genome (746 proteins; 6.2%) **(Figure 6A**). These proteins may be involved in the breakdown of host plant cells as well in the acquisition and uptake of nutrients. No extracellular proteins (for any species) were annotated as being involved in "cell cycle control, cell division, chromosome partitioning", "transcription", "replication, recombination and repair", "cell motility", "defense mechanisms", "extracellular structures" or " nuclear structure" (**Figure 6A–C**). This suggests that hyphal lysis is unlikely to have occurred during protein extraction as these annotations are associated with intracellular processes or cell structures. A large number of known effector families were also detected, including proteins with PAN/Apple domains (10), transglutaminase elicitors (6), elicitins (5), proteins belonging to the cysteine-rich secretory protein family (4), and NLPs (2) (**Table 6**). An extracellular berberine-like protein was also detected (**Table 6**). Berberine-like proteins were previously reported as putative virulence factors in *Ph. infestans* [69] and are thought to be involved in infection by the biosynthesis of alkaloids and the production of reactive oxygen species. Berberine-like proteins were also detected in the extracellular

proteomes of *Ph. infestans* and *Ph. plurivora* [15,16,69]. Additionally, an extracellular ribonuclease was detected (**Table 6**). Secreted ribonucleases have been reported as effectors in the fungal plant pathogen *Blumeria graminis* [126]. Ribonucleases were also detected in the *Ph. infestans* extracellular proteome [15]. In total, 140 extracellular proteins (40%) have homologs in PHI-Base (Table 6). A large number of extracellular CAZymes were also identified including glycoside hydrolases (60), polysaccharide lyases (7), carbohydrate esterases (4), auxiliary activities (8) and proteins with carbohydrate-binding modules (3) (**Table 6**).

Examining *Ph. gonapodyides*, 237 protein groups (268 proteins) were identified in the cV8 samples and 196 protein groups (230 proteins) were identified in the V8 samples (**Supplementary Table S8**). 17 protein groups (19 proteins) were unique to the cV8 samples and 9 protein groups (15 proteins) were unique to the V8 samples (**Supplementary Table S8**). In total across the two conditions, 246 protein groups (283 proteins) were identified (**Table 6 and Supplementary Table S8**), of which 133 proteins (47%) overlap with the predicted secretome of *Ph. gonapodyides*. 167 proteins (59%) have positive SignalP 3 HMM predictions, ignoring additional cut-offs. An additional 76 proteins (26.9%) have a positive prediction from SecretomeP, suggesting non-classical secretion. Functional annotation using eggNOG assigned 238 extracellular proteins (84.1%) to one or more COG categories. Overall the functional profile was similar to that of *Ph. chlamydospora*, with the most numerous functional categories being carbohydrate transport and metabolism (95 proteins, 39.9%), function unknown (69 proteins, 29.0%), posttranslational modification, protein turnover, chaperones (29 proteins, 12.2%), amino acid transport and metabolism (9 proteins, 3.8%), lipid transport and metabolism (8 proteins, 3.4%) and energy production and conversion (7 proteins, 2.9%) (**Figure 6B**). Identified effector families include proteins with PAN/Apple domains (8),

transglutaminase elicitors (6), elicitins (8), members of the cysteine-rich secretory protein family (4), NLPs (4), a berberine-like protein and a ribonuclease (**Table 6**). Some 96 (34%) extracellular proteins have homologs in PHI-Base (**Table 6**). A similar number of extracellular CAZymes were detected including glycoside hydrolases (68), polysaccharide lyases (9), carbohydrate esterases (4), auxiliary activities (5) and proteins with carbohydrate-binding modules (4) (Table 6).

Examining *Ph. pseudosyringae,* 280 protein groups (296 proteins) were identified in the cV8 samples and 247 protein groups (259 proteins) were identified in the V8 samples (**Supplementary Table S8**). 18 protein groups (22 proteins) were unique to the cV8 samples and 6 protein groups (6 proteins) were unique to the V8 samples (**Supplementary Table S8**). In total across the two conditions, 313 protein groups (331 proteins) were identified (**Table 6 and Supplementary Table S8**), of which 145 proteins (44%) overlap with the predicted secretome of *Ph. pseudosyringae*. 188 proteins (56.8%) proteins have positive SignalP 3 HMM predictions without applying cut offs. An additional 67 proteins (20.2%) have a positive prediction from SecretomeP. eggNOG functional annotation assigned 279 extracellular proteins (84.3%) to one or more COG functional categories. The high-level functional annotation of the *Ph. pseudosyringae* is similar to that of *Ph. chlamydospora* and *Ph. gonapodyides* (**Figure 6C**). The most numerous functional categories are carbohydrate transport and metabolism (93 proteins, 33.3%), function unknown (70 proteins, 25.1%), posttranslational modification, protein turnover, chaperones (42 proteins, 15.1%), energy production and conversion (13 proteins, 4.7%), amino acid transport and metabolism (10 proteins, 3.6%), translation, ribosomal structure and biogenesis (9 proteins, 3.2%) and lipid transport and metabolism (8 proteins, 2.9%) (**Figure 6C**). The number of effector families is also similar and includes proteins with PAN/Apple domains (5), transglutaminase elicitors (8), elicitins

(4), members of the cysteine-rich secretory protein family (4), NLPs (4) and ribonucleases (3) (**Table 6**). We also detected an extracellular PcF phytotoxin from *Ph. pseudosyringae*, which was absent in the extracellular proteomes of both *Ph. chlamydospora* and *Ph. gonapodyides* (**Table 6**). Unlike *Ph. chlamydospora* and *Ph. gonapodyides*, we did not identify any berberine-like extracellular proteins from *Ph. pseudosyringae*, although its predicted secretome encodes a copy (**Table 3**). Some 109 extracellular proteins (33%) have homologs in PHI-Base (**Table 6**). Overall the CAZyme content of the *Ph. pseudosyringae* extracellular proteome is also similar with glycoside hydrolases (61), polysaccharide lyases (5), carbohydrate esterases (2), auxiliary activities (4) and proteins with carbohydrate-binding modules (3) (**Table 6**).

Very few cytoplasmic effectors were detected in our analyses of all three species. CRNs were absent from the extracellular proteomes of all three species. No putative RxLRs were identified in the extracellular proteome of *Ph. chlamydospora* or *Ph. gonapodyides*. Only two putative RxLRs were identified in the extracellular proteome of *Ph. pseudosyringae* PHPS_09091 and PHPS_15662. PHPS_09091 was detected in all replicates of both V8 and cV8 media with 4 unique peptides (**Supplementary Table S8**) and was identified as an RxLR based on the Win, Regex and HMM methods (**Supplementary Table S5**). Similarly, PHPS_15662 was identified in all replicates of both V8 and cV8 media with 4 unique peptides (**Supplementary Table S8**). PHPS_15662 was identified only using the homology method and does not contain a RxLR-like motif (**Supplementary Table S5**), therefore it is not likely to be a legitimate RxLR effector. It is not surprising that so few cytoplasmic effectors were identified as it is possible that most cytoplasmic effectors are secreted from haustoria [127,128]. However, Meijer et al. (2014) report the detection of several *Ph. infestans* RxLRs and CRNs being released from hyphae in the absence of haustoria.

Previously, LC-MS/MS analysis of *Ph. infestans* identified 31 extracellular proteins that contained a single transmembrane domain [15]. These proteins are thought to be membrane proteins that are found in the extracellular medium due to proteolytic ectodomain shedding by sheddases. We also detected a large number of extracellular proteins that contain a single transmembrane domain. This included 33 proteins in *Ph. chlamydospora*, 25 in *Ph. gonapodyides* and 29 in *Ph. pseudosyringae* (Table 6). Of these, 17, 14 and 17 are homologous to the 31 *Ph. infestans* proteins. Similar to what was observed in *Ph. infestans*, the majority of identified transmembrane domains are found in the protein C-terminus.

Interestingly, we detected a number of extracellular proteins with KDEL or KDEL-like (HDEL or SDEL) C-terminal motifs. These are endoplasmic retention (ER) motifs that are usually associated with preventing protein secretion, signalling proteins to be retained in the ER lumen [129]. Proteins with KDEL motifs are usually excluded from in silico secretome studies. In *Ph. chlamydospora* we identified three such proteins PHCH_06832, PHCH_07252, and PHCH_15931. Both PHCH_06832 and PHCH_07252 are paralogs belonging to the same protein group. They were identified in all replicates of both V8 and cV8 media with four unique peptides (**Supplementary Table S8**). Both proteins contain a C-terminal HDEL motif and were annotated as belonging to heat shock protein (Hsp) 70 family. PHCH_15931 was identified in a total of five out of six replicates across the two conditions with five unique peptides and has a C-terminal KDEL motif (**Supplementary Table S8**). It was annotated as a calreticulin, which is an ER associated calcium-binding protein [130]. In *Ph. gonapodyides*, only one such protein was identified, PHGO_06390, which was detected in two out of three replicates of the cV8 samples and has a C-terminal SDEL motif (**Supplementary Table S8**). In *Ph. pseudosyringae*, three proteins were identified PHPS_03476, PHPS_04861, and PHPS_06172. PHPS_04861 is

orthologous to PHGO_06390, has a C-terminal SDEL motif and was identified in all replicates of both conditions with 3 unique peptides (**Supplementary Table S8**). PHPS_03476 has a C-terminal KDEL motif and is orthologous to PHCH_15931 and was identified in all replicates of both conditions, with a total of seven unique peptides (**Supplementary Table S8**). PHPS_06172, a Hsp90 protein, contains a C-terminal KDEL motif and was identified in a total of four replicates across the two conditions, with two peptides, only one of which was unique (**Supplementary Table S8**). Inspecting the *Ph. infestans* extracellular proteins identified by Meijer et al. (2014)[15], there were six extracellular proteins identified that contain C-terminal KDEL/HDEL/SDEL motifs, five of which are orthologous to those identified above. As we detected these KDEL/KDEL-like motif-containing proteins in the extracellular medium, it suggests that their ER retention motifs are masked or perhaps they escape ER retrieval due to saturation of KDEL receptors [130].

LC-MS/MS Identification of Mycelial Proteins

We used mass-spectrometry to characterise the mycelial proteomes of *Ph. chlamydospora*, *Ph. gonapodyides* and *Ph. pseudosyringae*, and to understand how they change in response to oxidative stress and high temperatures. Proteins were extracted from mycelia grown under three conditions: “normal”—mycelia grown for 10 days at optimum temperatures—“heat”—mycelia grown for 7 days at optimum temperatures then switched to 30 °C for 3 days—and “oxidative stress”—mycelia grown for 10 days at optimum temperatures followed by exposure to 1 mM H_2O_2 for 3 h.

Examining *Ph. chlamydospora*, a total of 2592 protein groups (2635 proteins) were identified across the three conditions (**Supplementary Table S9**). Under the normal condition, 2418 protein groups (2461 proteins) were identified (**Supplementary Table**

S9). 123 protein groups (130 proteins) were uniquely detected under the normal conditions (**Supplementary Table S9**) and were significantly enriched for oxidoreductase activity (GO:0016491). Only three protein groups (three proteins) were uniquely detected in the heat-treated samples. These included a protein with an FHA domain (PHCH_03368) and a histidine phosphatase (PHCH_17500) (**Supplementary Table S9**). Only 5 protein groups (5 proteins) were uniquely detected in the oxidative stress samples (**Supplementary Table S9**). These included a CAF1 ribonuclease (PHCH_05536), an integrator complex subunit (PHCH_09875), a protein kinase (PHCH_13661), a DNA photolyase (PHCH_14550) and a methyltransferase (PHCH_16784) (**Supplementary Table S9**). 328 (12.4%) of all identified proteins have one or more predicted transmembrane helices. Furthermore, 131 (5.0%) of all identified proteins belong to the predicted secretome, while 213 (8.1%) proteins were also were also identified in the extracellular proteome. Amongst the identified proteins were several effector families, including NLPs (1), transglutaminase elicitors (3), elicitins (5), PAN domain containing proteins (5), CAP family proteins (5), RxLRs (6), and CRNs (14) (Supplementary Table S9). Additionally, 81 CAZymes were identified from *Ph. chlamydospora* mycelium (**Supplementary Table S9**).

Examining *Ph. gonapodyides*, a total of 2745 protein groups (2,840 proteins) were identified across the three conditions (Supplementary Table S9). Under the normal condition, 2360 protein groups (2436 proteins) were detected (Supplementary Table S9). Only one protein was uniquely detected in this condition, a member of glycoside hydrolase family 31 (PHGO_11274). Under heat treatment 12 protein groups (12 proteins) were uniquely detected, 3 of these contain predicted transmembrane helices and 2 are predicted to be secreted. Included amongst these proteins, were an acetyltransferase (PHGO_14090), an ABC transporter (PHGO_22460), an auto-transporter adhesin

(PHGO_20954), a metallopeptidase (PHGO_13577), a starch-binding protein (PHGO_10964), an exoribonuclease (PHGO_00249), a Maf like protein (PHGO_22144) and an Arf GTPase activating protein (PHGO_17826) (**Supplementary Table S9**). In response to oxidative stress, 34 protein groups (37 proteins) were uniquely detected. Amongst these, 12 proteins (32.4%) have one or more predicted transmembrane helices and three belong to the predicted secretome. These proteins include two peroxidases (PHGO_01245 and PHGO_22132), both of which are predicted to be secreted. Similar to *Ph. chlamydospora*, 346 (12.2%) of all identified *Ph. gonapodyides* mycelial proteins have one or more predicted transmembrane helices. Furthermore, 131 (4.6%) of identified mycelial proteins also belong to the predicted secretome and 144 (5.1%) were also identified in the extracellular proteome (**Supplementary Table S9**). We also detected a number of effector families, including CRNs (2), RxLRs (2), transglutaminase elicitors (4), PAN domain-containing proteins (5), and elicitins (7) (Supplementary Table S9). Additionally, 89 CAZymes were identified from *Ph. gonapodyides* mycelium (**Supplementary Table S9**).

Examining *Ph. pseudosyringae*, a total of 3195 protein groups (3245 proteins) were identified across the three conditions (**Supplementary Table S9**). Under the normal condition, 2223 protein groups (2248 proteins) were detected, 22 protein groups (22 proteins) of which were uniquely detected in this condition (**Supplementary Table S9**). 32 protein groups (33 proteins) were uniquely detected in the heat-treated samples which included three proteins predicted to be secreted and four proteins with predicted transmembrane helices. Heat-treated samples were significantly enriched for chaperone binding (GO:0051087). We detected 103 unique protein groups (106 proteins) in response to H_2O_2 treatment, including three proteins predicted to be secreted and 30 proteins with at least one predicted transmembrane helix. Of the proteins identified across

all conditions, 445 (13.7%) of mycelial proteins contain one or more predicted transmembrane helices. Furthermore, 160 (4.9%) belong to the predicted secretome, and 204 (6.3%) were also identified in the extracellular proteome (**Supplementary Table S9**). We also detected effector families, including NLPs (2), CAP family proteins (3), elicitins (4), transglutaminase elicitors (5), PAN domain-containing proteins (5), RxLRs (8) and CRNs (18) (Supplementary Table S9). Additionally, 103 CAZymes were identified from *Ph. pseudosyringae* mycelium (**Supplementary Table S9**).

Overall, the functional annotation of all identified mycelial proteins is similar between each of the three species (**Figure 7A**). Clustering with MCL grouped identified mycelial proteins from the three species into 2577 protein families, of which 1554 families were shared by all three species (**Figure 7B**). More proteins were common between *Ph. chlamydospora* and *Ph. pseudosyringae* (205) and between *Ph. gonapodyides* and *Ph. pseudosyringae* (185) than between *Ph. chlamydospora* and *Ph. gonapodyides* (103) (**Figure 7B**). 93 mycelial protein families were unique to *Ph. chlamydospora*, 84 to *Ph. gonapodyides* and 353 to *Ph. pseudosyringae* (**Figure 7B**), indicating increased variation across *Phytophthora* clades.

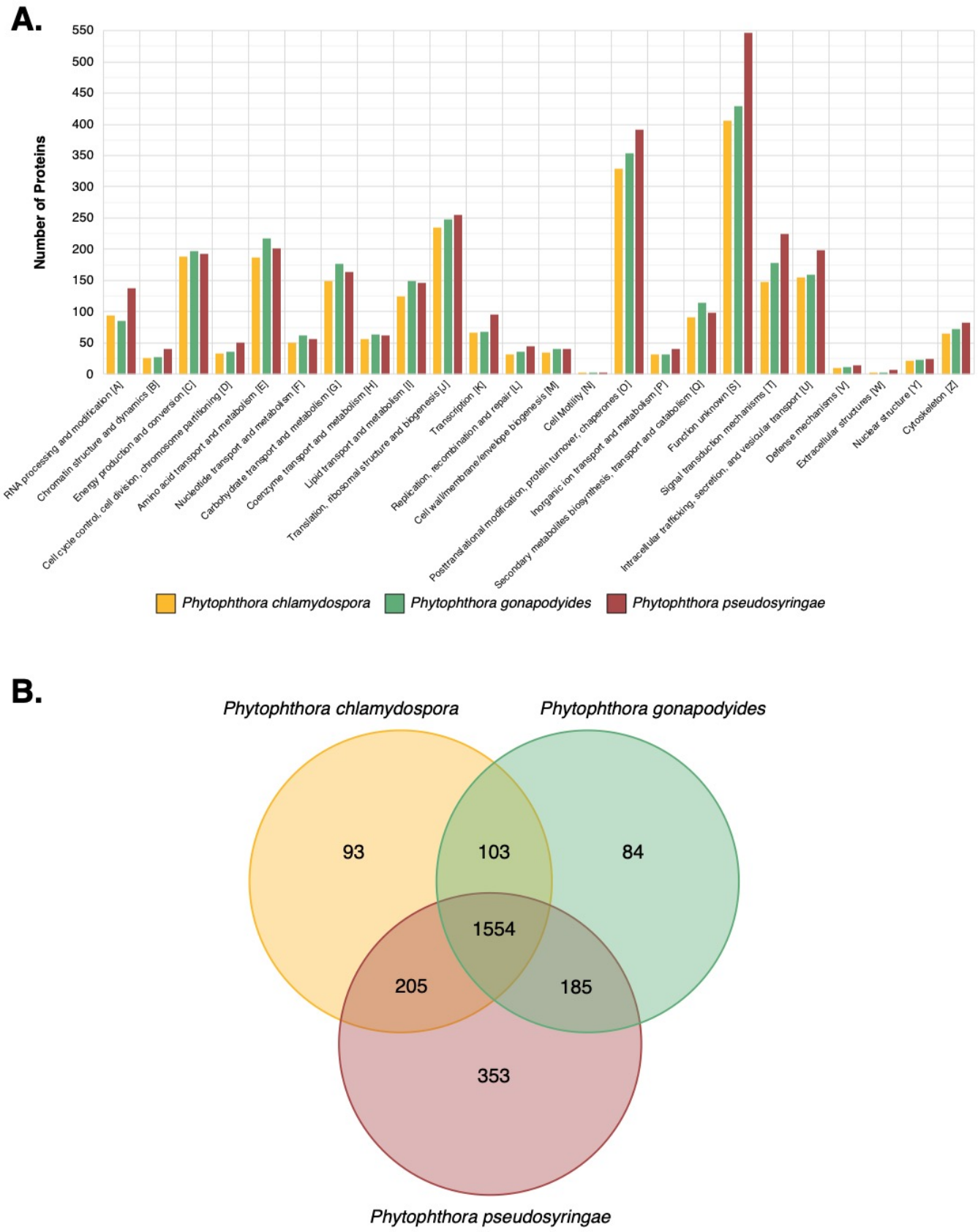


Figure 7. Functional annotation of identified mycelial proteins. **(A)** Identified mycelial proteins were functionally annotated using eggNOG-Mapper. **(B)** Venn diagram showing the number of identified mycelial protein families shared between each species.

Phylostratigraphy Analysis

Taxonomically restricted genes were identified using phylostratigraphy. Homologs were identified for each *Phytophthora* protein-coding gene by performing BLAST searches against a large protein database (18,084,866 proteins) with broad phyletic distribution [85]. Genes were assigned to one of seven phylostrata (cellular organisms, eukaryotes, Stramenopiles, oomycetes, Peronosporales, *Phytophthora* or species-specific orphans) based on their conservation in other taxonomic lineages.

Overall the proportion of genes assigned to each phylostrata are similar between the three *Phytophthora* genomes (**Figure 8**). On average, 33.4% of genes were assigned to the phylostratum cellular organisms (i.e., homologs are present in eukaryotes and prokaryotes), 29.9% are unique to eukaryotes, 1.6% are unique to stramenopiles, 23.6% are unique to oomycetes, 3.9% are unique to Peronosporales, 5.9% are unique to *Phytophthora* and 1.7% are orphans unique to one *Phytophthora* species (**Figure 8**). Individually, 194 orphans (1.1%) were identified for *Ph. chlamydospora*, 520 (2.2%) for *Ph. gonapodyides* and 312 (1.8%) for *Ph. pseudosyringae* (**Figure 8**). *Ph. gonapodyides* had a smaller proportion of genes identified as originating in cellular organisms (30.4%), compared to *Ph. chlamydospora* (35.8%) and *Ph. pseudosyringae* (34.1%) (**Figure 8B**). In addition, *Ph. gonapodyides* had a higher proportion of genes identified as being unique to oomycetes (25.1%), unique to *Phytophthora* (7.4%) and species-specific (2.2%) (**Figure 8B**). This suggests that the increased gene repertoire of *Ph. gonapodyides* is due to expansions of more recently evolved gene families, i.e., genes unique to oomycetes, *Phytophthora*, and *Ph. gonapodyides*, as opposed to expansions of more ancient genes that are conserved in other eukaryotes or prokaryotes. Overall the proportion of genes per phylostratum is similar to previous phylostratigraphic analyses for other *Phytophthora* genomes [124].

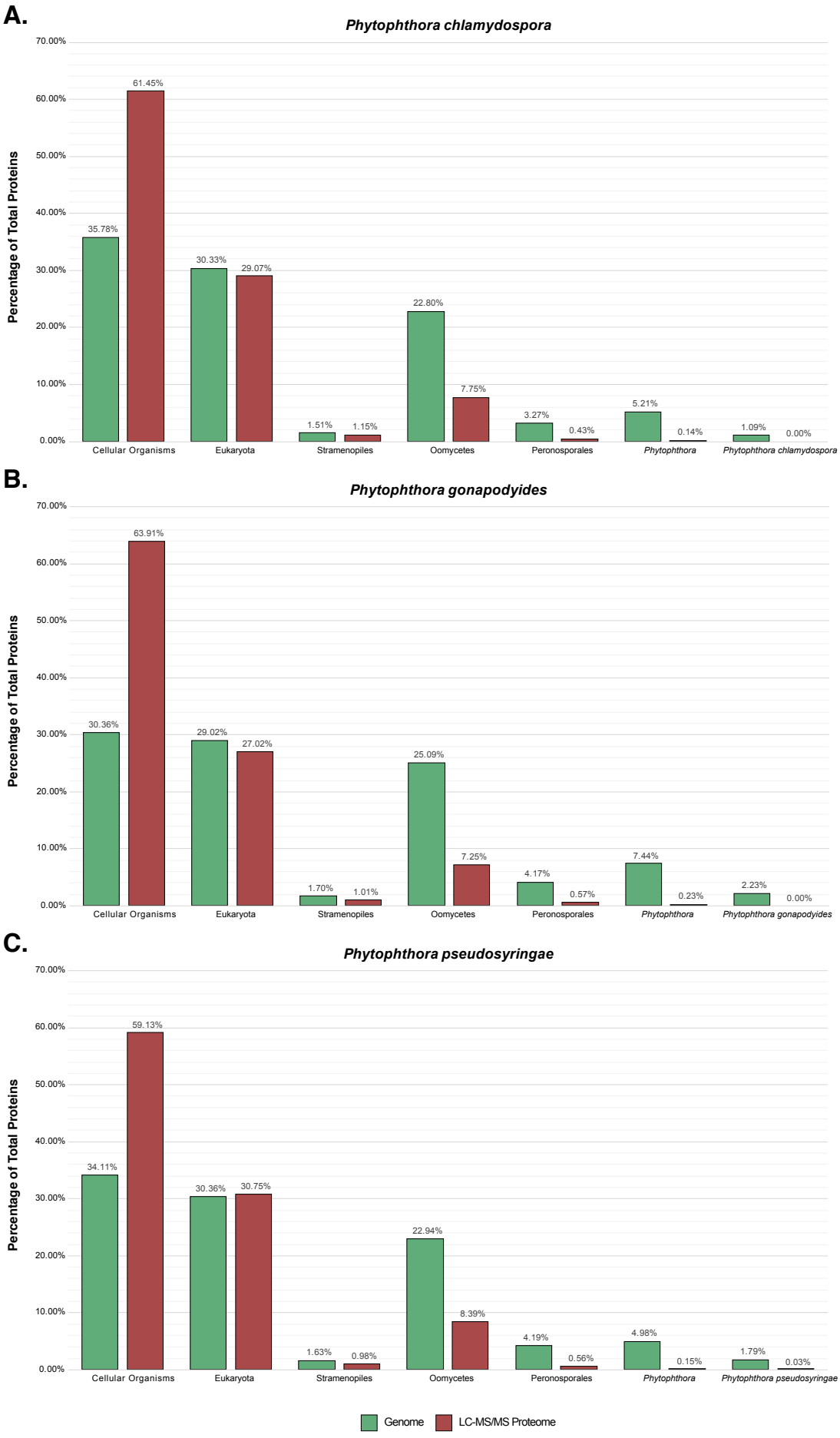


Figure 8. Phylostratigraphy analysis of *Ph. chlamydospora* **(A)**, *Ph. gonapodyides* **(B)**, and *Ph. pseudosyringae* **(C)**, showing the proportion of genes assigned to each phylostratum for both the genome and the LC-MS/MS proteome.

We coupled our phylostratigraphy analysis with the mass spectrometry data. Compared to the overall genome, a much larger proportion (approximately 61.5%) of identified proteins belongs to the phylostratum "cellular organisms" (**Figure 8**). This suggests that the majority of identified proteins are evolutionarily conserved proteins that possibly play roles in conserved housekeeping functions. Only 8.1% to 9.1% of identified proteins belong to the phylostratum oomycetes or "younger" (**Figure 8**). Furthermore, only 0.57% to 0.80% of identified proteins belong to the phylostratum Peronosporales or "younger" (**Figure 8**). This suggests that the more recently evolved genes may be under tighter transcriptional control or expressed only in specific scenarios. Additionally, some of the proteins identified as being species-specific may not be legitimate genes.

Conclusions

Here, we have sequenced the genomes for three ubiquitous *Phytophthora* species—*Ph. chlamydospora*, *Ph. gonapodyides* and *Ph. pseudosyringae*. Using bioinformatics methods, comparative genomics, and mass spectrometry, we provide a comprehensive characterization of the nuclear genomes, mitochondrial genomes, in silico secretomes, extracellular proteomes, and mycelial proteomes of each species. These genome resources will be useful for future studies to understand the lifestyles of these widespread *Phytophthora* species.

Acknowledgments

J.M. is funded by a postgraduate scholarship from the Irish Research Council, Government of Ireland (grant number: GOIPG/2016/1112). Mass spectrometry facilities were funded by a Science Foundation Ireland infrastructure award (SFI 12/RI/2346(3)). We acknowledge the DJEI/DES/SFI/HEA Irish Centre for High-End Computing (ICHEC) for the provision of computational facilities and support.

Supplementary Data

Supplementary data for this chapter is available from the publisher website: https://www.mdpi.com/2076-2607/8/5/653

Supplementary data is also available at the following GitHub repository: https://github.com/jamiemcg/ThesisSupplementaryMaterial

Supplementary Figures

Supplementary Figure S1: Nucleotide alignment of *Phytophthora* mitochondrial genomes.

Supplementary Figure S2: Distribution of CAZyme families across 44 oomycete genomes.

Supplementary Tables

Supplementary Table S1: List of oomycete genomes used in this study.

Supplementary Table S2: Repetitive element annotation.

Supplementary Table S3: Functional annotation of gene models.

Supplementary Table S4: Annotation of mitochondrial genes.

Supplementary Table S5: Annotation of RxLR effectors.

Supplementary Table S6: Annotation of CAZymes using dbCAN2.

Supplementary Table S7: List of clusters of tandemly duplicated genes.

Supplementary Table S8: List of identified extracellular protein groups.

Supplementary Table S9: List of identified mycelial protein groups.

References

1. Burki, F.; Roger, A.J.; Brown, M.W.; Simpson, A.G.B. The New Tree of Eukaryotes. *Trends Ecol. Evol.* **2020**, *35*, 43–55.
2. Yang, X.; Tyler, B.M.; Hong, C. An expanded phylogeny for the genus Phytophthora. *IMA Fungus* **2017**, *8*, 355–384.
3. Haas, B.J.; Kamoun, S.; Zody, M.C.; Jiang, R.H.Y.; Handsaker, R.E.; Cano, L.M.; Grabherr, M.; Kodira, C.D.; Raffaele, S.; Torto-Alalibo, T.; et al. Genome sequence and analysis of the Irish potato famine pathogen Phytophthora infestans. *Nature* **2009**, *461*, 393–398.
4. Tyler, B.M. Phytophthora Genome Sequences Uncover Evolutionary Origins and Mechanisms of Pathogenesis. *Science (80-.).* **2006**, *313*, 1261–1266.
5. Vetukuri, R.R.; Tripathy, S.; Mathu, M.C.; Panda, A.; Kushwaha, S.K.; Chawade, A.; Andreasson, E.; Grenville-Briggs, L.J.; Whisson, S.C. Draft genome sequence for the tree pathogen phytophthora plurivora. *Genome Biol. Evol.* **2018**, *10*, 2432–2442.
6. Armitage, A.D.; Lysøe, E.; Nellist, C.F.; Lewis, L.A.; Cano, L.M.; Harrison, R.J.; Brurberg, M.B. Bioinformatic characterisation of the effector repertoire of the strawberry pathogen Phytophthora cactorum. *PLoS One* **2018**, *13*, 1–24.
7. Feau, N.; Taylor, G.; Dale, A.L.; Dhillon, B.; Bilodeau, G.J.; Birol, I.; Jones, S.J.M.; Hamelin, R.C. Genome sequences of six Phytophthora species threatening forest ecosystems. *Genomics Data* **2016**, *10*, 85–88.
8. Studholme, D.J.; McDougal, R.L.; Sambles, C.; Hansen, E.; Hardy, G.; Grant, M.; Ganley, R.J.; Williams, N.M. Genome sequences of six Phytophthora species associated with forests in New Zealand. *Genomics Data* **2015**, *7*, 54–56.
9. Ali, S.S.; Shao, J.; Lary, D.J.; Kronmiller, B.A.; Shen, D.; Strem, M.D.; Amoako-Attah, I.; Akrofi, A.Y.; Begoude, B.A.D.; ten Hoopen, G.M.; et al. Phytophthora megakarya and Phytophthora palmivora, Closely Related Causal Agents of Cacao Black Pod Rot, Underwent Increases in Genome Sizes and Gene Numbers by Different Mechanisms. *Genome Biol. Evol.* **2017**, *9*, 536–557.
10. Richards, T.A.; Soanes, D.M.; Jones, M.D.M.; Vasieva, O.; Leonard, G.; Paszkiewicz, K.; Foster, P.G.; Hall, N.; Talbot, N.J. Horizontal gene transfer facilitated the evolution of plant parasitic mechanisms in the oomycetes. *Proc. Natl. Acad. Sci.* **2011**, *108*, 15258–15263.

11. Kamoun, S. A catalogue of the effector secretome of plant pathogenic oomycetes. *Annu. Rev. Phytopathol.* **2006**, *44*, 41–60.
12. McGowan, J.; Fitzpatrick, D.A. Genomic, Network, and Phylogenetic Analysis of the Oomycete Effector Arsenal. *mSphere* **2017**, *2*, e00408-17.
13. Wawra, S.; Belmonte, R.; Löbach, L.; Saraiva, M.; Willems, A.; van West, P. Secretion, delivery and function of oomycete effector proteins. *Curr. Opin. Microbiol.* **2012**, *15*, 685–691.
14. Stassen, J.H.M.; Van den Ackerveken, G. How do oomycete effectors interfere with plant life? *Curr. Opin. Plant Biol.* **2011**, *14*, 407–414.
15. Meijer, H.J.G.; Mancuso, F.M.; Espadas, G.; Seidl, M.F.; Chiva, C.; Govers, F.; Sabido, E. Profiling the Secretome and Extracellular Proteome of the Potato Late Blight Pathogen Phytophthora infestans. *Mol. Cell. Proteomics* **2014**, *13*, 2101–2113.
16. Severino, V.; Farina, A.; Fleischmann, F.; Dalio, R.J.D.; Di Maro, A.; Scognamiglio, M.; Fiorentino, A.; Parente, A.; Osswald, W.; Chambery, A. Molecular profiling of the Phytophthora plurivora secretome: a step towards understanding the cross-talk between plant pathogenic oomycetes and their hosts. *PLoS One* **2014**, *9*, e112317.
17. Hosseini, S.; Resjö, S.; Liu, Y.; Durling, M.; Heyman, F.; Levander, F.; Liu, Y.; Elfstrand, M.; Funck Jensen, D.; Andreasson, E.; et al. Comparative proteomic analysis of hyphae and germinating cysts of Phytophthora pisi and Phytophthora sojae. *J. Proteomics* **2015**, *117*, 24–40.
18. Resjö, S.; Brus, M.; Ali, A.; Meijer, H.J.G.; Sandin, M.; Govers, F.; Levander, F.; Grenville-Briggs, L.; Andreasson, E. Proteomic Analysis of Phytophthora infestans Reveals the Importance of Cell Wall Proteins in Pathogenicity. *Mol. Cell. Proteomics* **2017**, *16*, 1958–1971.
19. Savidor, A.; Donahoo, R.S.; Hurtado-Gonzales, O.; Land, M.L.; Shah, M.B.; Lamour, K.H.; McDonald, W.H. Cross-species Global Proteomics Reveals Conserved and Unique Processes in *Phytophthora sojae* and *Phytophthora ramorum*. *Mol. Cell. Proteomics* **2008**, *7*, 1501–1516.
20. Jung, T.; Nechwatal, J.; Cooke, D.E.L.; Hartmann, G.; Blaschke, M.; Oßwald, W.F.; Duncan, J.M.; Delatour, C. Phytophthora pseudosyringae sp. nov., a new species causing root and collar rot of deciduous tree species in Europe. *Mycol. Res.* **2003**, *107*, 772–789.

21. Hansen, E.M.; Reeser, P.W.; Sutton, W. Ecology and pathology of Phytophthora ITS clade 3 species in forests in western Oregon, USA. *Mycologia* **2017**, *109*, 100–114.
22. O'Hanlon, R.; Choiseul, J.; Corrigan, M.; Catarame, T.; Destefanis, M. Diversity and detections of Phytophthora species from trade and non-trade environments in Ireland. *EPPO Bull.* **2016**, *46*, 594–602.
23. O'Hanlon, R.; McCracken, A.; Cooke, L. Diversity and ecology of Phytophthora species on the island of Ireland. *Biol. Environ. Proc. R. Irish Acad.* **2016**, *116B*, 27.
24. Fajardo, S.N.; Valenzuela, S.; Dos Santos, A.F.; González, M.P.; Sanfuentes, E.A. Phytophthora pseudosyringae associated with the mortality of Nothofagus obliqua in a pure stand in central-southern Chile. *For. Pathol.* **2017**, *47*, e12361.
25. Varela, C.P.; Vázquez, J.P.M.; Casal, O.A.; Martínez, C.R. First Report of Phytophthora pseudosyringae on Chestnut Nursery Stock in Spain. *Plant Dis.* **2007**, *91*, 1517–1517.
26. Redondo, M.Á.; Boberg, J.; Stenlid, J.; Oliva, J. First Report of Phytophthora pseudosyringae Causing Basal Cankers on Horse Chestnut in Sweden. *Plant Dis.* **2016**, *100*, 1024.
27. Scanu, B.; Webber, J.F. Dieback and mortality of Nothofagus in Britain: ecology, pathogenicity and sporulation potential of the causal agent Phytophthora pseudosyringae. *Plant Pathol.* **2016**, *65*, 26–36.
28. Riddell, C.E.; Frederickson-Matika, D.; Armstrong, A.C.; Elliot, M.; Forster, J.; Hedley, P.E.; Morris, J.; Thorpe, P.; Cooke, D. EL; Pritchard, L.; et al. Metabarcoding reveals a high diversity of woody host-associated Phytophthora spp. in soils at public gardens and amenity woodlands in Britain. *PeerJ* **2019**, *7*, e6931.
29. Brasier, C.M.; Cooke, D.E.L.; Duncan, J.M.; Hansen, E.M. Multiple new phenotypic taxa from trees and riparian ecosystems in Phytophthora gonapodyides-P. megasperma ITS Clade 6, which tend to be high-temperature tolerant and either inbreeding or sterile. *Mycol. Res.* **2003**, *107*, 277–290.
30. Hansen, E.M.; Reeser, P.W.; Sutton, W. Phytophthora Beyond Agriculture. *Annu. Rev. Phytopathol.* **2012**, *50*, 359–378.
31. Hansen, E.M.; Reeser, P.; Sutton, W.; Brasier, C.M. Redesignation of Phytophthora taxon Pgchlamydo as Phytophthora chlamydospora sp. nov. *North Am. Fungi* **2015**, *10*, 1–14.
32. Kurbetli, İ.; Aydoğdu, M.; Sülü, G. Phytophthora chlamydospora and P.

megasperma associated with root and crown rot of sour cherry in Turkey. *J. Plant Dis. Prot.* **2017**, *124*, 403–406.

33. Ginetti, B.; Carmignani, S.; Ragazzi, A.; Moricca, S. Phytophthora Taxon Pgchlamydo is a Cause of Shoot Blight and Root and Collar Rot of Viburnum tinus in Italy. *Plant Dis.* **2014**, *98*, 1432.
34. Blomquist, C.L.; Yakabe, L.E.; Soriano, M.C.; Negrete, M.A. First Report of Leaf Spot Caused by Phytophthora taxon Pgchlamydo on Evergreen Nursery Stock in California. *Plant Dis.* **2012**, *96*, 1691.
35. Derviş, S.; Türkölmez, Ş.; Çiftçi, O.; Ulubaş Serçe, Ç. First Report of Phytophthora chlamydospora Causing Root Rot on Walnut (Juglans regia) Trees in Turkey. *Plant Dis.* **2016**, *100*, 2336.
36. Türkölmez, Ş.; Derviş, S.; Çiftçi, O.; Ulubaş Serçe, Ç. First Report of Phytophthora chlamydospora Causing Root and Crown Rot on Almond (Prunus dulcis) Trees in Turkey. *Plant Dis.* **2016**, *100*, 1796.
37. Latorre, B.A.; Rioja, M.E.; Wilcox, W.F. Phytophthora Species Associated with Crown and Root Rot of Apple in Chile. *Plant Dis.* **2001**, *85*, 603–606.
38. Belisario, A.; Luongo, L.; Vitale, S.; Galli, M.; Haegi, A. Phytophthora gonapodyides Causes Decline and Death of English (Persian) Walnut (Juglans regia) in Italy. *Plant Dis.* **2016**, *100*, 2537.
39. Reeser, P.W.; Sutton, W.; Hansen, E.M. Phytophthora Species Causing Tanoak Stem Cankers in Southwestern Oregon. *Plant Dis.* **2008**, *92*, 1252.
40. Corcobado, T.; Cubera, E.; Pérez-Sierra, A.; Jung, T.; Solla, A. First report of Phytophthora gonapodyides involved in the decline of Quercus ilex in xeric conditions in Spain. *New Dis. Reports* **2010**, *22*, 33.
41. O'Hanlon, R.; Choiseul, J.; Grogan, H.; Brennan, J.M. In-vitro characterisation of the four lineages of Phytophthora ramorum. *Eur. J. Plant Pathol.* **2016**, *2*.
42. O'Hanlon, R.; Choiseul, J.; Brennan, J.M.; Grogan, H. Assessment of the eradication measures applied to Phytophthora ramorum in Irish Larix kaempferi forests. *For. Pathol.* **2018**, *48*, e12389.
43. Jeffers, S.N. Comparison of Two Media Selective for Phytophthora and Pythium Species. *Plant Dis.* **1986**, *70*, 1038.
44. Werres, S.; Marwitz, R.; Man In't veld, W.A.; De Cock, A.W.A.M.; Bonants, P.J.M.; De Weerdt, M.; Themann, K.; Ilieva, E.; Baayen, R.P. Phytophthora ramorum sp. nov., a new pathogen on Rhododendron and Viburnum. *Mycol. Res.*

2001, *105*, 1155–1165.

45. Gallegly, M.E.; ChuanXue, H. *Phytophthora: identifying species by morphology and DNA fingerprints.*; American Phytopathological Society (APS Press): St. Paul, 2008; ISBN 9780890543641.
46. Erwin, D.C.; Ribeiro, O.K. *Phytophthora diseases worldwide.*; American Phytopathological Society (APS Press): St. Paul, Minnesota, 1996; ISBN 0890542120.
47. Altschul, S.F.; Madden, T.L.; Schäffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402.
48. Marçais, G.; Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **2011**, *27*, 764–770.
49. Vurture, G.W.; Sedlazeck, F.J.; Nattestad, M.; Underwood, C.J.; Fang, H.; Gurtowski, J.; Schatz, M.C. GenomeScope: Fast reference-free genome profiling from short reads. In Proceedings of the Bioinformatics; 2017; Vol. 33, pp. 2202–2204.
50. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A. a.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Prjibelski, A.D.; et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477.
51. Boetzer, M.; Henkel, C. V.; Jansen, H.J.; Butler, D.; Pirovano, W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **2011**, *27*, 578–579.
52. Boetzer, M.; Pirovano, W. Toward almost closed genomes with GapFiller. *Genome Biol.* **2012**, *13*, R56.
53. Mikheenko, A.; Prjibelski, A.; Saveliev, V.; Antipov, D.; Gurevich, A. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* **2018**, *34*, i142–i150.
54. Waterhouse, R.M.; Seppey, M.; Simão, F.A.; Manni, M.; Ioannidis, P.; Klioutchnikov, G.; Kriventseva, E. V; Zdobnov, E.M. BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Mol. Biol. Evol.* **2018**, *35*, 543–548.
55. Flynn, J.M.; Hubley, R.; Goubert, C.; Rosen, J.; Clark, A.G.; Feschotte, C.; Smit, A.F. RepeatModeler2: automated genomic discovery of transposable element families. *bioRxiv* **2019**.

56. Dierckxsens, N.; Mardulyn, P.; Smits, G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* **2016**, *45*, gkw955.
57. Greiner, S.; Lehwark, P.; Bock, R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* **2019**, *47*, W59–W64.
58. Hoff, K.J.; Lange, S.; Lomsadze, A.; Borodovsky, M.; Stanke, M. BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS: Table 1. *Bioinformatics* **2016**, *32*, 767–769.
59. Bruna, T.; Lomsadze, A.; Borodovsky, M. GeneMark-EP and -EP+: automatic eukaryotic gene prediction supported by spliced aligned proteins. *bioRxiv* **2020**, 2019.12.31.891218.
60. Ter-Hovhannisyan, V.; Lomsadze, A.; Chernoff, Y.O.; Borodovsky, M. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Res.* **2008**, *18*, 1979–1990.
61. Buchfink, B.; Xie, C.; Huson, D.H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **2015**, *12*, 59–60.
62. Iwata, H.; Gotoh, O. Benchmarking spliced alignment programs including Spaln2, an extended version of Spaln that incorporates additional species-specific features. *Nucleic Acids Res.* **2012**, *40*, e161–e161.
63. Hoff, K.J.; Stanke, M. Predicting Genes in Single Genomes with AUGUSTUS. *Curr. Protoc. Bioinforma.* **2018**, *65*, e57.
64. Jones, P.; Binns, D.; Chang, H.-Y.; Fraser, M.; Li, W.; McAnulla, C.; McWilliam, H.; Maslen, J.; Mitchell, A.; Nuka, G.; et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **2014**, *30*, 1236–1240.
65. Huerta-Cepas, J.; Forslund, K.; Coelho, L.P.; Szklarczyk, D.; Jensen, L.J.; von Mering, C.; Bork, P. Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Mol. Biol. Evol.* **2017**, *34*, 2115–2122.
66. Bendtsen, J.D.; Nielsen, H.; Von Heijne, G.; Brunak, S. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **2004**, *340*, 783–795.
67. Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E.L.. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **2001**, *305*, 567–80.
68. Sperschneider, J.; Williams, A.H.; Hane, J.K.; Singh, K.B.; Taylor, J.M. Evaluation of Secretion Prediction Highlights Differing Approaches Needed for Oomycete and

Fungal Effectors. *Front. Plant Sci.* **2015**, *6*, 1–14.

69. Raffaele, S.; Win, J.; Cano, L.M.; Kamoun, S. Analyses of genome architecture and gene expression reveal novel candidate virulence factors in the secretome of Phytophthora infestans. *BMC Genomics* **2010**, *11*, 637.
70. Sperschneider, J.; Dodds, P.N.; Singh, K.B.; Taylor, J.M. ApoplastP: prediction of effectors and plant proteins in the apoplast using machine learning. *New Phytol.* **2018**, *217*, 1764–1778.
71. Zhang, H.; Yohe, T.; Huang, L.; Entwistle, S.; Wu, P.; Yang, Z.; Busk, P.K.; Xu, Y.; Yin, Y. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **2018**, *46*, W95–W101.
72. Urban, M.; Cuzick, A.; Rutherford, K.; Irvine, A.; Pedro, H.; Pant, R.; Sadanadan, V.; Khamari, L.; Billal, S.; Mohanty, S.; et al. PHI-base: a new interface and further additions for the multi-species pathogen–host interactions database. *Nucleic Acids Res.* **2017**, *45*, D604–D610.
73. Enright, A.J.; Van Dongen, S.; Ouzounis, C.A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **2002**, *30*, 1575–84.
74. Klopfenstein, D. V.; Zhang, L.; Pedersen, B.S.; Ramírez, F.; Vesztrocy, A.W.; Naldi, A.; Mungall, C.J.; Yunes, J.M.; Botvinnik, O.; Weigel, M.; et al. GOATOOLS: A Python library for Gene Ontology analyses. *Sci. Rep.* **2018**, *8*, 1–17.
75. Win, J.; Morgan, W.; Bos, J.; Krasileva, K. V.; Cano, L.M.; Chaparro-garcia, A.; Ammar, R.; Staskawicz, B.J.; Kamoun, S. Adaptive Evolution Has Targeted the C-Terminal Domain of the RXLR Effectors of Plant Pathogenic Oomycetes © American Society of Plant Biologists Adaptive Evolution Has Targeted the C-Terminal Domain of the RXLR Effectors of Plant Pathogenic Oomycetes. *Plant Cell* **2007**, *19*, 2349–2369.
76. Eddy, S. Profile hidden Markov models. *Bioinformatics* **1998**, *14*, 755–763.
77. 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. A translocation signal for delivery of oomycete effector proteins into host plant cells. *Nature* **2007**, *450*, 115–118.
78. Boutemy, L.S.; King, S.R.F.F.; Win, J.; Hughes, R.K.; Clarke, T.A.; Blumenschein, T.M.A.A.; Kamoun, S.; Banfield, M.J. Structures of Phytophthora RXLR effector proteins: A conserved but adaptable fold underpins functional diversity. *J. Biol.*

Chem. **2011**, *286*, 35834–35842.

79. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797.
80. Capella-Gutierrez, S.; Silla-Martinez, J.M.; Gabaldon, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **2009**, *25*, 1972–1973.
81. Nguyen, L.T.; Schmidt, H.A.; Von Haeseler, A.; Minh, B.Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274.
82. Kalyaanamoorthy, S.; Minh, B.Q.; Wong, T.K.F.; von Haeseler, A.; Jermiin, L.S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **2017**, *14*, 587–589.
83. Lartillot, N.; Rodrigue, N.; Stubbs, D.; Richer, J. PhyloBayes MPI: Phylogenetic Reconstruction with Infinite Mixtures of Profiles in a Parallel Environment. *Syst. Biol.* **2013**, *62*, 611–615.
84. Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* **2019**, *47*, W256–W259.
85. Drost, H.-G.; Gabel, A.; Grosse, I.; Quint, M. Evidence for Active Maintenance of Phylotranscriptomic Hourglass Patterns in Animal and Plant Embryogenesis. *Mol. Biol. Evol.* **2015**, *32*, 1221–1231.
86. Quint, M.; Drost, H.-G.; Gabel, A.; Ullrich, K.K.; Bönn, M.; Grosse, I. A transcriptomic hourglass in plant embryogenesis. *Nature* **2012**, *490*, 98–101.
87. Morrin, S.T.; Owens, R.A.; Le Berre, M.; Gerlach, J.Q.; Joshi, L.; Bode, L.; Irwin, J.A.; Hickey, R.M. Interrogation of Milk-Driven Changes to the Proteome of Intestinal Epithelial Cells by Integrated Proteomics and Glycomics. *J. Agric. Food Chem.* **2019**, *67*, 1902–1917.
88. Owens, R.A.; O'Keeffe, G.; Smith, E.B.; Dolan, S.K.; Hammel, S.; Sheridan, K.J.; Fitzpatrick, D.A.; Keane, T.M.; Jones, G.W.; Doyle, S. Interplay between Gliotoxin Resistance, Secretion, and the Methyl/Methionine Cycle in Aspergillus fumigatus. *Eukaryot. Cell* **2015**, *14*, 941–957.
89. Cox, J.; Neuhauser, N.; Michalski, A.; Scheltema, R.A.; Olsen, J. V.; Mann, M. Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. *J. Proteome Res.* **2011**, *10*, 1794–1805.
90. Tyanova, S.; Temu, T.; Cox, J. The MaxQuant computational platform for mass

spectrometry-based shotgun proteomics. *Nat. Protoc.* **2016**, *11*, 2301–2319.

91. Delgado, J.; Núñez, F.; Asensio, M.A.; Owens, R.A. Quantitative proteomic profiling of ochratoxin A repression in Penicillium nordicum by protective cultures. *Int. J. Food Microbiol.* **2019**, *305*, 108243.
92. Tyanova, S.; Temu, T.; Sinitcyn, P.; Carlson, A.; Hein, M.Y.; Geiger, T.; Mann, M.; Cox, J. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **2016**, *13*, 731–740.
93. Fletcher, K.; Gil, J.; Bertier, L.D.; Kenefick, A.; Wood, K.J.; Zhang, L.; Reyes-Chin-Wo, S.; Cavanaugh, K.; Tsuchida, C.; Wong, J.; et al. Genomic signatures of heterokaryosis in the oomycete pathogen Bremia lactucae. *Nat. Commun.* **2019**, *10*, 1–13.
94. Fang, Y. "Francis"; Coelho, M.A.; Shu, H.; Schotanus, K.; Thimmappa, B.C.; Yadav, V.; Chen, H.; Malc, E.P.; Wang, J.; Mieczkowski, P.A.; et al. Long transposon-rich centromeres in an oomycete reveal divergence of centromere features in Stramenopila-Alveolata-Rhizaria lineages. *PLoS Genet*. **2020**, *16*, e1008646.
95. Malar C, M.; Yuzon, J.D.; Panda, A.; Kasuga, T.; Tripathy, S. Updated Assembly of Phytophthora ramorum pr102 Isolate Incorporating Long Reads from PacBio Sequencing. *Mol. Plant-Microbe Interact.* **2019**, *32*, 1472–1474.
96. McCarthy, C.G.P.; Fitzpatrick, D.A. Phylogenomic Reconstruction of the Oomycete Phylogeny Derived from 37 Genomes. *mSphere* **2017**, *2*, e00095-17.
97. Blair, J.E.; Coffey, M.D.; Park, S.Y.; Geiser, D.M.; Kang, S. A multi-locus phylogeny for Phytophthora utilizing markers derived from complete genome sequences. *Fungal Genet. Biol.* **2008**, *45*, 266–277.
98. Martin, F.N.; Blair, J.E.; Coffey, M.D. A combined mitochondrial and nuclear multilocus phylogeny of the genus Phytophthora. *Fungal Genet. Biol.* **2014**, *66*, 19–32.
99. Yang, X.; Hong, C. Differential Usefulness of Nine Commonly Used Genetic Markers for Identifying Phytophthora Species. *Front. Microbiol.* **2018**, *9*, 2334.
100. Fitzpatrick, D.A.; Logue, M.E.; Stajich, J.E.; Butler, G. A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis. *BMC Evol. Biol.* **2006**, *6*, 99.
101. Fletcher, K.; Klosterman, S.J.; Derevnina, L.; Martin, F.; Bertier, L.D.; Koike, S.; Reyes-Chin-Wo, S.; Mou, B.; Michelmore, R. Comparative genomics of downy

mildews reveals potential adaptations to biotrophy. *BMC Genomics* **2018**, *19*, 8–10.

102. Karlovsky, P.; Fartmann, B. Genetic code and phylogenetic origin of oomycetous mitochondria. *J. Mol. Evol.* **1992**, *34*, 254–258.

103. Martin, F.N.; Bensasson, D.; Tyler, B.M.; Boore, J.L. Mitochondrial genome sequences and comparative genomics of Phytophthora ramorum and P. sojae. *Curr. Genet.* **2007**, *51*, 285–296.

104. Derevnina, L.; Dagdas, Y.F.; De la Concepcion, J.C.; Bialas, A.; Kellner, R.; Petre, B.; Domazakis, E.; Du, J.; Wu, C.; Lin, X.; et al. Nine things to know about elicitins. *New Phytol.* **2016**, *212*, 888–895.

105. Jiang, R.H.Y.; Tyler, B.M.; Whisson, S.C.; Hardham, A.R.; Govers, F. Ancient Origin of Elicitin Gene Clusters in Phytophthora Genomes. *Mol. Biol. Evol.* **2006**, *23*, 338–351.

106. Seidl, M.F.; Van den Ackerveken, G. Activity and Phylogenetics of the Broadly Occurring Family of Microbial Nep1-Like Proteins. *Annu. Rev. Phytopathol.* **2019**, *57*, 367–386.

107. Orsomando, G.; Lorenzi, M.; Raffaelli, N.; Dalla Rizza, M.; Mezzetti, B.; Ruggieri, S. Phytotoxic Protein PcF, Purification, Characterization, and cDNA Sequencing of a Novel Hydroxyproline-containing Factor Secreted by the Strawberry Pathogen Phytophthora cactorum. *J. Biol. Chem.* **2001**, *276*, 21578–21584.

108. Raaymakers, T.M.; Van den Ackerveken, G. Extracellular Recognition of Oomycetes during Biotrophic Infection of Plants. *Front. Plant Sci.* **2016**, *7*, 1–12.

109. Brunner, F.; Rosahl, S.; Lee, J.; Rudd, J.J.; Geiler, C.; Kauppinen, S.; Rasmussen, G.; Scheel, D.; Nürnberger, T. Pep-13, a plant defense-inducing pathogen-associated pattern from Phytophthora transglutaminases. *EMBO J.* **2002**, *21*, 6681–8.

110. Gaulin, E.; Jauneau, A.; Villalba, F.; Rickauer, M.; Esquerré-Tugayé, M.-T.; Bottin, A. The CBEL glycoprotein of Phytophthora parasitica var-nicotianae is involved in cell wall deposition and adhesion to cellulosic substrates. *J. Cell Sci.* **2002**, *115*, 4565–75.

111. Schneiter, R.; Di Pietro, A. The CAP protein superfamily: function in sterol export and fungal virulence. *Biomol. Concepts* **2013**, *4*, 519–525.

112. Fries, M.; Ihrig, J.; Brocklehurst, K.; Shevchik, V.E.; Pickersgill, R.W. Molecular basis of the activity of the phytopathogen pectin methylesterase. *EMBO*

J. **2007**, *26*, 3879–3887.

113. Vercauteren, I.; de Almeida Engler, J.; De Groodt, R.; Gheysen, G. An Arabidopsis thaliana Pectin Acetylesterase Gene Is Upregulated in Nematode Feeding Sites Induced by Root-knot and Cyst Nematodes. *Mol. Plant-Microbe Interact.* **2002**, *15*, 404–407.

114. Win, J.; Krasileva, K. V.; Kamoun, S.; Shirasu, K.; Staskawicz, B.J.; Banfield, M.J. Sequence Divergent RXLR Effectors Share a Structural Fold Conserved across Plant Pathogenic Oomycete Species. *PLoS Pathog.* **2012**, *8*, e1002400.

115. Yin, J.; Gu, B.; Huang, G.; Tian, Y.; Quan, J.; Lindqvist-Kreuze, H.; Shan, W. Conserved RXLR Effector Genes of Phytophthora infestans Expressed at the Early Stage of Potato Infection Are Suppressive to Host Defense. *Front. Plant Sci.* **2017**, *8*, 1–11.

116. Wang, S.; McLellan, H.; Bukharova, T.; He, Q.; Murphy, F.; Shi, J.; Sun, S.; van Weymers, P.; Ren, Y.; Thilliez, G.; et al. Phytophthora infestans RXLR effectors act in concert at diverse subcellular locations to enhance host colonization. *J. Exp. Bot.* **2019**, *70*, 343–356.

117. Schornack, S.; van Damme, M.; Bozkurt, T.O.; Cano, L.M.; Smoker, M.; Thines, M.; Gaulin, E.; Kamoun, S.; Huitema, E. Ancient class of translocated oomycete effectors targets the host nucleus. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 17421–17426.

118. Stam, R.; Jupe, J.; Howden, A.J.M.; Morris, J.A.; Boevink, P.C.; Hedley, P.E.; Huitema, E. Identification and Characterisation CRN Effectors in Phytophthora capsici Shows Modularity and Functional Diversity. *PLoS One* **2013**, *8*, 1–13.

119. Zerillo, M.M.; Adhikari, B.N.; Hamilton, J.P.; Buell, C.R.; Lévesque, C.A.; Tisserat, N. Carbohydrate-Active Enzymes in Pythium and Their Role in Plant Cell Wall and Storage Polysaccharide Degradation. *PLoS One* **2013**, *8*, e72572.

120. Sharma, R.; Xia, X.; Cano, L.M.; Evangelisti, E.; Kemen, E.; Judelson, H.; Oome, S.; Sambles, C.; van den Hoogen, D.J.; Kitner, M.; et al. Genome analyses of the sunflower pathogen Plasmopara halstedii provide insights into effector evolution in downy mildews and Phytophthora. *BMC Genomics* **2015**, *16*, 741.

121. Benhamou, N.; le Floch, G.; Vallance, J.; Gerbore, J.; Grizard, D.; Rey, P. Pythium oligandrum: An example of opportunistic success. *Microbiol. (United Kingdom)* **2012**, *158*, 2679–2694.

122. Rodenburg, S.Y.A.; de Ridder, D.; Govers, F.; Seidl, M.F. Oomycete

metabolism is highly dynamic and reflects lifestyle adaptations. *bioRxiv* **2020**, 2020.02.12.941195.

123. Thines, M.; Sharma, R.; Rodenburg, S.Y.A.; Gogleva, A.; Judelson, H.S.; Xia, X.; van den Hoogen, J.; Kitner, M.; Klein, J.; Neilen, M.; et al. The genome of Peronospora belbahrii reveals high heterozygosity, a low number of canonical effectors and CT-rich promoters. *bioRxiv* **2019**, 721027.

124. McGowan, J.; Byrne, K.P.; Fitzpatrick, D.A. Comparative Analysis of Oomycete Genome Evolution Using the Oomycete Gene Order Browser (OGOB). *Genome Biol. Evol.* **2019**, *11*, 189–206.

125. Bendtsen, J.D.; Jensen, L.J.; Blom, N.; von Heijne, G.; Brunak, S. Feature-based prediction of non-classical and leaderless protein secretion. *Protein Eng. Des. Sel.* **2004**, *17*, 349–356.

126. Pennington, H.G.; Jones, R.; Kwon, S.; Bonciani, G.; Thieron, H.; Chandler, T.; Luong, P.; Morgan, S.N.; Przydacz, M.; Bozkurt, T.; et al. The fungal ribonuclease-like effector protein CSEP0064/BEC1054 represses plant immunity and interferes with degradation of host ribosomal RNA. *PLOS Pathog.* **2019**, *15*, e1007620.

127. Wang, S.; Boevink, P.C.; Welsh, L.; Zhang, R.; Whisson, S.C.; Birch, P.R.J. Delivery of cytoplasmic and apoplastic effectors from Phytophthora infestans haustoria by distinct secretion pathways. *New Phytol.* **2017**, *216*, 205–215.

128. Wang, S.; Welsh, L.; Thorpe, P.; Whisson, S.C.; Boevink, P.C.; Birch, P.R.J. The Phytophthora infestans Haustorium Is a Site for Secretion of Diverse Classes of Infection-Associated Proteins. *MBio* **2018**, *9*, 1–14.

129. Munro, S.; Pelham, H.R.B. A C-terminal signal prevents secretion of luminal ER proteins. *Cell* **1987**, *48*, 899–907.

130. Papp, S.; Opas, M. Sub-Cellular Distribution of Calreticulin. In *Calreticulin: Second Edition*; Eggleton, P., Michalak, M., Eds.; Springer US: Boston, MA, 2003; pp. 38–48 ISBN 978-1-4419-9258-1.

Chapter 5

General Discussion

Overview

Oomycetes are osmotrophic microorganisms that commonly inhabit terrestrial and aquatic habitats worldwide. Oomycetes are highly diverse in terms of their lifestyles, pathogenicity and host ranges. Some of the most destructive pathogens of plants and animals belong to the oomycete class. It is estimated that global food production must increase by 70% to feed the growing global population (M. Carvajal-Yepes et al. 2020). However, global food supply is threatened by plant pathogens and pests which are estimated to cause up to 30% crop yield losses annually (Savary et al. 2019). Phytopathogenic oomycetes represent some of the biggest threats to global food security. Examples of infamous oomycete species include *Phytophthora infestans*, the causative agent of late potato blight and cause of the Irish potato famine (Haas et al. 2009). *Phytophthora sojae* is another highly destructive species that causes up to $2 billion worth of soybean crop loss each year (Tyler 2007). Oomycetes also threaten forestry and natural ecosystems. For example, *Phytophthora ramorum*, the sudden oak death pathogen, has decimated forests across North America and Europe (Grünwald et al. 2019). There is an increasing risk of invasive pathogens spreading to nurseries, forests and natural ecosystems worldwide due to the increased global movement of people and trade of plants (e.g. nursery plant trade) (Brasier 2008; Jung et al. 2016; Hulbert et al. 2017).

Sequencing the genomes of oomycete species has significantly advanced our understanding of oomycete biology and evolution, and revealed insights into the mechanisms of oomycete-host interactions. The results presented in this thesis use comparative and evolutionary genomics methods to investigate genomic evolution across the oomycete class, with a particular focus on oomycete secretomes and effector arsenals. The main themes of research conducted in this thesis are summarised in **Figure 5.1**.

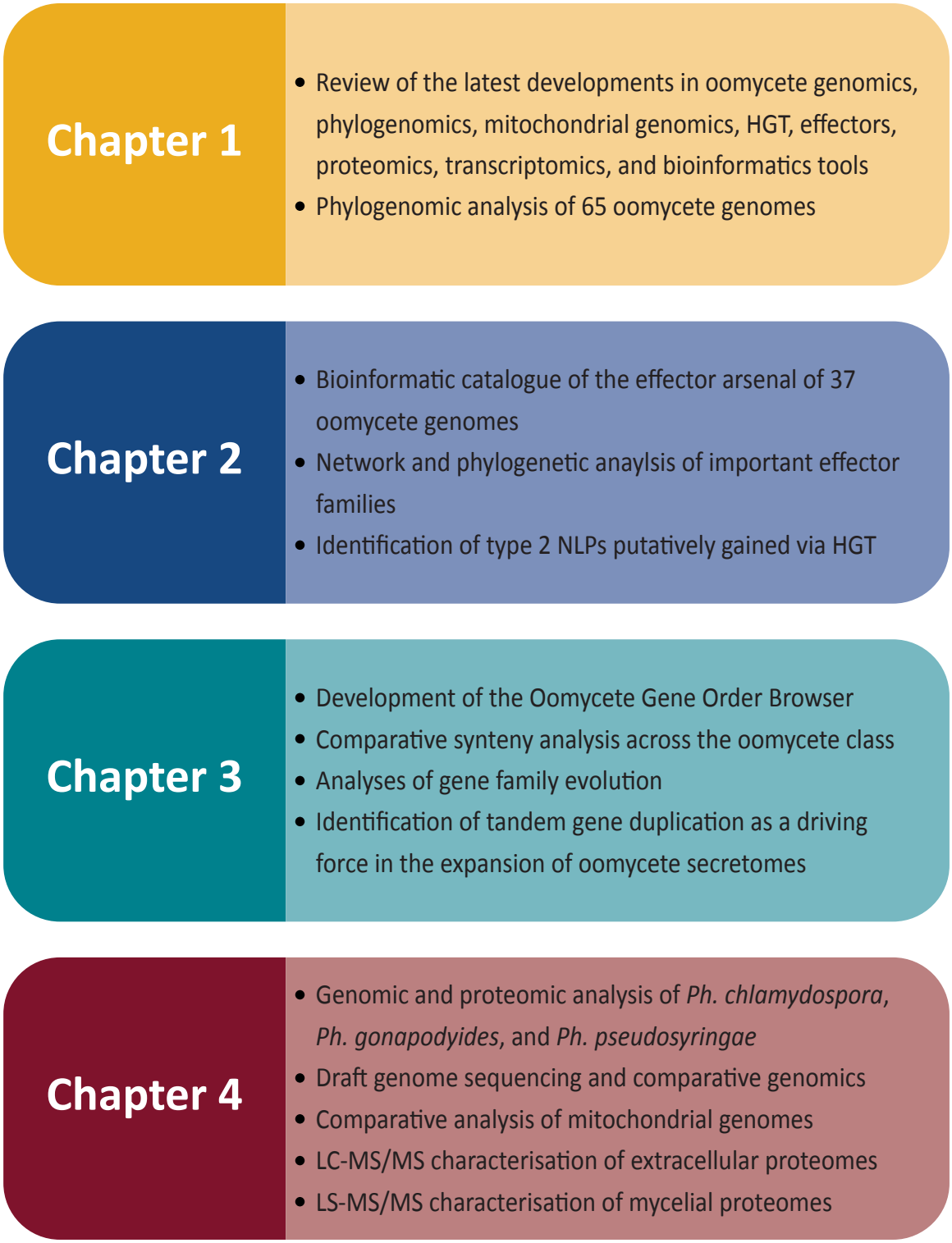


Figure 5.1. Summary of the research conducted in each chapter.

This thesis begins with a review of the latest developments in oomycete genomics (**Chapter 1)**(McGowan and Fitzpatrick 2020). This introductory chapter discusses some of the important findings that were facilitated by oomycete genome sequencing projects since the first oomycete genome sequences (*Ph. ramorum* and *Ph. sojae*) were published in 2006 (Tyler et al. 2006). This chapter also presents an updated phylogeny of the oomycete class using a phylogenomic approach based on a supermatrix alignment of 102 BUSCO proteins from the genomes available at the time of writing (McGowan and Fitzpatrick 2020).

The availability of a large number of oomycete genome sequences facilitated bioinformatic characterisation of the secretomes and effector arsenals of 37 species (**Chapter 2**)(McGowan and Fitzpatrick 2017). Large numbers of known effector families were significantly enriched across the secretomes, for example, elicitins, necrosis-inducing proteins, RxLRs and plant cell wall degrading enzymes. Comparative genomic analyses identified species-specific and lineage-specific expansions of effector families. The repertoire of cytoplasmic RxLR effectors in each oomycete was also catalogued leading to the identification of 4,131 putative RxLRs across the 37 genomes. A homology network was constructed of putative RxLRs revealing numerous disconnected clusters, indicating that some RxLR families do not share significant sequence similarity. This suggests that some RxLR families may not be related. A number of secreted proteins were putatively annotated as IgA peptidases, suggesting that they may represent a novel effector family. Network and phylogenetic analyses identified a putative horizontal gene transfer event involving a type 2 necrosis-inducing protein. This chapter represents an up-to-date *in silico* catalogue of the effector arsenal of diverse oomycete species (McGowan and Fitzpatrick 2017).

Compared to other taxonomic groups, such as fungi, there has been a dearth of dedicated tools available to study oomycete genomes. In an attempt to negate this somewhat, the Oomycete Gene Order Browser (OGOB) was developed (**Chapter 3**)(McGowan et al. 2019). OGOB is a novel database and tool that facilitates comparative genomic and syntenic analyses of oomycete genomes. The version described herein currently hosts genomic data for 20 diverse oomycete species and is available at https://ogob.ie. OGOB also hosts a number of bioinformatics tools that facilitates common analyses such as performing multiple sequence alignments, estimating the rate of evolution of genes and performing phylogenetic reconstruction of gene families. Comparative syntenic analyses using OGOB highlighted the high degree of syntenic conservation within oomycete genera. On average, when comparing any pair of the 20 species, 34% of total genes or 65% of orthologs were microsyntenically conserved. As expected, syntenically conserved genes were enriched for housekeeping functions. Analysis of gene duplications within individual species revealed that up to 15% of all genes are located in tandem clusters. Proteins predicted to be secreted were significantly enriched in tandem clusters, with up to 35% of tandemly duplicated genes bearing putative signal peptides. This suggests that tandem gene duplication has played a major role in the expansion and evolution of oomycete secretomes (McGowan et al. 2019).

Draft genome assemblies were generated for three widespread *Phytophthora* species - *Ph. chlamydospora*, *Ph. gonapodyides* and *Ph. pseudosyringae* (**Chapter 4**)(McGowan et al. 2020). *Ph. gonapodyides* and *Ph. chlamydospora* are thought to be the two most globally widespread *Phytophthora* species. They are both important opportunistic pathogens that can cause significant damage when spread from natural ecosystems to managed forests. *Ph. pseudosyringae* is an important forest pathogen that is abundant in Europe and North America. The genomes were sequenced to high completeness (87.2% to 97.8% BUSCO completeness) with genome sizes ranging from

45 Mb to 61 Mb. The mitochondrial genome of each species was also assembled and circularised revealing genome sizes ranging from 38,329 bp to 43,974 bp. Mitochondrial gene content and gene order were highly conserved across the three species, except for two inversions present in *Ph. pseudosyringae*. Comparative genomic analyses of 44 oomycete genomes suggest that *Ph. chlamydospora* and *Ph. gonapodyides* have distinctive CAZymes arsenals relative to other *Phytophthora* species and that oomycete lifestyles may be linked to their CAZyme repertoires, in particular, their glycoside hydrolase repertoires. Gene prediction led to the annotation of 17,439 to 23,348 protein-coding genes per species. Availability of high-quality gene sets facilitated proteomics experiments to further interrogate the three *Phytophthora* species. Previously, mass spectrometric analyses of oomycete extracellular proteomes were limited to only two *Phytophthora* species – *Ph. infestans* and *Ph. plurivora* (Meijer et al. 2014; Severino et al. 2014). This work extends this number to five by performing LC-MS/MS analysis of the extracellular proteomes of *Ph. chlamydospora*, *Ph. gonapodyides* and *Ph. pseudosyringae*. This led to the identification of approximately 300 extracellular proteins per species. The majority of identified extracellular proteins were putatively involved in osmotrophy or infection and included several known effector families, such as necrosis-inducing proteins, elicitins, transglutaminase elicitors and plant cell wall degrading enzymes. Furthermore, LC-MS/MS analysis identified differences in the mycelial proteome of each species, with 84 to 353 mycelial proteins being uniquely detected in one species. In total, the expression of approximately 3,000 genes per species was validated at the protein level (McGowan et al. 2020).

Overall the body of work presented in this thesis provides new insights into the evolution of oomycete species. In particular, the data presented herein significantly expands our knowledge of the biology of *Ph. chlamydospora*, *Ph. gonapodyides* and *Ph. pseudosyringae*.

Future Work

In line with continuous advances in sequencing technologies and drastic decreases in the cost of whole-genome sequencing, there has been an increased pace of oomycete genome sequencing (**Figure 5.2**). At the time of writing (April 2020), there are 215 publicly available oomycete genome assemblies deposited in NCBI GenBank (Benson et al. 2012), corresponding to 75 species (**Figure 5.2**). The increasing number of high-quality genome assemblies, in particular those that have been assembled to high contiguity with long-read technologies, represent new genomes that could be added to future versions of OGOB. Furthermore, as more high-quality genome sequences become available for multiple representatives of each *Phytophthora* clade/subclade, future phylogenomic analyses may be performed to further resolve the contentious relationships between *Phytophthora* clades.

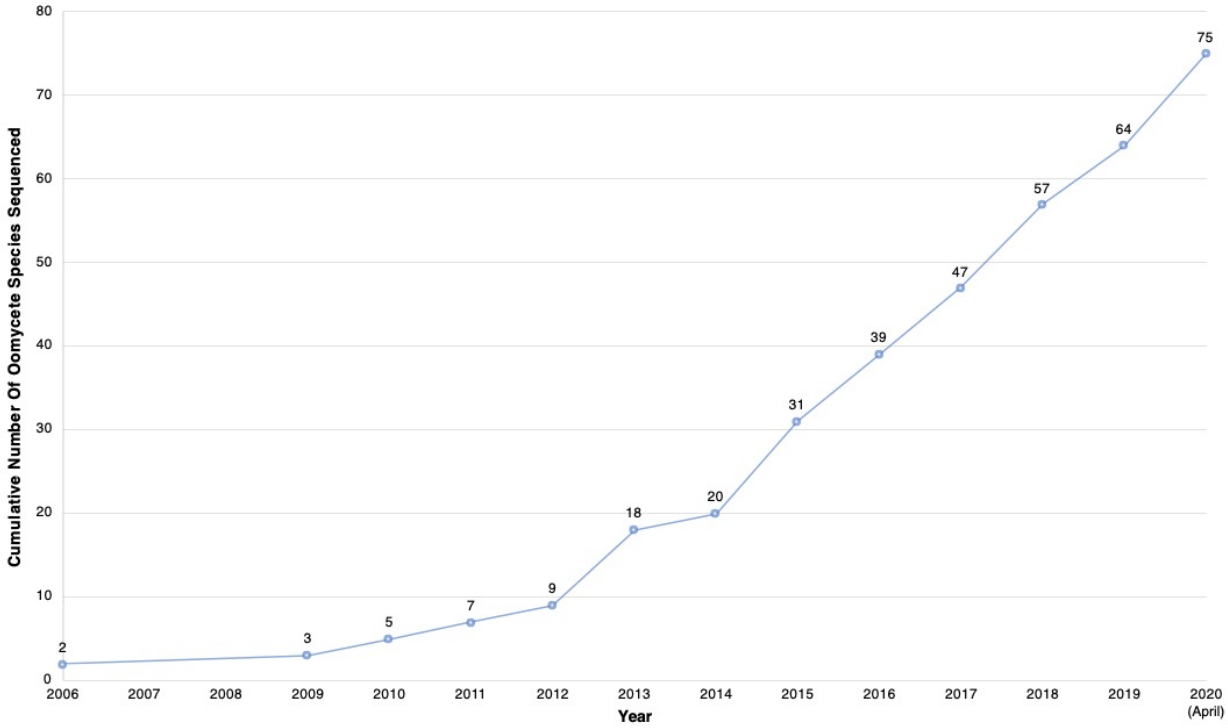


Figure 5.2. Total number of oomycete genomes sequenced up to April 2020. Data is based on either year of publication or year deposited on NCBI GenBank.

The ever-increasing number of oomycete genomes will require the development of dedicated tools to fully exploit the genomic resources. As discussed in **Chapter 1** (McGowan and Fitzpatrick 2020), BUSCO (Waterhouse et al. 2018) is one of the most commonly used tools to assess genome completeness. In BUSCO v3 the most specialised BUSCO dataset available for oomycete genomes is the "Alveolata-Stramenopiles" dataset, which contains only 234 target BUSCO proteins. The low number of target proteins may result in an inaccurate assessment of genome completeness. BUSCO v4 was recently released and includes a new "Stramenopiles" dataset, which incorporates orthologs from 27 stramenopiles, including 15 oomycete species. However, this dataset includes even fewer target BUSCOs with only 100 target proteins, due to the inclusion of diverse stramenopiles in the dataset, such as brown algae and diatoms. With a large number of oomycete genome sequences available, it would be more appropriate to have dedicated datasets for the oomycetes or for individual oomycete genera, for example using ubiquitous single-copy orthologs identified by OGOB. This would allow for more accurate assessment of oomycete genome assembly completeness.

LC-MS/MS analysis of the extracellular proteomes of *Ph. chlamydospora*, *Ph. gonapodyides* and *Ph. pseudosyringae* identified a total of 965 extracellular proteins. Only 551 (57%) of these proteins were predicted to contain a signal peptide. Future bioinformatics analyses of the extracellular proteins lacking signal peptides may reveal divergent signal peptides or novel motifs involved in non-canonical secretory pathways. SignalP is the most commonly used bioinformatics tool to predict protein secretion and is often the first step performed to identify putative effector proteins. The latest version, SignalP v5, uses a deep neural network-based approach to identify signal peptides, trained with a dataset of 20,758 protein sequences (Almagro Armenteros et al. 2019). The majority of proteins in the training dataset come from model organisms, such as *Homo sapiens* (4,258), *Arabidopsis thaliana* (2,491), *Saccharomyces cerevisiae* (2,061),

Schizosaccharomyces pombe (1,888), *Mus musculus* (1,841) and *Drosophila melanogaster* (540). Only one protein in the training data comes from the oomycetes. The inclusion of additional training data from oomycete species in future tools, for example, the extracellular proteins identified by mass spectrometry in this thesis, might improve the performance of oomycete signal peptide prediction. Similarly, with the growing number of experimentally verified oomycete effector proteins and databases such as PHI-Base (Urban et al. 2017), machine learning classifiers could be trained to predict oomycete effectors, using methods similar to EffectorP, which was developed to predict fungal effector proteins (Sperschneider et al. 2018).

Several oomycete species have had multiple strains sequenced, including *Ph. fragariae*, *Ph. kernoviae*, *Ph. lateralis*, *Ph. ramorum* and *Ph. rubi* (Quinn et al. 2013; Studholme et al. 2015; Adams et al. 2019; Dale et al. 2019; Studholme et al. 2019). Sequencing multiple isolates of a species facilitates population genetics studies to understand the population structure and diversity of different strains. Furthermore, it facilitates pan-genomic studies (Tettelin et al. 2005) which can reveal insights into strain-level variation of gene content which may be linked to differences in infection aggressiveness, differences in host range and fungicide resistance. Availability of genome sequences for *Ph. chlamydospora*, *Ph. gonapodyides* and *Ph. pseudosyringae* facilitates future resequencing studies to understand the population structure of these species.

The two clade 6 species sequenced in this thesis, *Ph. chlamydospora* and *Ph. gonapodyides*, are generally regarded as opportunistic pathogens that primarily act as saprophytes in riparian ecosystems, where their populations survive by colonising plant debris (Hansen et al. 2012). The newly sequenced genomes will facilitate future studies to determine if these opportunistic pathogens behave the same as other "typical" *Phytophthora* species during infection. This could be achieved by performing comparative proteomics or transcriptomics analyses of these species with another

Phytophthora species, such as *Ph. pseudosyringae*, during infection of a common host, for example, *Fagus sylvatica* or apple fruit.

The majority of oomycetes genomics research conducted to date has focused on plant pathogens. For example, 62 (83%) of the sequenced oomycete species are phytopathogens. This is not surprising, given the threat oomycete plant pathogens pose to global food security. In comparison to phytopathogens, there is a lack of genomic data for non-pathogenic oomycetes and little is known about saprophytic oomycetes that play key ecological roles in natural environments. Furthermore, there is a dearth of genomic data for more basal oomycete species. Future sequencing of basal oomycete species will potentially help to further elucidate the origin and evolution of pathogenicity in the oomycetes.

Concluding Remarks

To conclude, the results presented here help elucidate some of the mechanisms of genomic evolution across the oomycete class. In particular, this thesis focuses on the genomic and proteomic characterisation of oomycete secretomes and effector arsenals to better understand the evolution of pathogenicity in economically important members of the class Oomycota. The development of the Oomycete Gene Order Browser, a novel database and tool for analysing oomycete genes and genomes, is also described. Furthermore, this thesis presents the first large-scale genomic and proteomic investigation of three ubiquitous *Phytophthora* species, *Ph. chlamydospora*, *Ph. gonapodyides* and *Ph. pseudosyringae*. The data presented herein will be invaluable for future studies to further understand the biology of these ubiquitous microorganisms.

References

Adams TM, Armitage AD, Sobczyk MK, Bates HJ, Tabima JF, Kronmiller BA, Tyler BM, Grünwald NJ, Dunwell JM, Nellist CF, et al. 2019. Genomic investigation of the strawberry pathogen Phytophthora fragariae indicates pathogenicity is determined by transcriptional variation in three key races. bioRxiv:860619.

Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat. Biotechnol. 37:420–423.

Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2012. GenBank. Nucleic Acids Res. 41:D36–D42.

Brasier CM. 2008. The biosecurity threat to the UK and global environment from international trade in plants. Plant Pathol. 57:792–808.

Dale AL, Feau N, Everhart SE, Dhillon B, Wong B, Sheppard J, Bilodeau GJ, Brar A, Tabima JF, Shen D, et al. 2019. Mitotic recombination and rapid genome evolution in the invasive forest pathogen phytophthora ramorum. MBio 10:1–19.

Grünwald NJ, LeBoldus JM, Hamelin RC. 2019. Ecology and Evolution of the Sudden Oak Death Pathogen Phytophthora ramorum. Annu. Rev. Phytopathol. 57:1–21.

Haas BJ, Kamoun S, Zody MC, Jiang RHY, Handsaker RE, Cano LM, Grabherr M, Kodira CD, Raffaele S, Torto-Alalibo T, et al. 2009. Genome sequence and analysis of the Irish potato famine pathogen Phytophthora infestans. Nature 461:393–398.

Hansen EM, Reeser PW, Sutton W. 2012. Phytophthora Beyond Agriculture. Annu. Rev. Phytopathol. 50:359–378.

Hulbert JM, Agne MC, Burgess TI, Roets F, Wingfield MJ. 2017. Urban environments provide opportunities for early detections of Phytophthora invasions. Biol. Invasions 19:3629–3644.

Jung T, Orlikowski L, Henricot B, Abad-Campos P, Aday AG, Aguín Casal O, Bakonyi J, Cacciola SO, Cech T, Chavarriaga D, et al. 2016. Widespread Phytophthora infestations in European nurseries put forest, semi-natural and horticultural ecosystems at high risk of Phytophthora diseases. For. Pathol. 46:134–163.

M. Carvajal-Yepes KC, A. Nelson, K. A. Garrett BG, D. G. O. Saunders, S. Kamoun JPL, V. Verdier, J. Lessel RAN, R. Day, P. Pardey MLG, A. R. Records, B. Bextine JEL, S. Staiger JT. 2020. A global surveillance system for crop diseases.

Science (80-.). 364:1237–1240.

McGowan J, Byrne KP, Fitzpatrick DA. 2019. Comparative Analysis of Oomycete Genome Evolution Using the Oomycete Gene Order Browser (OGOB).Baldauf S, editor. Genome Biol. Evol. 11:189–206.

McGowan J, Fitzpatrick DA. 2017. Genomic, Network, and Phylogenetic Analysis of the Oomycete Effector Arsenal. mSphere 2:e00408-17.

McGowan J, Fitzpatrick DA. 2020. Recent advances in oomycete genomics. In: Advances in Genetics. Academic Press.

McGowan J, O'Hanlon R, Owens RA, Fitzpatrick DA. 2020. Comparative Genomic and Proteomic Analyses of Three Widespread Phytophthora Species: Phytophthora chlamydospora, Phytophthora gonapodyides and Phytophthora pseudosyringae. Microorganisms 8:653.

Meijer HJG, Mancuso FM, Espadas G, Seidl MF, Chiva C, Govers F, Sabido E. 2014. Profiling the Secretome and Extracellular Proteome of the Potato Late Blight Pathogen Phytophthora infestans. Mol. Cell. Proteomics 13:2101–2113.

Quinn L, O'Neill PA, Harrison J, Paskiewicz KH, Mccracken AR, Cooke LR, Grant MR, Studholme DJ. 2013. Genome-wide sequencing of Phytophthora lateralis reveals genetic variation among isolates from Lawson cypress (Chamaecyparis lawsoniana) in Northern Ireland. FEMS Microbiol. Lett. 344:179–185.

Savary S, Willocquet L, Pethybridge SJ, Esker P, McRoberts N, Nelson A. 2019. The global burden of pathogens and pests on major food crops. Nat. Ecol. Evol. 3:430–439.

Severino V, Farina A, Fleischmann F, Dalio RJD, Di Maro A, Scognamiglio M, Fiorentino A, Parente A, Osswald W, Chambery A. 2014. Molecular profiling of the Phytophthora plurivora secretome: a step towards understanding the cross-talk between plant pathogenic oomycetes and their hosts. PLoS One 9:e112317.

Sperschneider J, Dodds PN, Gardiner DM, Singh KB, Taylor JM. 2018. Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. Mol. Plant Pathol. 19:2094–2110.

Studholme DJ, McDougal RL, Sambles C, Hansen E, Hardy G, Grant M, Ganley RJ, Williams NM. 2015. Genome sequences of six Phytophthora species associated with forests in New Zealand. Genomics Data 7:54–56.

Studholme DJ, Panda P, Sanfuentes Von Stowasser E, González M, Hill R, Sambles C, Grant M, Williams NM, McDougal RL. 2019. Genome sequencing of oomycete

isolates from Chile supports the New Zealand origin of Phytophthora kernoviae and makes available the first Nothophytophthora sp. genome. Mol. Plant Pathol. 20:423–431.

Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli S V., Crabtree J, Jones AL, Durkin AS, et al. 2005. Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: Implications for the microbial "pan-genome." Proc. Natl. Acad. Sci. 102:13950–13955.

Tyler BM. 2007. Phytophthora sojae: root rot pathogen of soybean and model oomycete. Mol. Plant Pathol. 8:1–8.

Tyler BM, Tripathy S, Zhang X, Dehal P, Jiang RHY, Aerts A, Arredondo FD, Baxter L, Bensasson D, Beynon JL, et al. 2006. Phytophthora genome sequences uncover evolutionary origins and mechanisms of pathogenesis. Science 313:1261–1266.

Urban M, Cuzick A, Rutherford K, Irvine A, Pedro H, Pant R, Sadanadan V, Khamari L, Billal S, Mohanty S, et al. 2017. PHI-base: a new interface and further additions for the multi-species pathogen–host interactions database. Nucleic Acids Res. 45:D604–D610.

Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva E V, Zdobnov EM. 2018. BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. Mol. Biol. Evol. 35:543–548.