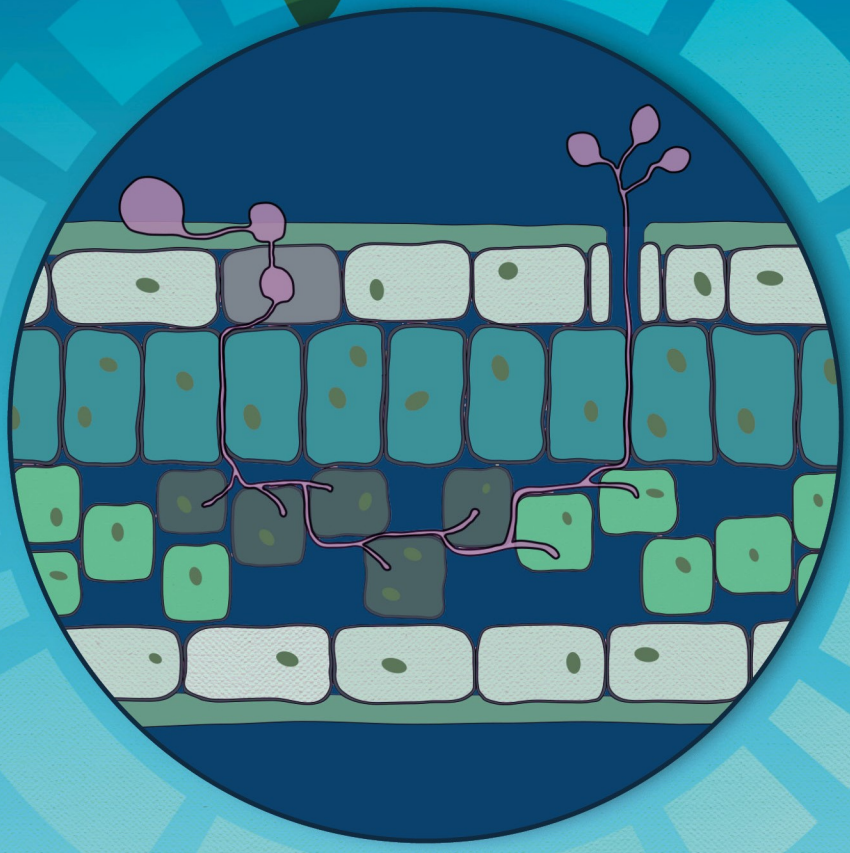
Uncovering oomycete metabolism using systems biology

Sander Y.A. Rodenburg



Propositions

1. Oomycete metabolism is underexplored.
   (this thesis)

2. A systems biology approach is needed to accelerate innovation in plant pathogen control.
   (this thesis)

3. Systems biology would benefit from the search-engine indexation of pre-internet literature.

4. Scientific journals should be more strict on the FAIRness of the supporting data.

5. 2020 is not the international year of plant health.

6. A truly sustainable society cannot be achieved without limiting population size.

7. Color blindness is an under-acknowledged handicap.

Propositions belonging to the thesis, entitled
"Uncovering oomycete metabolism using systems biology"

Sander Y.A. Rodenburg
Wageningen, 15 September 2020

Uncovering oomycete metabolism using systems biology

Sander Y.A. Rodenburg

Thesis committee

Promotors
Prof. Dr F.P.M. Govers
Personal chair at the Laboratory of Phytopathology
Wageningen University & Research

Prof. Dr D. de Ridder
Professor of the Bioinformatics Group
Wageningen University & Research

Co-promotors
Dr M.F. Seidl
Assistant Professor, Theoretical Biology & Bioinformatics
Utrecht University

Other members
Prof. Dr J.M. Wells, Wageningen University & Research
Prof. Dr J.M. McDowell, Virginia Tech, Blacksburg CA, USA
Prof. Dr V. van Noort, Leiden University
Dr M. Suarez Diez, Wageningen University & Research

This work was conducted under the auspices of the Graduate School Experimental Plant Sciences.

Uncovering oomycete metabolism using systems biology

Sander Y.A. Rodenburg

Thesis

submitted in the fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. Dr A.P.J. Mol
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Tuesday 15 September
at 16:00 p.m. in the Aula

Sander Y.A. Rodenburg
Uncovering oomycete metabolism using systems biology
174 pages

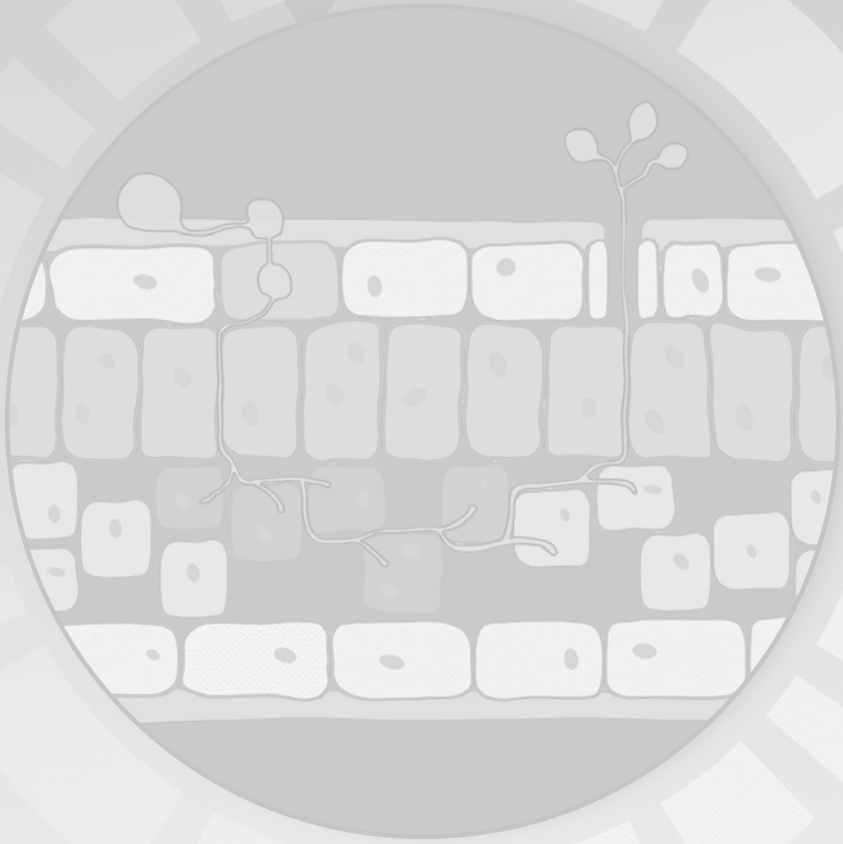
PhD thesis, Wageningen University, Wageningen, the Netherlands (2020)
With references, with summaries in English and Dutch

ISBN: 978-94-6395-494-5
DOI: 10.18174/528798

Table of contents

Chapter 1

General introduction



Metabolism is essential for life. The metabolism of an organism defines its capabilities to take up nutrients from the environment and to convert these into its essential building blocks, such as nucleic acids and amino acids (Lazar and Birnbaum, 2012). Cellular metabolism can be described as a system of biochemical conversions (reactions), most of which are catalyzed by metabolic enzymes. Typically, a genome encodes a few thousand metabolic enzymes (Yilmaz and Walhout, 2017). Each enzyme acts on a selection of substrates and converts these into products, typically by adding or removing reactive groups. Reactions that share substrates or products can be considered functionally connected. Consequently, the collection of biochemical reactions within a cell forms a large interconnected network, representing the routes by which the organism converts simple nutrients into complex metabolites and vice-versa. This network is distributed over different subcellular compartments (organelles). Transporter proteins and channels facilitate the transport of metabolites across the lipid bilayers that surround the cell and the organelles (Sahoo *et al.*, 2014). The overall system is subject to many parameters, such as variability in substrates, temperature, or the pH, not only between cells and the extracellular space but also within cells. Cells regulate this system to maintain homeostasis, i.e. the ability to perform important cellular functions despite variations (perturbations), and this provides robustness (Eberl, 2018; Nijhout *et al.*, 2019). The ability to sense environmental variations and metabolic cues and to adapt metabolism accordingly, depends on a tightly interlinked regulatory system (Watson *et al.*, 2015). This involves regulatory feedback loops embedded in interaction networks crossing metabolic, protein, transcript, and (epi)genetic levels (**Figure 1**).

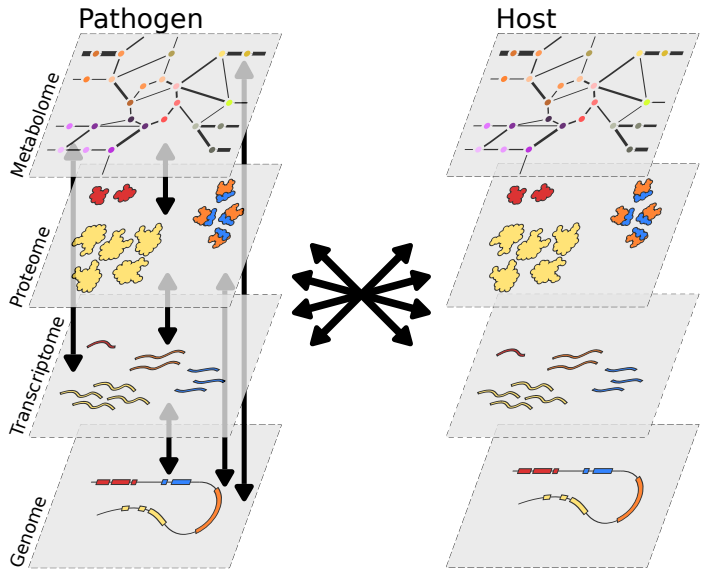


FIGURE 1 | The molecular layers of a cell are all interconnected, and form a complex and integrated system. In the symbiosis of a pathogen and a host, their systems are connected, and interactions occur between all molecular layers.

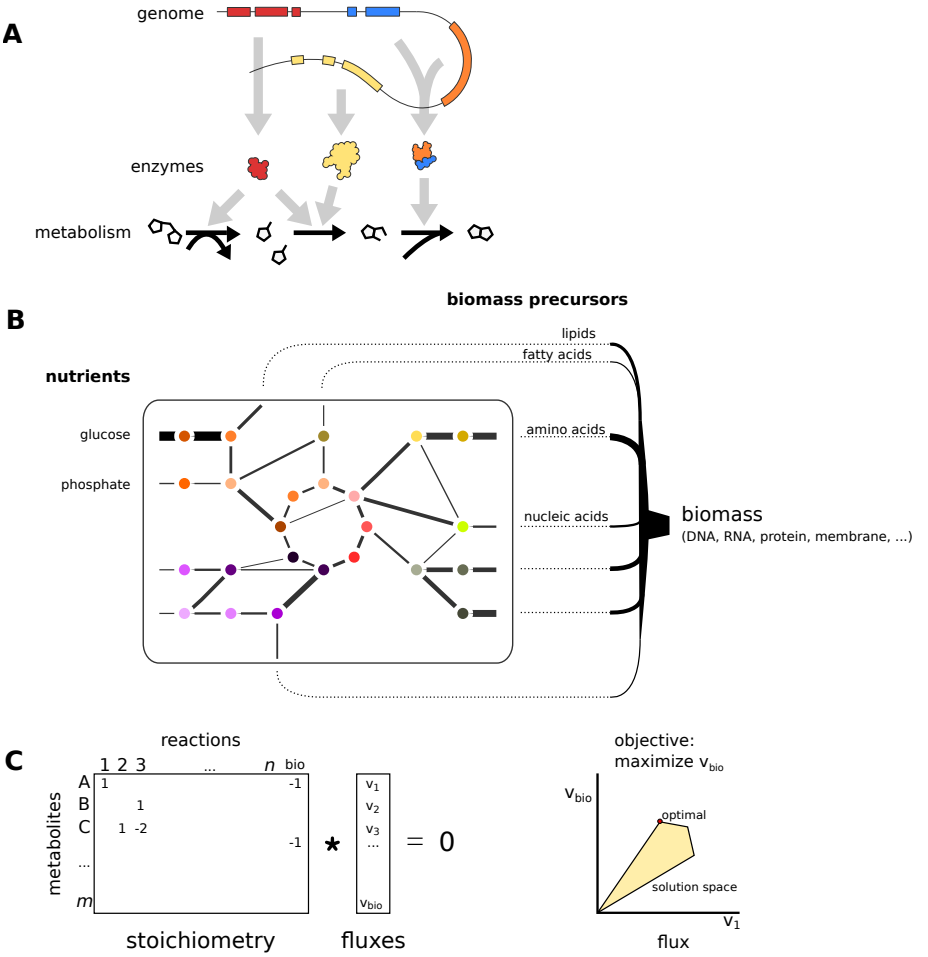
As such, the phenotype of the cell is an emergent property of the system's complexity (Aderem, 2005). The rates of individual metabolic reactions are a direct consequence of the overall state

of cellular metabolism, and therefore understanding a small part of the system (e.g. a single enzyme or pathway) provides only limited insight into the complete system. Thus, a more holistic view is essential to understand how the state of a system can lead to the complex phenotype of an organism.

Systems biology

The rise of computational biology and high-throughput measurement technologies in the early 21st century enabled biologists to measure the presence, quantity, and/or interactions between molecules in the cell at genome-scale (Reed *et al.*, 2006). Today, numerous species have their genome sequenced, and transcriptome sequencing has become routine. Additionally, high-throughput methods such as mass spectrometry allow the analyses of the proteome and metabolome. Data analyses typically focus on the differential abundances of these molecules (transcripts, proteins, or metabolites) between samples. Alternatively, omics data can be used to construct models of molecular systems, as part of a scientific discipline that is known as systems biology (Bordbar *et al.*, 2014), a relatively young scientific discipline that integrates — among others — bioinformatics, mathematics, biochemistry, molecular biology, physics and engineering (Breitling, 2010). In systems biology, omics data is integrated with prior experimental knowledge into models of molecular interactions at genome-scale, aiming to capture the complete molecular systems that result in the phenotype of an organism (Yurkovich and Palsson, 2016). This includes models of protein-protein interactions, metabolic fluxes, regulatory interactions, or signaling pathways (Albert, 2007). Interaction networks can be investigated for topological features, for instance to identify critical connectors or central elements, or can be analyzed for potential biological functions, such as the presence of particular molecular pathways (Winterbach *et al.*, 2013). Importantly, these types of molecular networks are typically static and do not provide insights into the dynamics of these interactions. In contrast, systems modelling involves the mathematical abstraction of a system's behavior to study its dynamic properties. Many different types of models exist, such as ordinary differential equation (ODE) based models for enzyme kinetics, Bayesian, Boolean, or rule-based models for signaling and regulatory networks, or constraint-based models of metabolism (genome-scale metabolic models or GEMs) (**Box 1**) (Bartocci and Lió, 2016; Bordbar *et al.*, 2014). In particular GEMs are nowadays popular tools to simulate the cellular metabolism of a species, and have shown remarkable successes in modelling metabolic fluxes within an organism (Gu *et al.*, 2019). GEMs are based on the predicted enzyme repertoire based on the genome sequence, and the associated network of biochemical reactions with substrates and products. This property enables the integration of miscellaneous omics data and prior knowledge about the metabolic properties of an organism (Zhang and Hua, 2015). It also allows the *in silico* identification of the essential genes/reactions/metabolites of the cell, by means of predicting the effect of enzyme knockouts on the functioning of the system (Chavali *et al.*, 2012). As such GEMs can serve as a knowledge base and guide future research.

Box 1 – Genome-scale metabolic models



Basic principles of GEMs. **A**: GEMs are reconstructed from a genome sequence, and connect enzymes to reactions. **B**: A GEM is a network that simulates fluxes from nutrient uptake to the production of biomass precursors. **C**: The mathematical representation of a GEM. Reactions are stored in a stoichiometric matrix, that is multiplied by a vector of fluxes. The solution space is limited by constraints. Here only two fluxes are shown, but the optimization is of *n* dimensions.

Metabolism is a complex system of thousands of biochemical reactions, most of which are catalyzed by metabolic enzymes. Genome-scale metabolic models (GEMs) have been developed as an effective way to generate hypotheses about cellular metabolism (Yurkovich and Palsson, 2016). A GEM is based on the repertoire of metabolic enzymes encoded in the genome, which are cross-referenced with biochemical databases or template models to obtain a set of biochemical reactions that can transform substrate metabolites into products. These reactions are interconnected by shared substrates or products, forming a complex network that is typically divided over several subcellular compartments, such as the mitochondria and the cytosol. The metabolic network can be represented by a sparse integer matrix, denoting the stoichiometry of substrate and product metabolites of each reaction.

Box 1 – Genome-scale metabolic models (*Continued*)

GEMs are assumed to be in steady-state, which means the net uptake of nutrient mass is equal to the net production of biomass, implying there is no accumulation of metabolites. This simplifies the model to a system of linear equations. Each reaction in the GEM is considered to have a flux, i.e. a steady-state reaction rate. The most popular method to model metabolism in this framework is called flux balance analysis (FBA) (Orth *et al.*, 2010). An assumption is that cells have a specific objective, often maximization of growth (production of biomass) or minimal usage of energy (García Sánchez and Torres Sáez, 2014). Linear optimization can find values for all fluxes that attain the specified objective, for instance maximal flux towards biomass precursors. The production of biomass is modelled as a pseudo-reaction that consumes all biomass precursors (e.g. amino acids for proteins) with appropriate stoichiometry, often manually implemented based on experimental data (Feist and Palsson, 2010). As there is an infinite number of solutions that allow biomass production, constraints need to be implemented to find a single set of fluxes.

Thermodynamic constraints can implement an upper and lower bound on each flux, specifying that reactions are either bidirectional (flux can be either negative or positive) or irreversible (either the upper or lower bound is zero). These constraints significantly limit the number of possible outcomes of the optimization problem. The flux values that yield a maximal value for the specified objective, within the constraints, are then selected as an optimal solution.

The integration of omics data can be used to impose additional model constraints or objective functions (Bordbar *et al.*, 2014). Since most reactions of a GEM are associated with one or more genes, optimization can take into account gene expression and calculate the fluxes that concur with underly ing gene expression. For instance, the INIT algorithm (Agren *et al.*, 2012) maximizes a global score, which increases when "expressed" reactions have a flux and decreases when "non-expressed" reactions have a flux. As a result, the optimal solution depends on the expression of genes (i.e. context-specific submodels), which enables the comparison of fluxes between transcriptome conditions. Similarly, metabolomics data can be used to calculate fluxes that attain the presence of measured metabolites. Note that these are only two specific examples of omics integration into GEMs, in a rapidly expanding field of methods (Lewis *et al.*, 2012).

The linear program of a GEM can often be solved in a matter of seconds on modern computers, making it a powerful computational tool to predict the effect of perturbations to GEMs (Peyraud *et al.*, 2018). For instance, the impact of a different growth medium can be analyzed (i.e. change of nutrient uptake reactions), or genes/reactions can be iteratively removed from the model to investigate the effect on the metabolic fluxes. Reactions that have large effects on the flux distribution or biomass production upon removal suggest biological relevance (Chavali *et al.*, 2012). Essential genes or reactions can be predicted when their elimination yields an infeasible problem, i.e. no solutions respecting the implemented constraints (e.g. no biomass flux possible after removal) (Pratapa *et al.*, 2015). Similarly, synthetic lethal gene or reaction pairs can be identified that will only impair biomass production upon simultaneous deletion.

Systems biology on pathogens

Systems biology offers a powerful toolbox to study pathogens and their relation with their hosts (Cesur *et al.*, 2018; Dix *et al.*, 2016; Durmus *et al.*, 2015; Horn *et al.*, 2012; Peyraud *et al.*, 2017). Pathogens and hosts often interact extensively on all molecular levels, i.e. metabolic, protein and DNA/RNA (**Figure 1**). Co-evolution shapes a host's immune system to be able to recognize the presence or action of a pathogen and to activate immune responses (Cook *et al.*, 2015). To counter these processes and to facilitate infection, a pathogen secretes

virulence factors in the form of proteins (effectors), small RNAs, or (secondary) metabolites (Frantzeskakis *et al.*, 2020). The symbiosis between pathogen and host can be regarded as a single intertwined system that is merely separated into different compartments and connected by their interactions (Olive and Sassetti, 2016). The reconstruction of a model for such a system can help to characterize the pathogen-host interactions by generating testable hypotheses. Systems biology has been used in the study of various pathogens or pathogen-host interactions to identify drug targets or key factors that allow pathogens to interact with their host (Durmus *et al.*, 2015). In pathogen-host interactions, protein-protein or small RNA interaction networks have been investigated using graph theory to identify pathogen effectors and their host interactors, in which network centrality or degree is considered a proxy for functional importance (Tekir and Ülgen, 2013). Additionally, GEMs can be used to simulate the system-wide metabolic fluxes of a pathogen, and are therefore useful tools to predict the important genes/reactions/metabolites of the cell, which can inspire control strategies (Chavali *et al.*, 2012). Not surprisingly, the first GEM ever generated was for a microbial pathogen, i.e. the bacterium *Haemophilus influenzae* that causes disease in humans (Edwards and Palsson, 1999). Since then, GEMs have been reconstructed for many more pathogens, such as the tuberculosis bacterium *Mycobacterium tuberculosis* (Kavvas *et al.*, 2018; Rienksma *et al.*, 2018) and the human and animal parasites of the genera *Plasmodium* (Plata *et al.*, 2010; Stanway *et al.*, 2019) and *Leishmania* (Chauhan and Singh, 2019; Sharma *et al.*, 2017; Subramanian *et al.*, 2015). Some GEMs integrated pathogen and host, thereby providing insight into the metabolic fluxes throughout infection (Bazzani *et al.*, 2012; Bordbar *et al.*, 2010; Huthmacher *et al.*, 2010).

In contrast to human pathogens, plant pathogens have hardly been studied taking a systems biology approach (Peyraud *et al.*, 2017). Yet, similar to human pathogens, plant pathogens have a major negative impact on the well being of their hosts. Plants are crucial for generating the oxygen (O_2) we breathe, for sequestering CO_2 and maintaining the balance in the global ecosystem, and for the production of food and feed. However, plants are under constant threat of pathogens, such as fungi, oomycetes, bacteria, and viruses. In agriculture, the resulting yield losses can be substantial, reaching up to 30% (Savary *et al.*, 2019). To combat plant diseases, a better understanding of plant-pathogen interactions is important. There are however only a few examples where a systems biology approach was applied to provide insight into the molecular mechanisms underlying plant-pathogen interactions. In one study a protein-protein interaction network of *Arabidopsis thaliana* and various pathogens of different kingdoms uncovered that effectors from different pathogens convergently target the same host proteins (Weßling *et al.*, 2014). In other studies, GEMs have been reconstructed for the bacterial plant pathogens *Ralstonia solanacearum*, *Xanthomonas oryzae,* and *Pectobacterium parmentieri* (Koduru *et al.*, 2020; Peyraud *et al.*, 2016; Zoledowska *et al.*, 2019), and for the fungus *Sclerotinia sclerotiorum* (Peyraud *et al.*, 2019). However, despite the abundance of omics data for many pathogens, very few have been analyzed from a systems biology perspective.

Oomycete-host interactions

Oomycetes are filamentous eukaryotes that resemble fungi in terms of morphology but evolved independently from fungi (McGowan and Fitzpatrick, 2020). In the tree of life, oomycetes group together with the brown algae and diatoms in the Stramenopile lineage (Beakes *et al.*, 2012). Many oomycetes are plant pathogens, but some can infect animals or other microbes, or live as saprophytes (Derevnina, Petre, *et al.*, 2016). Their lifestyles vary. Some are necrotrophic pathogens that swiftly kill their host plant and feed off dead plant material (Fawke *et al.*, 2015). Others are biotrophs that need living host tissue to infect, feed and proliferate. Hemibiotrophs such as the members of the genus *Phytophthora* have a biotrophic phase followed by a necrotrophic phase. Most of the biotrophic oomycetes, such as the white rusts and downy mildews, are obligate pathogens implying that they cannot grow outside a living host (Baxter *et al.*, 2010). One of the best studied oomycetes is *Phytophthora infestans* that causes late blight disease on potato and tomato (Kamoun *et al.*, 2015). Shortly after the first genome sequences of oomycetes were published and high-throughput omics data for oomycetes became more abundant (Ali *et al.*, 2017; Tyler *et al.*, 2006), it was already proposed to reconstruct predictive models for *P. infestans* to reveal its mechanisms of host interaction (Pinzón *et al.*, 2009; Pinzón *et al.*, 2011).

Biotrophic plant pathogens typically adhere to leaves or roots, break physical barriers and scavenge nutrients from their host, while suppressing the host's immune system (McDowell, 2011). They achieve this by secreting a large variety of enzymes and effector proteins that end up in the apoplast or inside the plant cell and pave the way for a successful infection. In biotrophic oomycetes such as *Phytophthora* spp. and downy mildews, typically hundreds of effector genes are identified in a genome (McGowan and Fitzpatrick, 2017; Whisson *et al.*, 2016). During infection, these biotrophs grow as filamentous hyphae inside their hosts. They colonize the apoplastic space in the leaf mesophyll and form feeding structures, so called haustoria, inside host cells. At the site of the haustoria and the apoplastic hyphae, there is a close interface between pathogen and host, through which effectors, enzymes and small molecules can be exchanged (Judelson and Ah-Fong, 2018). Oomycetes typically reproduce by forming sporangia on hyphae that emerge from stomata a few days after infection (Judelson and Blanco, 2005). These are dispersed by wind or water and after landing on a host surface they germinate directly or release zoospores that encyst and germinate. The tip of the germ tube then develops into an appressorium, an infection structure that the pathogen uses to penetrate host cells.

The ability of oomycetes to live in close symbiosis with a host drives continuous adaptations of both, pathogen and host. Oomycetes have dynamic genomes that allow swift adaptation (Leesutthiphonchai *et al.*, 2018). Comparative genomics has revealed that obligate biotrophic pathogens have suffered extensive gene loss as a result of their biotrophic lifestyle (Fletcher *et al.*, 2018; Kemen and Jones, 2012). This adaptive capacity facilitates the evolutionary "arms race" between oomycete effectors and host resistance genes (Y. Wang *et al.*, 2019),

but also allows adaptation of their core cellular machinery including metabolism and signal transduction, leading to various unique properties (Judelson, 2017). For instance, oomycetes have several genes encoding unique proteins with novel domain combinations (Seidl *et al.*, 2011), as well as a number of horizontally transferred genes that encode proteins with functions in metabolism (Richards *et al.*, 2011). Oomycetes are osmotrophs, which means they secrete enzymes to digest large molecules (polymers) extracellularly and import the resulting small molecules as nutrients (Richards and Talbot, 2013). This process is facilitated by a broad array of transporter proteins suggesting that a plethora of host compounds can be taken up during infection (Abrahamian *et al.*, 2016). However, some nutrients are indispensable for oomycetes. Many oomycetes lost the ability to synthesize sterols and therefore secrete elicitin proteins that are thought to be sterol carriers (Dahlin *et al.*, 2017; Derevnina, Dagdas, *et al.*, 2016). Moreover, most oomycetes are auxotrophic for thiamine, a vitamin that acts as a cofactor in carbohydrate catabolism (Hohl, 1991). Culturable oomycetes can be grown *in vitro* and seem to prefer amino acids as a substrate (Ah-Fong *et al.*, 2017b; Hodgson, 1958), but can utilize a wide variety of substances. Due to the complex nature of cellular metabolism, it is currently unclear which nutrients are more important than others, and how the differential usage of nutrients influences infection. Their adaptive capacity makes oomycete pathogens very challenging to control. It is therefore of importance to reconstruct holistic models that provide mechanistic insight into the molecular systems that allow oomycetes to proliferate and infect their hosts. Ultimately, a system-wide understanding of oomycete-host interactions might provide novel leads for control (Dunphy and Papin, 2018).

Scope of this thesis

The overall aim of this thesis is to investigate how the available omics data of oomycetes can be exploited to unravel the molecular mechanisms that allow oomycetes to proliferate and infect their hosts. We take a systems biology approach, integrating these data into holistic models of cell metabolism and putative factors for host interaction (**Figure 2**).

To address the question to what extent metabolic networks are representative for lifestyle we compare the metabolic gene content and evolutionary patterns of animal and plant pathogenic oomycetes with various trophic lifestyles (**Chapter 2**). The metabolic networks of obligate biotrophic oomycetes are compared with those of other plant pathogenic oomycetes, showing nonrandom patterns of gene loss related to obligate biotrophy from a network topological aspect.

To gain more insight into the metabolism of oomycetes, we choose *P. infestans* as a representative to reconstruct a GEM. We reconstruct the GEM using available biochemical information on *P. infestans* and modern model reconstruction techniques (**Chapter 3**). We then simulate the metabolic fluxes of *P. infestans* and those throughout its life cycle, by integration of transcriptome data. We also pinpoint several essential reactions and study the pathways involved.

As a pathogen *P. infestans* lives in close association with its host. To investigate the dynamics of its metabolism in symbiosis we take the *P. infestans* GEM one step further, and integrate it with a GEM of tomato (**Chapter 4**). This leads to a multi-compartment model of the *P. infestans*-tomato interaction. We employ various model-based methods to investigate the topology, connectivity and metabolic fluxes of the model, to reveal host-pathogen metabolic dependencies. We also integrate infection time-course transcriptome data to simulate the differential usage of nutrients by *P. infestans* throughout infection.

For pathogenesis oomycetes produce a large variety of effectors and other putative virulence factors among which enzymes. In **Chapter 5** we focus on a specific class of enzymes, the metalloproteases (MPs), We screen Stramenopile genomes for the presence of MP genes, create orthology groups, compare the MP repertoire in oomycete and non oomycete species, and analyse MP domain compositions. For the *P. infestans* MPs we perform gene model curations and select the ones that have a signal peptide for an *in planta* assay, to reveal potential virulence functions.

In **Chapter 6** we provide an outlook on how systems biology can be applied to study pathogens, and how models can be used as a driver of innovation. We discuss the limitations we encountered and the prospects of the models presented in this thesis, and reflect on the interpretation of the models. Finally, we propose a framework for improving the models to establish a high-quality model for *P. infestans* and late blight disease.

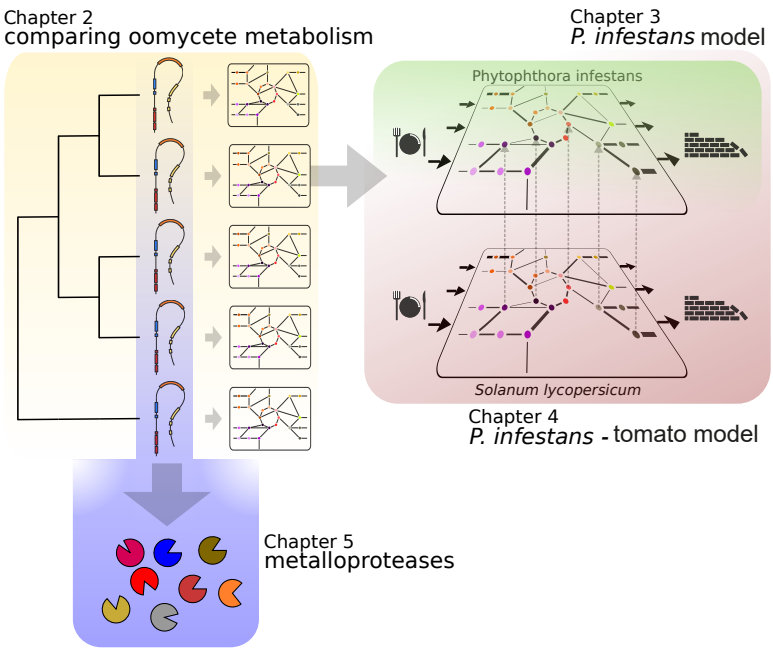


FIGURE 2 | Graphical abstract of this thesis.

Chapter 2

Oomycete metabolism is highly dynamic and reflects lifestyle adaptations

Sander Y.A. Rodenburg,
Dick de Ridder,
Francine Govers,
Michael F. Seidl

This chapter is a slightly modified version of the manuscript posted on BioRxiv.org
doi: https://doi.org/10.1101/2020.02.12.941195

Abstract

The selection pressure of a pathogen-host symbiosis drives adaptations. Animal and plant pathogenic oomycetes are thought to adapt their metabolism to facilitate interactions with their hosts. Here, we performed a large-scale comparison of oomycete metabolism to relate metabolic adaptations to lifestyle differences. We uncovered considerable variation in oomycete metabolism that could be linked to lifestyle differences. Comparisons of metabolic gene content revealed that plant pathogenic oomycetes have a bipartite metabolism; a conserved part and an accessory part. The accessory part could be associated with the degradation of plant compounds produced during defence responses. Obligate biotrophic oomycetes in particular have smaller metabolic networks, and taxonomically distantly related biotrophic lineages display converged evolution by repeated gene losses in both the conserved as well as the accessory parts of metabolism. We investigated differences between the metabolic networks of obligate biotrophic oomycetes and those of hemibiotrophic oomycetes, and revealed that the losses of metabolic enzymes in biotrophs are not random and that the networks contract from the periphery inwards. Our analyses represent the first metabolism-focused comparison of oomycetes at this scale and will contribute to a better understanding of the evolution and relationship between oomycete metabolism and lifestyle adaptation.

Introduction

To grow, organisms need to assimilate nutrients from the environment and assemble these into the building blocks of life – i.e. polymers such as proteins, DNA, carbohydrates and lipids. This is known as cellular metabolism, a process encompassing a network of thousands of biochemical reactions, many of which are catalyzed by metabolic enzymes (J. Nielsen, 2017). The repertoire of metabolic enzymes encoded in the genome of a specific species is indicative for its biochemical capabilities. Autotrophic organisms have the capability to assimilate inorganic compounds from the environment and to assimilate these into organic compounds. In contrast, heterotrophic organisms need organic nutrients to generate complex compounds and rely on other organisms for nutrition (Nelson and Cox, 2017). Parasitism is typically a heterotrophic lifestyle; the parasite lives in symbiosis with a host and takes up nutrients from it (Poulin, 2007). This symbiosis, which requires a close interface between parasite and host, is the result of a co-evolutionary process that has led to adaptations in the metabolism of both symbionts (Divon and Fluhr, 2007; Poulin and Randhawa, 2015). Many obligate parasites show an evolutionary pattern called reductive evolution (Casadevall, 2008), which manifests in the losses of primary metabolic enzymes and pathways (Corradi, 2015). This results in a dependency of the parasite on the host for nutrition (Kemen and Jones, 2012). Many other parasites evolved a more offensive strategy and have selectively expanded and diversified specific gene families to degrade host compounds, either for defence or for nutrition (Carere *et al.*, 2016; Lowe-Power *et al.*, 2016; Morales-Cruz *et al.*, 2015; Richards and Talbot, 2013). Together, these processes shape the metabolic network of a parasite such that it is optimized for symbiosis with its host(s) and for utilizing host nutrients to achieve maximal growth and reproduction.

Oomycetes form a diverse group of successful parasites (Kamoun *et al.*, 2015), many of which are serious pathogens of plants or animals (Fawke *et al.*, 2015; van West and Beakes, 2014). Even though oomycetes share many morphological and lifestyle characteristics with fungi (Judelson and Blanco, 2005), they evolved independently. Oomycetes are phylogenetically related to diatoms and brown algae, and united with these organisms in the Stramenopile lineage. It has been proposed that approximately 400-800 million years ago oomycetes evolved from an autotrophic algae-like marine ancestor, and subsequently lost the capacity to perform photosynthetic metabolism in adaptation to a heterotrophic lifestyle (Beakes *et al.*, 2012; Matari and Blair, 2014). Most oomycetes described so far are plant pathogens. Based on their lifestyle, they can be divided into different groups (Fawke *et al.*, 2015). Obligate biotrophs fully depend on a living host to proliferate. By contrast, necrotrophs kill their host upon infection and feed saprophytically on the decaying tissue. Lastly, hemibiotrophs have an initial biotrophic phase followed by a necrotrophic phase.

Sequencing and comparative analyses of oomycete genomes have revealed large differences in genome sizes and gene content, suggesting highly dynamic genomes (McGowan *et al.*, 2019; Raffaele and Kamoun, 2012; Seidl *et al.*, 2011; Seidl *et al.*, 2012). Apparently, oomycetes rapidly evolved into highly divergent species that vary in growth substrates and lifestyle

(Danies *et al.*, 2013; Hodgson, 1958; Leesutthiphonchai *et al.*, 2018) and they have the capacity to readily adapt to the environment, to overcome host resistance, or to develop fungicide resistance (Derevnina, Petre, *et al.*, 2016; Matson *et al.*, 2015). It is conceivable that many of these adaptations are driven by a pathogenic lifestyle (Ah-Fong *et al.*, 2019). The genome dynamics also translate into fundamental differences in metabolism (Judelson, 2017). Notably, oomycetes acquired several metabolic enzymes from bacteria and fungi by horizontal gene transfer to support the breakdown and uptake of host nutrients (Richards *et al.*, 2011; Richards and Talbot, 2013; Savory *et al.*, 2015; Savory *et al.*, 2018). Furthermore, oomycetes classified as obligate biotrophs, such as downy mildews and white rusts (e.g. *Hyaloperonospora*, *Albugo*, *Peronospora, Plasmopara*), have lost various enzymes of assimilation pathways (e.g. for nitrate and sulfate). Such losses are generally regarded as characteristic of an obligate parasitic lifestyle (Dean *et al.*, 2012; Fletcher *et al.*, 2018; McDowell, 2011; Spanu, 2012; Thines *et al.*, 2020). An understanding of pathogen metabolism in relation to lifestyle can yield important information about the growth conditions and preferred substrates and, ultimately, might point to novel leads for disease control (Garavito *et al.*, 2019; Peyraud *et al.*, 2017; Warrilow *et al.*, 2014).

In this study we performed large-scale analyses of the metabolism of oomycetes and closely related Stramenopiles, with the aim to shed light on lifestyle- and lineage-specific adaptations in metabolism in oomycetes during diversification from their last common ancestor. We analyzed the conservation of metabolic enzymes through pan- and core-genome analyses and translated the metabolic enzyme content into metabolic networks. Comparison of these networks revealed the impact of gene losses on the metabolism of obligate biotrophs.

Results & Discussion

Metabolic gene content in Stramenopiles

Metabolic enzymes catalyze most of the biochemical reactions in the cell, collectively orchestrating cellular metabolism (J. Nielsen, 2017). To compare the metabolism of oomycetes and related sister-species, we identified the metabolic enzyme repertoires in the complete predicted proteomes of 54 species, encompassing nearly all Stramenopiles for which a publicly available genome sequence with associated gene annotation was available (as of early 2019). This collection included 42 oomycetes, six diatoms, four brown algae, a Labyrinthulomycete (*Hondaea fermentalgiana*), and a hyphochytrid (*Hypochytrium catenoides*) (**Table S1; Figure 1A**). The latter two species are members of two sister clades to the oomycetes (Leonard *et al.*, 2018). The oomycetes in our species selection have a variety of lifestyles: one mycoparasite (*Pythium oligandrum*), two saprotrophs (*Thraustotheca clavata*, *Aphanomyces stellatus*) (Misner *et al.*, 2014), six animal pathogens (two *Saprolegnia* spp., *Achlya hypogyna, Pythium insidiosum,* two *Aphanomyces* spp.) and 33 plant pathogens including 19 hemibiotrophic *Phytophthora* spp., eight necrotrophs (*Aphanomyces euteiches* and seven *Pythium* spp.), and six biotrophs (*Albugo laibachii*, two *Peronospora* spp., two *Plasmopara* spp. and

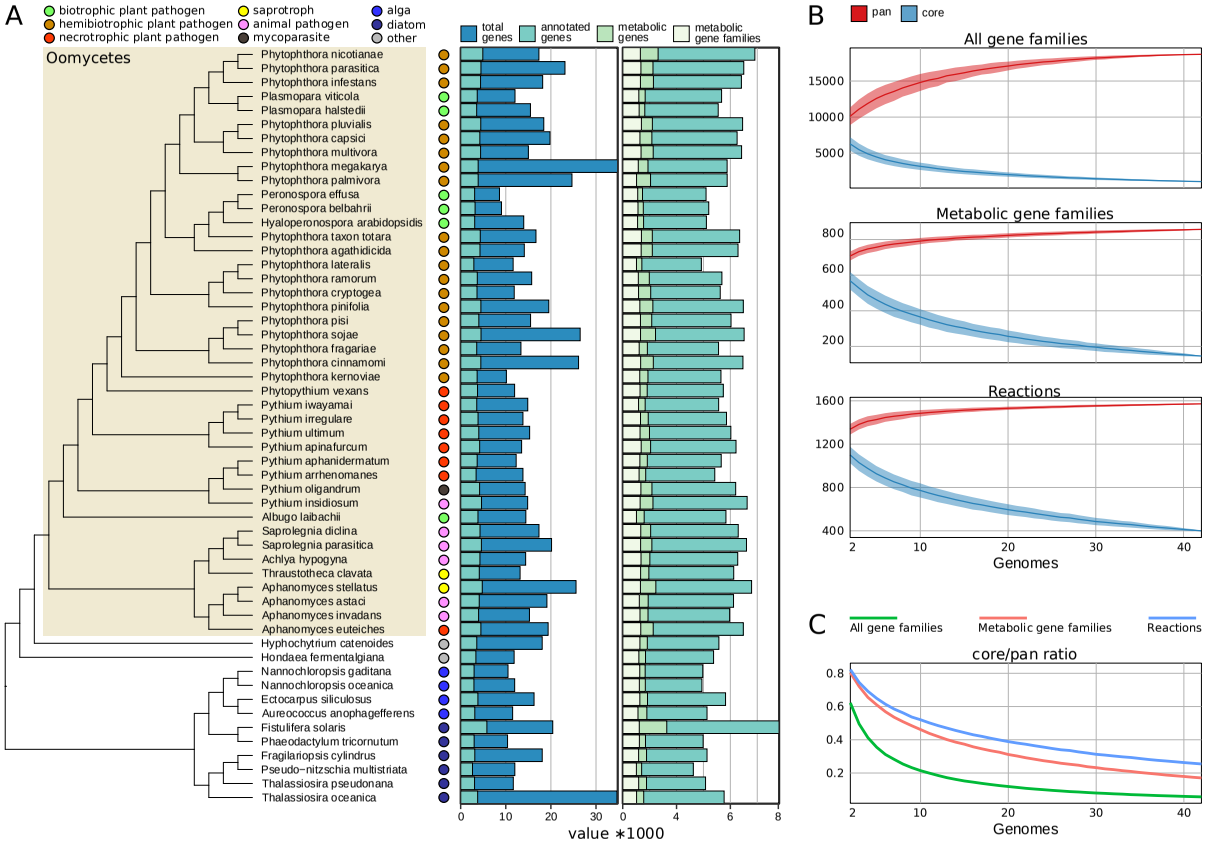


FIGURE 1 | Phylogeny, genome content and conservation of gene content in Stramenopiles. A: Species phylogeny of 54 Stramenopiles reconstructed from single-copy orthologous gene clusters (branch lengths not calibrated). Oomycetes are shaded, and the lifestyle of each species is displayed by colored dots (key: top left). Bar charts indicate the total number of (annotated) genes per species and the total number of genes per species annotated as (unique) metabolic genes. **B:** Pan and core size estimates of 2-42 oomycete genomes. The size of the core (blue, present in all) and pan (red, present in any) genome based on the overall set of gene families, metabolic gene families, and metabolic reactions, with ribbons indicating standard deviation, based on 1,000 permutations. **C:** The core/pan ratio for overall gene families, metabolic gene families, and metabolic reactions.

Hyaloperonospora arabidopsis) (**Figure 1A**). The obligate biotrophic plant pathogens in the genera *Hyaloperonospora*, *Plasmopara* and *Peronospora* are nested within the *Phytophthora* genus (Thines and Choi, 2015), a genus which exclusively contains hemibiotrophic plant pathogens. In contrast, the genera *Pythium* and *Aphanomyces* are more diverse with respect to lifestyle; apart from necrotrophic plant pathogens, they also contain saprotrophs and animal pathogens, and even a mycoparasite (*P. oligandrum*) (Adhikari *et al.*, 2013; Benhamou *et al.*, 2012; Diéguez-Uribeondo *et al.*, 2009; Marano *et al.*, 2016). The Stramenopiles analyzed here have on average 16,345 predicted protein-coding genes, but display considerate variations between individual species (standard deviation of 5,586 annotated genes). *Phytophthora megakarya* has the highest number of predicted genes (34,804), likely driven by extreme gene family expansions in particular in virulence-related genes (Ali *et al.*, 2017). In contrast, the two *Peronospora* spp., *P. effusa* and *P. belbahrii*, have the lowest number of genes (8,603 and 9,049, respectively).

To learn about the gene content of each genome, we selected one representative protein sequence for each gene, and annotated all proteins with KofamScan (Aramaki *et al.*, 2019), using profile hidden Markov models of the KEGG Orthology (KO) database (Mao *et al.*, 2005). The KO database contains manually defined clusters of protein families with a validated enzymatic function. KofamScan uses an adaptive score threshold for assigning proteins to a KO, determined by the minimal F1-measure (a measure for accuracy) when matching positive and negative training data (Aramaki *et al.*, 2019). KofamScan thus aims to find an optimal trade-off between incorrectly assigning proteins to a KO (false positives) and missing putative orthologs (false negatives). To evaluate the performance of KofamScan for ortholog detection, we randomly selected 100 KOs and matched all associated proteins in KEGG to their respective HMMs (**Figure S1**). The majority of the HMMs (82%) picked up at least 80% of its orthologs, although we also noted that some KO HMMs were far less sensitive, likely as a result of a higher score threshold. As we assume that these biases affect our analyses in a uniform manner, we considered KofamScan a suitable method to provide a global overview of gene content.

On average, 25% of the proteins for each species could be assigned a KO identifier (referred to as KO), ranging from 11% to 36% (2,633 to 5,815 proteins, respectively). The total number of genes and the number of KOs per species were correlated (Pearson correlation r=0.49; P=1.3×10^{-4}), yet with considerable variation. The majority of the proteins do not show homology to any KO, i.e. their specific function is largely unknown. However, the differences in number of KOs between species reflect expansions and reductions of gene families within the oomycetes (McGowan *et al.*, 2019; Seidl *et al.*, 2012). For example, it is known that oomycete genomes have large variations in the number of genes encoding secreted virulence-related proteins (effectors), many of which show little to no homology to known proteins or domains outside the oomycetes (McGowan and Fitzpatrick, 2017).

To investigate to which extent differences in total gene numbers influence the metabolic capacity of oomycetes, we selected all KOs representing a metabolic enzyme, which we define

as enzymes associated with a biochemical conversion of small molecules (**Figure 1A**). With respect to the correlation between KO numbers and total gene numbers that we observed earlier, the numbers of metabolic enzymes showed even less correlation with the total gene numbers (Pearson correlation r=0.39; P=0.0031), suggesting that the metabolic gene content is more stable than other genes. On average, the species in our collection contained 981 metabolic enzymes. An outlier is the diatom *Fistulifera solaris* (1637 metabolic enzymes), which underwent a whole-genome duplication (Tanaka *et al.*, 2015). In contrast, obligate biotrophic species of the genera *Peronospora, Plasmopara*, *Hyaloperonospora,* and *Albug*o have a reduced number of metabolic enzymes (on average 787) (one-tailed rank-sum test P=2.4×10^{-4}). For the other oomycetes however, we could not relate the metabolic enzyme numbers to lifestyle; the pathogenic oomycete species appear to have a similar proportion of metabolic enzymes as free-living autotrophic species, such as the diatoms and algae. This suggests a similar magnitude of expansions/contractions of gene families encoding metabolic enzymes (metabolic gene families) for most species, resulting in similar-sized sets of metabolic enzymes in the proteomes. It should be noted, however, that these numbers do not provide insight into the metabolic capacity of each species, as the size of a metabolic gene family is just one of the many factors that determine the metabolic fluxes of an organism (Ah-Fong *et al.*, 2019). Factors such as (post-)transcriptional/translational regulation, reaction stoichiometry, enzyme activity, pH, and temperature all contribute to the physiological state of an organism (Wegner *et al.*, 2015). To compare metabolic capacity among species it is therefore more relevant to compare the number of metabolic gene families, i.e the number of unique metabolic enzymes, without taking into account copy number variation. This analysis revealed a similar number of metabolic gene families for each species (622±51) (**Figure 1A**). Yet, obligate biotrophs have a consistently lower number of metabolic gene families than the other species in our collection (on average 566 versus 629, respectively; one-tailed rank-sum test; P=1.8×10^{-3}), which reflects the metabolic gene losses in their adaptation to an obligate biotrophic lifestyle (Thines *et al.*, 2020).

Dynamics of metabolic gene content

To uncover the evolutionary patterns causing the differences in metabolic gene content, we used a maximum-likelihood approach to reconstruct gains and losses of metabolic genes families for each branch in the phylogenetic tree (**Figure S2**). This reconstruction revealed that novel gains and losses of metabolic gene families occurred frequently throughout the evolution of oomycetes. Especially within the *Phytophthora* lineage evolutionary events are numerous, which is striking considering their close evolutionary relationship. This clearly reflects the dynamics of oomycete metabolic gene content, potentially leading to differences in metabolic capacity among these species (Seidl *et al.*, 2012). The numbers of gene family losses are higher than the numbers of gene family gains (on average 61 and 42 respectively; paired samples rank-sum test; P=1.4×10^{-6}). This pattern is especially prevalent for the obligate biotrophs, which have lost on average 101 metabolic gene families. *Albugo laibachii* is the most extreme example, with 8 family gains and 168 losses, but *Phytophthora ramorum, Phytophthora lateralis* and *Pythium iwayamai* also displayed extensive losses (179, 110 and 113, respectively). Taken together, these

results highlight the highly variable metabolic gene content in oomycetes that likely evolved in relation to lifestyle and/or (host) environment.

Next to these dynamics, it is anticipated that certain metabolic functions that are essential for viability (e.g. TCA cycle or glycolysis) are shared as a core set of metabolic enzymes. To investigate the degree of conservation of metabolic capacity between oomycetes, we calculated the core sets (intersection, present in all species) and pan sets (union, present in any species) for metabolic gene families (**Figure 1B**). To obtain the overall set of gene families for reference, we clustered the 882,655 proteins predicted in the 42 oomycete species into 28,463 gene families, which contained 89% of the proteins (**Table S2**). We calculated accumulation curves that show the pan/core size as a function of a random sample of oomycete genomes. As expected, the core size decreases when the sample contains more genomes and the pan size increases. Consequently, the metabolic core in all 42 oomycetes comprises only 146 metabolic gene families (6% of the pan). We observed that the core and pan curves of metabolic gene families flatten later than the curves of the overall gene families, and the core/pan ratio of metabolic gene families is also consistently higher (**Figure 1C**). Thus, even though we observed frequent gains and losses of metabolic gene families, metabolic gene families are generally more conserved within oomycete genomes than gene families overall. Since metabolic enzymes can have a broad range of substrates (i.e. enzyme promiscuity) and non-homologous enzymes can catalyze the same reactions (i.e. isozymes), it is anticipated that biochemical function is generally better conserved than metabolic genes (Carbonell *et al.*, 2011). To validate this, we also calculated the core and pan of the reaction sets (i.e. all biochemical reactions associated with the metabolic enzymes). Indeed, the core/pan ratio was consistently higher for reactions than for metabolic enzymes (**Figure 1C**), corroborating that non-essential (redundant) enzymes are generally less conserved than those playing a central role in metabolic pathways.

To investigate whether oomycetes of the same lineage or with a similar lifestyle have a different degree of conservation of metabolic capacity, we calculated the core and pan sets for subsets of species grouped by their respective taxonomic lineage and lifestyle (**Figure S3**). The *Saprolegniales* have a core of 549 metabolic genes, which is significantly larger than that of eight randomly selected genomes (**Figure 1B**) (average 397±46; $P(X \geq 549)=4.6\times10^{-4}$). It is conceivable that a closely related subset of genomes has a larger core than a random selection of genomes. However, to our surprise the core sets of metabolic genes in the *Pythiaceae* and *Peronosporaceae* (445 and 222 respectively) were not significantly larger than the core sets of randomly selected species ($P(X \geq 445)=0.089$ and $P(X \geq 222)=0.63$ respectively). Even though the overrepresentation of *Peronosporaceae* in our species selection may influence the random genome selection (i.e. the background distribution), our results suggest that metabolic gene content even within a lineage varies considerably. We did not observe a significant difference in core size for any of the lifestyle-groups compared with that of a random sample. For all groups the overall core sets were highly enriched for metabolic enzymes (hypergeometric test; *P*-values ranging from 3.6×10^{-49} for *Saprolegniales* to 7.0×10^{-13} for *Pythiaceae*), in line with

our earlier observation that between species metabolic gene families are more conserved than overall gene families, regardless of lineage or lifestyle (**Table S3**). In addition, all subgroups have a consistently higher core/pan ratio for reactions than for metabolic gene families (two-tailed paired T-test; P=1.0×10^{-4}), in line with the previously observed oomycete-wide pattern (**Figure 1C**). These results make it plausible that oomycetes have a variable metabolism that reflects adaptations to their host and environment (Klein *et al.*, 2012).

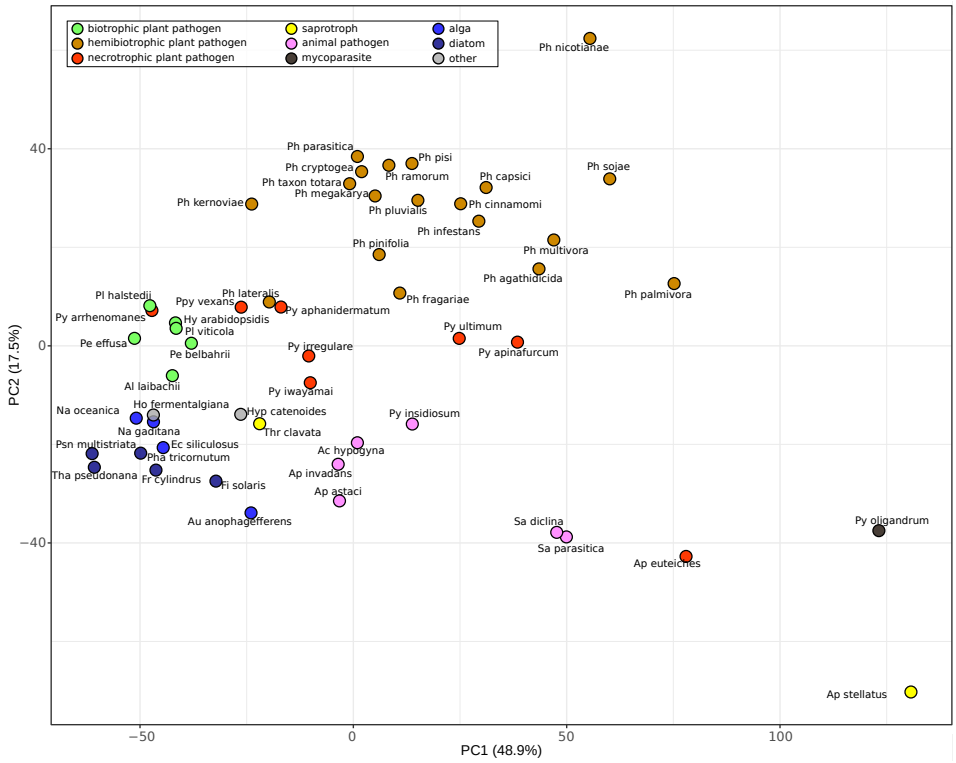


FIGURE 2 | Stramenopiles with a similar lifestyle cluster based on their metabolic reactions. Principal component analysis (PCA) of metabolic reaction sets (i.e. metabolic networks) predicted for each species. The first two principal components are shown, and the percentage of explained variance is indicated. The individual species are colored according to lifestyle (key: top left). *Thalassiosira oceanica* was a strong outlier and is therefore not displayed. Genus names are abbreviated; *Achlya*: Ac, *Albugo*: Al, *Aphanomyces*: Ap, *Aureococcus*: Au, *Ectocarpus*: Ec, *Fistulifera*: Fi, *Fragilariopsis*: Fr, *Hondaea*: Ho, *Hyaloperonospora*: Hy, *Hypochytrium*: Hyp, *Nannochloropsis*: Na, *Peronospora*: Pe, *Phaeodactylum*: Pha, *Phytophthora*: Ph, *Phytopythium*: Ppy, *Plasmopara*: Pl, *Pseudo-nitzschia*: Psn, *Pythium*: Py, *Saprolegnia*: Sa, *Thalassiosira*: Tha, *Thraustrotheca*: Thr.

Metabolic networks reflect differences in lifestyle and host adaptations

To investigate whether or not the metabolic capacity of oomycetes relates to lifestyle, we performed a principal component analysis (PCA) on the metabolic reaction sets of all species (hereinafter referred to as their metabolic networks) (**Figure 2**). The first two

principal components, which accounted for 66.4% of the variance in the data, revealed a clear separation between species with different lifestyles while species with similar lifestyles typically cluster together. Clustering is most prominent for pathogens with an obligate biotrophic lifestyle, even though this lifestyle has evolved several times independently within oomycetes (Thines and Choi, 2015). In the PCA *Pythium* and *Phytophthora* species are more spread, which corroborates the small set of core metabolic genes (**Figure S2**). In addition, most necrotrophic plant pathogenic *Pythium* species cluster, yet the animal pathogen *Pythium insidiosum* is separated and co-localizes with other animal pathogens of the genera *Aphanomyces* and *Achlya*. The sole mycoparasite in our collection, *Pythium oligandrum* is localized distantly from all other (*Pythium*) species. Similarly, the *Aphanomyces* species are clearly separated according to their lifestyle. Collectively, these results highlight that species with similar lifestyles have similar metabolic networks, irrespective of their taxonomic association.

To gain more insight into the conserved and accessory metabolic processes related to specific lineages and lifestyles, we divided the metabolic networks into 91 pathways, based on the classification of reactions by KEGG (**Table S4**). Subsequently, we calculated the pathway coverage, i.e. the proportion of reactions present in a species relative to the set of reactions present in at least one species (**Figure 3**). Visualisation of the pathway coverage revealed a core set of conserved pathways present in all species and accessory pathways with a substantial number of reactions only occurring in few species. The most conserved pathways by average coverage are 'lipoic acid metabolism' (7 reactions), for which 50 out of 54 species have 100% coverage; and 'valine, leucine and isoleucine biosynthesis' (19 reactions), for which 40 species have 100% coverage. Additionally, 'fatty acid degradation' and 'fatty acid elongation' are among the pathways with highest average coverage (95% and 91% respectively). Pathways that have on average the least coverage include among others 'glycosaminoglycan degradation', 'cutin, suberine and wax biosynthesis' and 'retinol biosynthesis'. In general, in the pattern of pathway coverage the oomycetes can be clearly distinguished from the non-oomycetes. This is reflected in a reduced coverage of photosynthesis-related pathways ('carotenoid biosynthesis', 'porphyrin, and chlorophyll metabolism') (Nisar *et al.*, 2015), cell wall (degradation) related pathways ('galactose metabolism'), and pathways related to secondary metabolism and pathogen defense ('glucosinolate biosynthesis', 'terpenoid biosynthesis') (Li *et al.*, 2015). By contrast, pathways that have higher coverage in oomycetes include many lipid-related pathways, such as 'sphingolipid', 'glycerolipid', 'glycerophospholipid', and 'ether lipid' metabolism (63-78% coverage), emphasizing that lipids and fatty acids play a central role in oomycetes (Fernandes *et al.*, 2019; Griffiths *et al.*, 2003b; Rodenburg *et al.*, 2018a; Rodenburg *et al.*, 2019). The *Saprolegniales* differ from other oomycetes as they have a higher coverage for various pathways including thiamine and vitamin B6 metabolism. The most prominent difference is the relatively high coverage of the steroid biosynthesis pathway, which is almost completely absent in other oomycetes (Dahlin *et al.*, 2017). The conservation of various enzymes in these pathways is unique to *Saprolegniales* and contributes to the relatively large core set of metabolic genes in *Saprolegnia* spp. (**Figure S3**). *Phytophthora* and *Pythium* show a higher coverage for the tyrosine, phenylalanine, and phenylpropanoid pathways than other

species. The phenylpropanoids are derivatives of phenylalanine and tyrosine, and in plants phenylpropanoids are strongly associated with pathogen defense and cell-wall formation (Miedes *et al.*, 2014). Recently it has been observed that even the most ancient oomycete-plant interaction triggers a phenylpropanoid defense response, indicating that phenylpropanoids are conserved and involved with the prime defenses against pathogens (Carella *et al.*, 2019). *Phytophthora* spp. have a high coverage of the 'naphthalene degradation' pathway, and this includes an enzyme with salicylate hydroxylase activity (EC 1.14.13.1) that degrades salicylic acid, a compound strongly associated with defense in plants against *Phytophthora* (Halim *et al.*, 2007). Moreover, salicylate precursors in plants were found to interact with oomycete effectors (Liu *et al.*, 2014). These results support our hypothesis that the metabolism of oomycetes, especially *Phytophthora*, is highly adapted to their interaction with plant hosts.

Another distinguishing pattern is observed for the obligate biotrophs, which show reduced coverage of the same or similar metabolic networks despite their evolutionary distance. Hierarchical clustering (UPGMA on pairwise Spearman correlation distances) revealed that metabolic networks of obligate biotrophs have converged by losing the same sets of metabolic enzymes multiple times during adaptation to a biotrophic lifestyle (**Figure 3**). This is reflected in the patterns of 78 of the 91 pathways for which the obligate biotrophs have a lower coverage compared to other oomycetes (one-tailed paired T-test; $P=4.7\times10^{-14}$). We observed the largest differences for steroid biosynthesis and arachidonic acid (ARA) metabolism. Whereas most *Phytophthora* and *Pythium* spp., despite many gene losses retained 10 reactions (27% coverage) (**Figure 3**), obligate biotrophs have lost all but one reaction for sterol biosynthesis and that is phospholipase A2 (EC 3.1.1.4). This enzyme also plays a role in ARA production by desaturation of phosphatidylcholine, a membrane phospholipid that can facilitate energy storage in lipid droplets to fuel sporulation (Chen *et al.*, 2013). It has been proposed that obligate biotrophs have lost the capacity to synthesize ARA because it elicits defense responses in the host (Judelson, 2017; Robinson and Bostock, 2015). The detection of ARA in the biotroph *Plasmopara viticola* suggests that obligate biotrophs derive this compound from host membranes (Negrel *et al.*, 2018; Shanab *et al.*, 2018). Reactions in this pathway downstream of ARA for the conversion into fatty acids were lost by all obligate biotrophs. In summary, obligate biotrophs have lost metabolic enzymes in almost all pathways, yet seem to converge on similar metabolic traits, which suggests a selection pressure on particular metabolic properties.

FIGURE 3 | Core and accessory metabolic pathways in Stramenopiles relate to lifestyle and phylogeny. Heatmap showing pathway coverage per species. The colors of the boxes (key: top left) scale with pathway coverage (percentage of reactions present of the pan-pathway) for each metabolic pathway and in each species in the dataset. The bar plot (right) shows the number of reactions present in at least one species. Heatmap columns are ordered according to a hierarchical clustering of reaction frequency in the full metabolic networks (UPGMA on pairwise Spearman correlation distances). The species tree on top is drawn according to **Figure 1A**. The lines that are colored according to lifestyle (key: top right), connect the tips of the phylogeny with the corresponding column in the heatmap. Species names are abbreviated as described for **Figure 2**.

Metabolic network analyses highlight the differences that characterize biotrophs

To investigate how losses of metabolic enzymes in biotrophs influence the structure of the metabolic network, we transformed each network into a directed graph in which nodes represent compounds and edges represent reactions (see Methods). Thus, central nodes (hubs) represent metabolites that likely play an important role in metabolism as the substrate or product of an array of different reactions. As expected, the most central nodes (by betweenness-centrality) involved primary metabolites such as pyruvate, adenosine monophosphate (AMP), amino acids, and acetyl-CoA, the crucial precursor for the synthesis of fatty acids (Fernandes *et al.*, 2019). By contrasting the networks of obligate biotrophs to those of other plant pathogenic oomycetes in our collection we were unable to detect substantial differences in overall network properties such as average node degree or betweenness-centrality (**Table S5**). However, we observed that networks of obligate biotrophs were generally smaller and thus contain fewer nodes and edges (one-tailed rank-sum test; $P=9.9\times10^{-5}$ and $P=2.8\times10^{-4}$ respectively). To investigate whether these reduced networks have a different capacity to take up nutrients, we calculated for each network the "seed" (Borenstein and Feldman, 2009), which is defined as the minimal set of compounds that organisms need to import to produce (i.e. have a path to) all other compounds in the metabolic network (see Methods). The size of the reduced metabolic networks of biotrophs strongly correlated with the network seed size (Spearman correlation $\rho=0.77$; $P=7.2\times10^{-12}$) (**Figure S4**), which implies a smaller pool of substrates. This is in line with an overall lower number of transporters encoded in the genomes of obligate pathogens compared to free-living organisms, which may relate to highly specialized metabolism (Baxter *et al.*, 2010; Blume and Seeber, 2018; Dean *et al.*, 2014). To identify the specific differences that distinguish metabolic networks of biotrophs from those of other plant pathogenic oomycetes, we collapsed the metabolic networks of both groups to derive a pan-pathogen metabolic network (the union of the necrotroph, hemibiotroph, and biotroph networks) and a pan-biotroph metabolic network (the union of the biotroph networks), which is a subset of the former. We then compared these two networks to visualize the differences, indicating which nodes (compounds) occur in the overall plant pathogen network, but have been lost in the biotrophs (**Figure 4**). The plant pathogen network consists of 1,484 nodes (compounds), of which 931 (63%) are part of the largest graph component (i.e. the largest connected subnetwork). It also contains 205 additional small network components, often composed of only a few compounds and reactions (**Figure S4**). These components predominantly contained pathways that could be associated with plant compounds such as 'drug metabolism', 'naphthalene degradation', and 'terpenoid biosynthesis'. It is therefore conceivable that these components originate from secreted enzymes involved in interactions with the host or the environment (Judelson and Ah-Fong, 2018). In comparison to the plant-pathogen network, the obligate biotroph network lacks 209 compounds, of which 65 are seed compounds and 65 are part of the primary component. The nodes missing in the primary component comprise 18 seeds, including nitrate, nitrite, and 4-methyl-5-(2-phosphooxyethyl) thiazole (a thiamine precursor). This is in line with the type of genes that are lost in obligate biotrophs, as many of these encode proteins with a role in the assimilation pathways for

these compounds (**Figure 3**) (Spanu, 2012). As these compounds are missing from the biotroph metabolic network, an alternative set of seed compounds is introduced, including thiamine monophosphate, fructose, pseudouridine (a nucleoside), and two lysine catabolites, 2-oxoadipic and L-2-aminoadipic acid. It is possible that the (partial) loss of an assimilation pathway drives adaptations to allow the import of alternative compounds to compensate, e.g. through the diversification of transporter substrates (Dean *et al.*, 2018). In addition to the losses in the primary component, the large majority (144/209; 69%) of compound losses were part of the small network components, indicating that enzyme losses of biotrophs have largely occurred in the accessory parts of metabolism (hypergeometric test; P=1.1×10^{-23}). These losses involve many plant-related compounds such as p-coumaric and pipecolic acid, both strongly associated with plant defense (**Table S5**) (Chen *et al.*, 2019; Hartmann *et al.*, 2018). A successful biotrophic infection suppresses host defense responses (Asai *et al.*, 2014), thereby making the degradation of these compounds obsolete. Interestingly, the average betweenness of the lost compounds was almost eight-fold lower than that of all nodes (two-tailed rank-sum test; P=2.2×10^{-16}). Moreover, terminal nodes in the network (i.e. with a degree of 1) were highly enriched for lost compounds (hypergeometric test; P=1.63×10^{-20}) and often co-occurred with other lost compounds (**Figure 4**). Thus, lost compounds in biotrophs are not randomly distributed in the network but are predominantly located at the terminal branches of the network. We suggest that the prolonged symbiosis of obligate pathogens with their host leads to the evolution of 'metabolic shortcuts' as a result of redundant, parallel pathways with the host, leading to a loss of enzymes in the initial steps of the pathway (Albalat and Cañestro, 2016).

Conclusions

The co-evolution of a parasite with its host drives adaptations in the metabolism of both species. Here, we predicted and compared the metabolic capacity of oomycetes, a class mainly consisting of plant and animal pathogens, and with distinct lifestyles. We found that the metabolic gene repertoire among oomycetes varies widely, especially among the plant-pathogenic species within the *Peronosporacaea* lineage, and this indicates strong dynamics and selection pressure on metabolic gene content. When the metabolic networks of these species were divided into a core and an accessory part, the latter showed associations with plant compounds. This raises the hypothesis that the pathogens produce enzymes that use host components as substrate. A way to test this would be to grow axenic cultures on minimal medium enriched with particular host substrates and analyze the culture filtrate for the presence of the respective enzymes and/or metabolites (Meijer *et al.*, 2014). However, for obligate biotrophs such an approach is not feasible as they can not be cultured. The obligate biotrophs showed a profound reduction of coverage of almost all pathways, and in particular in their accessory metabolism. These reductions were most prominent at terminal branches in the network, suggesting that the metabolic networks of obligate biotrophs contract from the periphery to the core. It would be of interest to compare these networks to those of their

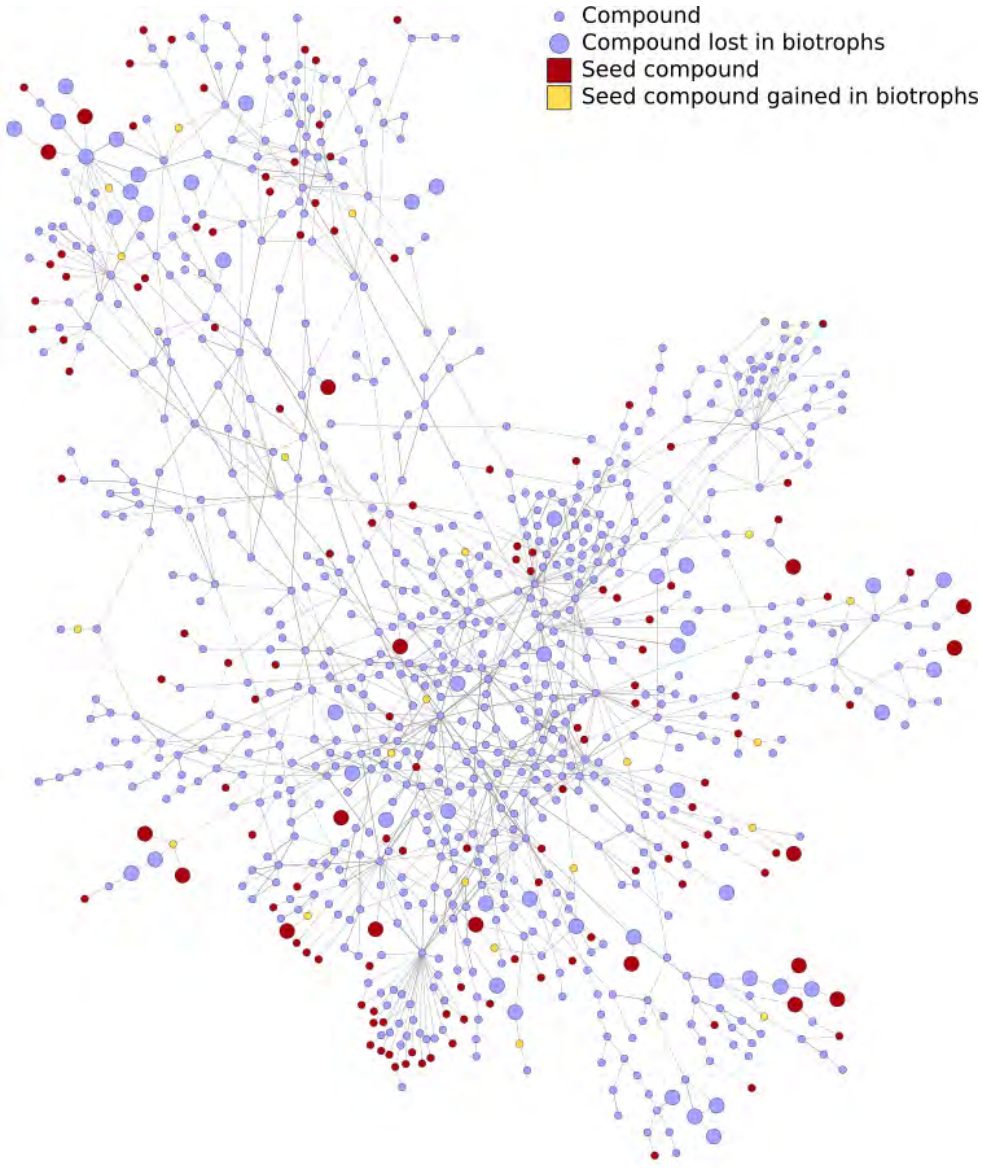


FIGURE 4 | Lost reactions in biotrophs are mainly located at terminal branches of the metabolic network. A graph representation of the main component of the metabolic network of plant pathogenic oomycetes, overlaid with the metabolic network of solely biotrophs. Nodes are compounds, edges are (partial) reactions. Larger nodes highlight nodes absent (lost) in the pan-network of biotrophs. Red nodes represent the network seed nodes, and green nodes network seed nodes that are gained in biotrophs. For the complete pan-metabolic network, including smaller components, see **Figure S5**.

hosts to investigate to what extent gene losses in obligate biotrophs are compensated via the metabolic networks of their hosts, and whether the lost enzymes are indeed dispensable (Duan *et al.*, 2013; Levy *et al.*, 2015). The detailed dissection of metabolic gene content, pathway coverage, and network topology presented in this study, will serve as a knowledge base to better understand metabolic adaptations in oomycetes in relation to lifestyle, host, and environment.

Methods

Species selection and proteome annotation

Genomes and annotation files were selected from literature and downloaded from FungiDB, NCBI Genome, Ensembl, aphanoDB or https://github.com/oomycetes/oomycetes.github.io (**Table S1**) (McGowan and Fitzpatrick, 2017). We only included genomes for which gene annotation was publicly available (early 2019). We excluded *Albugo candida* and *Phytophthora rubi* as we considered their genome assemblies and annotations of insufficient quality, showing patterns of contamination. Since available gene annotations varied in levels of detail (e.g. inclusion of UTRs or transcript isoforms), we only retained the longest coding sequences per gene. Predicted protein sequences were clustered into ortholog clusters (orthogroups) using OrthoFinder (v2.3.3) (Emms and Kelly, 2019) with DIAMOND (v0.9.24.125) for homology searches (Buchfink *et al.*, 2015), MAFFT v7.310 (Katoh and Standley, 2013) for multiple sequence alignments, and FastTree (v2.1.10) (Price *et al.*, 2010) to infer the maximum likelihood gene tree for each orthogroup. The species tree was inferred as part of the Orthofinder pipeline using the STAG algorithm and rooted by STRIDE (Emms and Kelly, 2017; Emms and Kelly, 2019).

The predicted protein sequences were annotated based on their orthology with metabolic enzymes using KofamScan (v1.0.0) (Aramaki *et al.*, 2019), which matches query protein sequences to pre-computed profile-hidden Markov models of KEGG Ortholog clusters. KofamScan uses an adaptive score threshold, heuristically determined by matching positive and negative training data to the KO HMM, and selecting the threshold with maximal F1-measure (accuracy).

Since our analyses depend on the completeness of the genome annotations, we performed additional screenings to identify missing gene annotations in the genomes. In brief, we identified missing KOs based on the presence of these KOs in neighboring species, and subsequently queried the genome for these missing KOs. We extracted all possible subtrees from the species tree and selected those containing three or more species. Subsequently, we iterated over all subtrees from small to large, and identified 'annotation gaps', i.e. KOs that were missing in at least one species and present in more than 50% of the remaining species (thus at least 2/3 for a minimal subtree of three species). For each annotation gap, we considered a target species (which contains the annotation gap), and a source species (its closest relative

having the KO). The protein sequence of the KO in the source species was then aligned to the target genome, masked for its original genome annotations. Alignments were performed using Exonerate (v2.2.0) (Slater and Birney, 2005) with the protein2genome model (default settings; maximum intron size of 20 kb) and only considering the (single) best alignment as candidate KO. To prevent spurious additions of KOs, various filters were applied. First of all, alignments were filtered for those containing internal stop codons. Secondly, we removed alignments that matched the query protein for less than 90% (query coverage). Thirdly, only alignments were used for which the sequence identity and alignment length was above the HSSP-curve (offset +5), which relates sequence identity to alignment length (**Figure S6**) (Rost, 1999). Singletons, i.e. KOs present in only a single species, were removed from the dataset to prevent the influence of contamination.

Evolutionary analyses

To reconstruct a phylogenetic tree from the predicted species tree, we synchronized the nodes of the species tree with the estimated divergence times of Stramenopiles, as predicted from a relaxed molecular clock model (Matari and Blair, 2014) accounting for variable evolutionary rates on different branches of the tree. Calibration was performed using the *chronos* function of the Ape (v5.3) package in R (Kim and Sanderson, 2008; Paradis and Schliep, 2019), which uses penalized likelihood to calculate the divergence times of all nodes in a tree based on given branch lengths and calibration points. We used the estimated confidence intervals of the calibration points as minimum and maximum bounds for the likelihood function (Matari and Blair, 2014). Subsequently, we used the *ace* function of the Ape package to reconstruct the ancestral states of the metabolic KOs using maximum likelihood, and subsequently inferred gene family gains or losses for each branch when either parent or child node had a gene family size of 0.

Pan and core genomes

The pan and core accumulation curves were determined for oomycetes using a custom R script, iteratively sampling *n* species in 1,000 permutations, where *n* ranged from 1 to 42. Subsequently, for all associated genes/reactions, the core sets (present in all species) and pan sets (present in any species) were determined.

Pathway analyses

To quantify pathway completeness for each species, we mapped reactions on KEGG pathway maps (https://www.kegg.jp/). Pathway coverage was calculated as the set of uniquely mapping reactions for a species, expressed as a fraction of the union reaction set of all species. We pruned 44 pathways with five or less reactions in the pan-pathway to reduce noise and emphasize larger pathways. The metabolic networks were clustered according to the pairwise Spearman correlation distance (1-ρ) of the reaction frequency table using UPGMA (**Table S4**).

Graph analyses

Graph analyses were performed using the R package *igraph* (v1.2.4.2) (Csardi and Nepusz, 2006). Metabolic networks were converted into directed graphs using the KEGG RCLASS database. This database contains for each reaction in KEGG the pairs of substrates and products that form the metabolic reaction. Reactions with multiple products or substrates are decomposed into sub-reactions, which helps to eliminate redundant edges. For instance, a graph representation of the reaction *(S)-Lactate + NAD+ <=> Pyruvate + NADH + H+* would connect all substrates with all products, introducing 6 edges. Since many reactions use NADH, this metabolite will be hyper-connected in the graph, leading to connections of many unrelated reactions. Using RCLASS annotation, we can subdivide this reaction into *(S)-Lactate <=> Pyruvate* and *NAD+ <=> NADH*. Many other reactions have the same sub-reaction of NAD^+ to NADH conversion, for which we do not introduce new edges in the graph. Graphs were visualized using Gephi (v0.9.2) (Bastian *et al.*, 2009). The network seeds were calculated using the NetSeed algorithm (Borenstein and Feldman, 2009), and only network seeds with a confidence score of 1 were used.

Acknowledgements

This work was funded by the Food-for-Thought campaign from Wageningen University Fund. We thank Jamie McGowan and David Fitzpatrick of Maynooth University for kindly sharing the oomycete annotation files upon request.

Supplemental Information

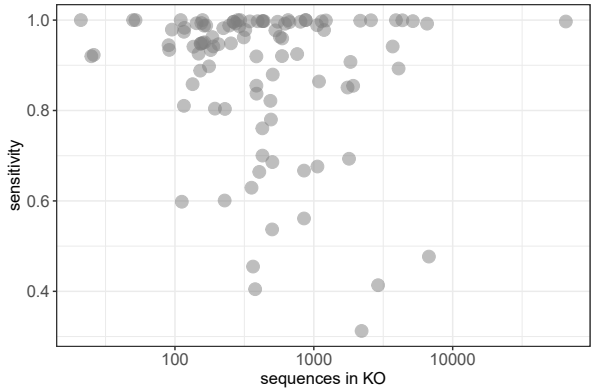


FIGURE S1 | Sensitivity of KofamScan (Aramaki *et al.*, 2019) in assigning the protein sequences of 100 randomly selected KEGG ortholog clusters to their respective HMMs: fraction of proteins assigned to their respective KO as a function of the size (number of sequences) of the KO.

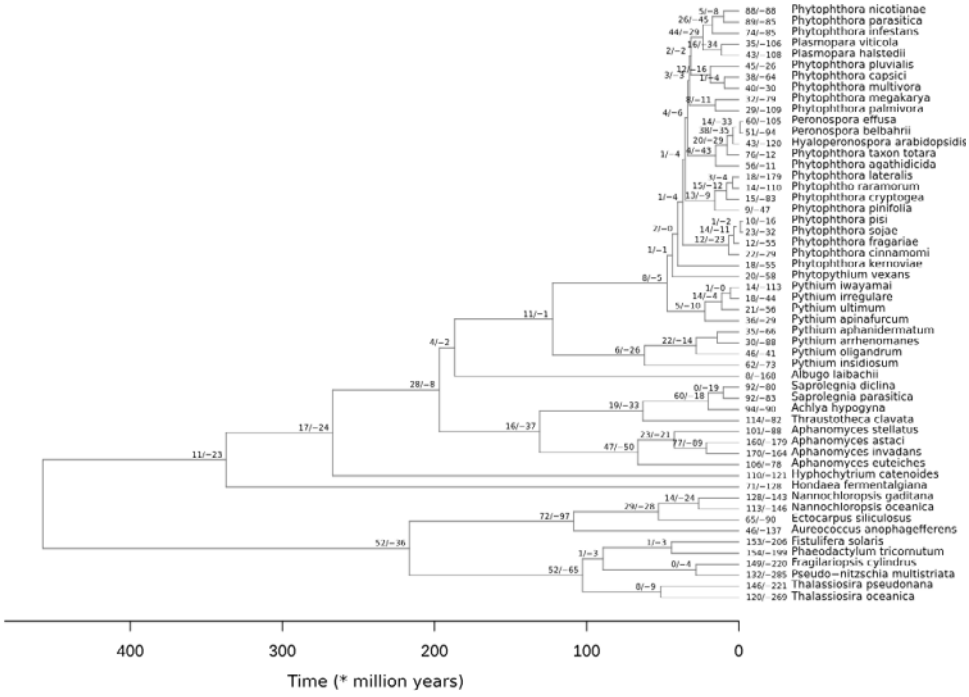


FIGURE S2 | Phylogenetic tree synchronized according to the molecular clock model (Matari and Blair, 2014). Each branch displays the metabolic gene family gains and losses as inferred by maximum likelihood analyses (see **Methods**).

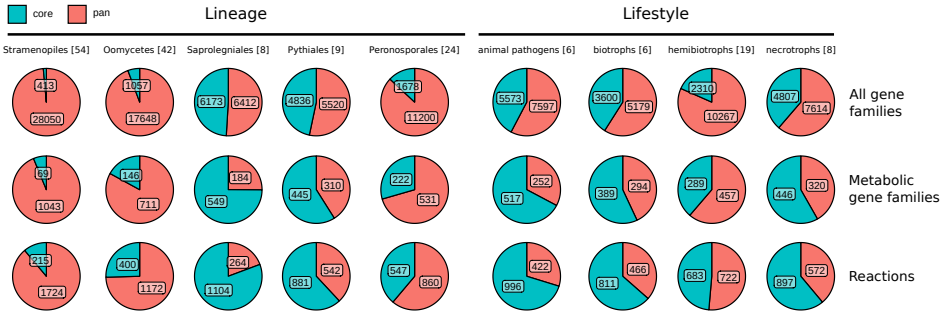


FIGURE S3 | Pie charts showing the pan (red) and core (blue) sizes of (metabolic) gene families and reaction sets in multiple taxonomic groups and lineages. The areas are overlapping, i.e. the total pie corresponds to the size of the pan-set, of which the core is a subset. The numbers in square brackets indicate the number of species in the respective group.

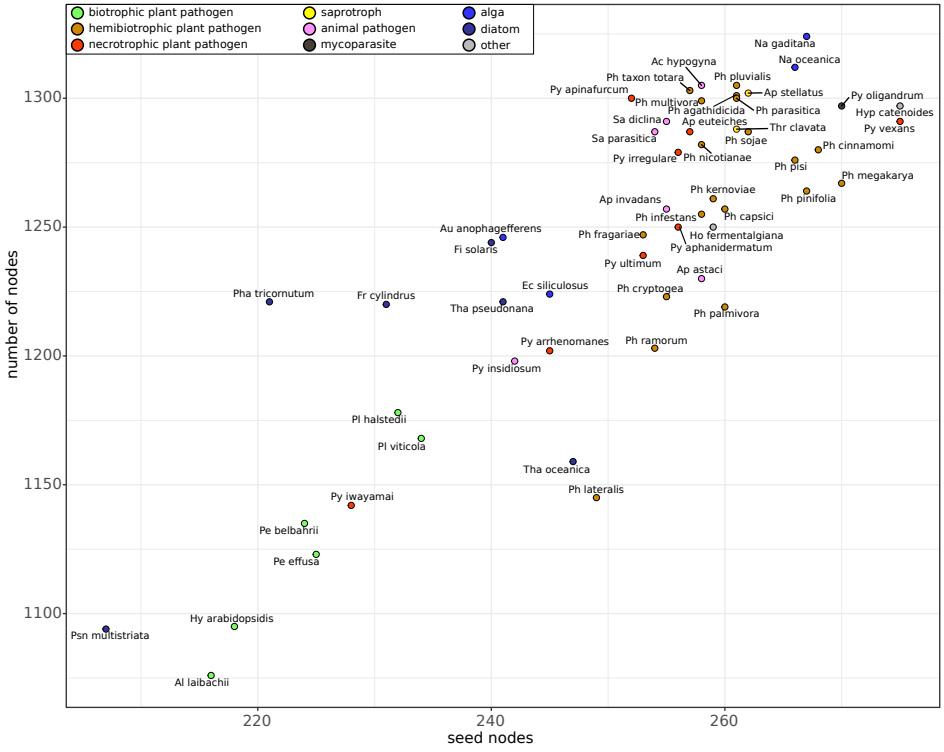


FIGURE S4 | Size of the metabolic network of each species in the dataset in number of nodes, related to the size of the network seed. Colors indicate the lifestyle of each species. Species names are abbreviated as described for **Figure 2**.

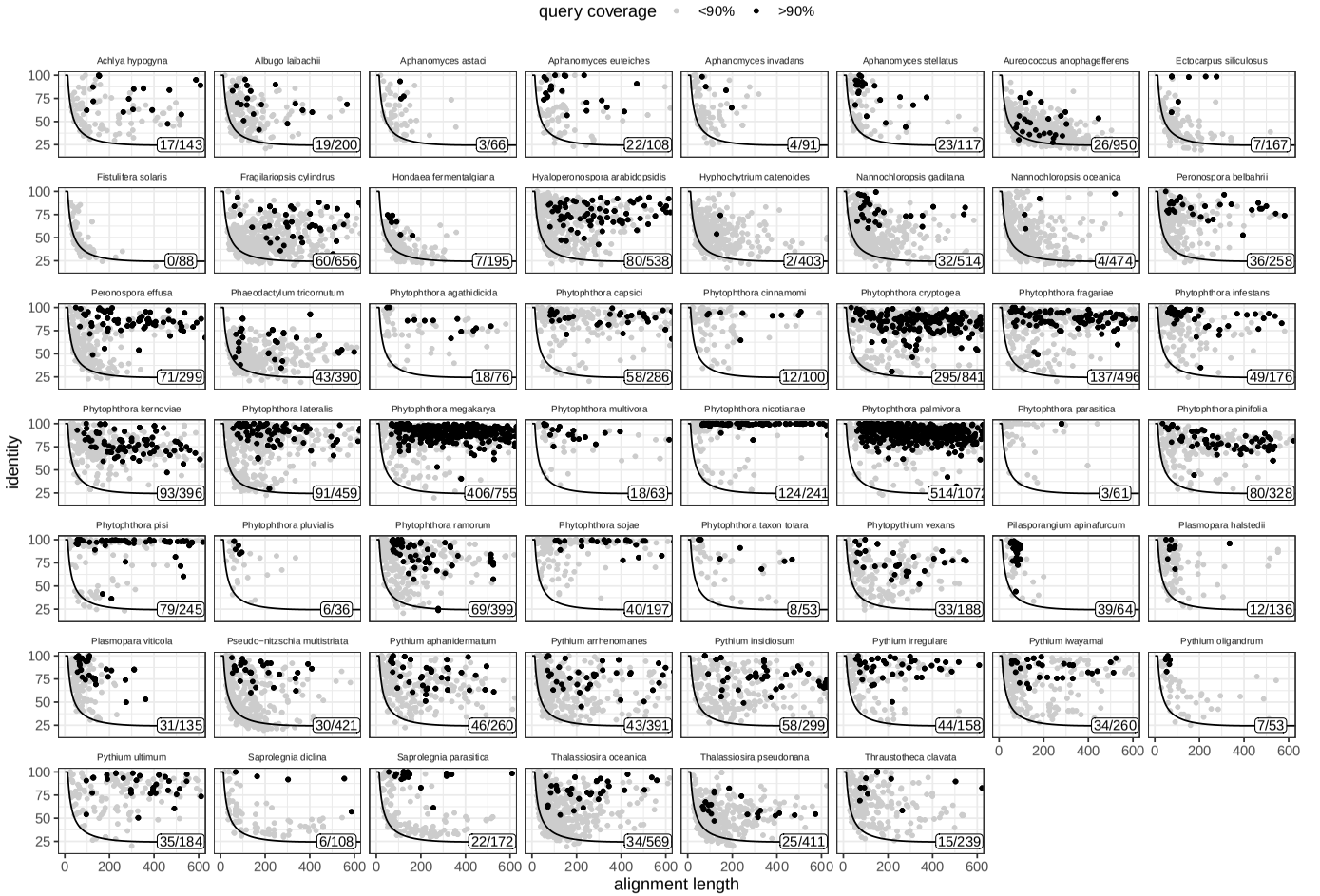


FIGURE S6 | Alignment results of the genome, screening for potentially missing genes in each species. The X-axis shows the alignment length and the Y-axis shows the sequence identity. The curve displays the HSSP-curve which was used as a cutoff for filtering. Black dots represent alignments with a query coverage >= 90%. The boxes in the bottom-right corner indicate the number of alignments passing all filters, and the number of total alignments.

TABLE S1A | Species selected for this study, with data sources and literature references.

Species	Lifestyle/Lineage	References	Download source
Achlya hypogyna	saprotroph	1	NCBI Genome
Albugo laibachii	biotroph	2	Ensembl protists
Aphanomyces astaci	animal pathogen	3	Ensembl protists
Aphanomyces euteiches	nectroph		aphanoDB
Aphanomyces invadans	animal pathogen	3	Ensembl protists
Aphanomyces stellatus	saprotroph	3	aphanoDB
Aureococcus anophagefferens	alga	4	NCBI Genome
Ectocarpus siliculosus	alga	5	NCBI Genome
Fistulifera solaris	diatom	6	NCBI Genome
Fragilariopsis cylindrus	diatom	7	NCBI Genome
Hondaea fermentalgiana	chytrid		NCBI Genome
Hyaloperonospora arabidopsis	biotroph	8	Ensembl protists
Hyphochytrium catenoides	chytrid	9	https://github.com/guyleonard/hyphochytrium
Nannochloropsis gaditana	alga	10	NCBI Genome
Nannochloropsis oceanica	alga	11	NCBI Genome
Peronospora effusa	biotroph	12	NCBI Genome
Peronospora belbahrii	biotroph	13	NCBI Genome
Phaeodactylum tricornutum	diatom	14	NCBI Genome
Phytophthora agathidicida	hemibiotroph	15	NCBI Genome
Phytophthora capsici	hemibiotroph	16	FungiDB
Phytophthora cinnamomi	hemibiotroph		FungiDB
Phytophthora cryptogea	hemibiotroph	17, 18	NCBI Genome*
Phytophthora fragariae	hemibiotroph	19, 18	NCBI Genome*
Phytophthora infestans	hemibiotroph	20	Ensembl protists
Phytophthora kernoviae	hemibiotroph	21	Ensembl protists
Phytophthora lateralis	hemibiotroph	22	Ensembl protists
Phytophthora megakarya	hemibiotroph	23	NCBI Genome
Phytophthora multivora	hemibiotroph	15, 18	NCBI Genome*
Phytophthora nicotianae	hemibiotroph	24	Ensembl protists
Phytophthora palmivora	hemibiotroph	23	NCBI Genome
Phytophthora parasitica	hemibiotroph		Ensembl protists
Phytophthora pinifolia	hemibiotroph	17, 18	NCBI Genome*
Phytophthora pisi	hemibiotroph	18	NCBI Genome
Phytophthora pluvialis	hemibiotroph	15, 18	NCBI Genome*
Phytophthora ramorum	hemibiotroph	25	Ensembl protists
Phytophthora sojae	hemibiotroph	25	Ensembl protists
Phytophthora taxon totara	hemibiotroph	15, 18	NCBI Genome*
Phytopythium vexans	necrotroph	26	FungiDB
Plasmopara viticola	biotroph	27	NCBI Genome
Plasmopara halstedii	biotroph	28	Ensembl protists
Pseudo-nitzschia multistriata	diatom	29	http://bioinfo.szn.it/pmultistriata/

Species	Lifestyle/Lineage	References	Download source
Pythium apinafurcum	necrotroph	30, 18	NCBI Genome*
Pythium aphanidermatum	necrotroph	26, 18	NCBI Genome*
Pythium arrhenomanes	necrotroph	26	Ensembl protists
Pythium insidiosum	animal pathogen	31	NCBI Genome
Pythium irregulare	necrotroph	26	FungiDB
Pythium iwayami	necrotroph	26	FungiDB
Pythium oligandrum	mycoparasite	32, 18	NCBI Genome*
Pythium ultimum	necrotroph	33	FungiDB
Saprolegnia diclina	animal pathogen		Ensembl genomes
Saprolegnia parasitica	animal pathogen	34	Ensembl genomes
Thalassiosira oceanica	diatom	35	NCBI Genome
Thalassiosira pseudonana	diatom	36	NCBI Genome
Thraustrotheca clavata	saprotroph	1	NCBI Genome

* Gene annotations derived from https://github.com/oomycetes/oomycetes.github.io (ref 18)

TABLE S1B | Bibliography for the data used in this study.

Key	Citation
1	Misner I, Blouin N, Leonard G, Richards TA, Lane CE. 2014. 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.
2	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:e1001094.
3	Gaulin E, Pel MJC, Camborde L, San-Clemente H, Courbier S, Dupouy MA, Lengellé J, Veyssiere M, Le Ru A, Grandjean F, Cordaux R, Moumen B, Gilbert C, Cano LM, Aury JM, 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.
4	Frischkorn KR, Harke MJ, Gobler CJ, Dyhrman ST. 2014. De novo assembly of *Aureococcus anophagefferens* transcriptomes reveals diverse responses to the low nutrient and low light conditions present during blooms. Front Microbiol 5:375.
5	Cock JM, Sterck L, Rouzé P, Scornet D, Allen AE, Amoutzias G, Anthouard V, Artiguenave F, Aury J-M, Badger JH, Beszteri B, Billiau K, Bonnet E, Bothwell JH, Bowler C, Boyen C, Brownlee C, Carrano CJ, Charrier B, Cho GY, Coelho SM, Collén J, Corre E, Silva CD, Delage L, Delaroque N, Dittami SM, Doulbeau S, Elias M, Farnham G, Gachon CMM, Gschloessl B, Heesch S, Jabbari K, Jubin C, Kawai H, Kimura K, Kloareg B, Küpper FC, Lang D, Bail AL, Leblanc C, Lerouge P, Lohr M, Lopez PJ, Martens C, Maumus F, Michel G, Miranda-Saavedra D, Morales J, Moreau H, Motomura T, Nagasato C, Napoli CA, Nelson DR, Nyvall-Collén P, Peters AF, Pommier C, Potin P, Poulain J, Quesneville H, Read B, Rensing SA, Ritter A, Rousvoal S, Samanta M, Samson G, Schroeder DC, Ségurens B, Strittmatter M, Tonon T, Tregear JW, Valentin K, Dassow P von, Yamagishi T, Peer YV de, Wincker P. 2010. The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. Nature 465:617–621.
6	Tanaka T, Maeda Y, Veluchamy A, Tanaka M, Abida H, Maréchal E, Bowler C, Muto M, Sunaga Y, Tanaka M, Yoshino T, Taniguchi T, Fukuda Y, Nemoto M, Matsumoto M, Wong PS, Aburatani S, Fujibuchi W. 2015. Oil accumulation by the oleaginous diatom *Fistulifera solaris* as revealed by the genome and transcriptome. Plant Cell 27:162–176.

Key	Citation
7	Mock T, Otillar RP, Strauss J, McMullan M, Paajanen P, Schmutz J, Salamov A, Sanges R, Toseland A, Ward BJ, Allen AE, Dupont CL, Frickenhaus S, Maumus F, Veluchamy A, Wu T, Barry KW, Falciatore A, Ferrante MI, Fortunato AE, Glöckner G, Gruber A, Hipkin R, Janech MG, Kroth PG, Leese F, Lindquist EA, Lyon BR, Martin J, Mayer C, Parker M, Quesneville H, Raymond JA, Uhlig C, Valas RE, Valentin KU, Worden AZ, Armbrust EV, Clark MD, Bowler C, Green BR, Moulton V, van Oosterhout C, Grigoriev IV. 2017. Evolutionary genomics of the cold-adapted diatom *Fragilariopsis cylindrus*. Nature 541:536–540.
8	Baxter L, Tripathy S, Ishaque N, Boot N, Cabral A, Kemen E, Thines M, Ah-Fong AM, Anderson RG, 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 HS, Kamoun S, Kyung K, Meijer HJG, Minx P, Morris PF, Nelson J, Phuntumart V, Qutob D, Rehmany A, Rougon-Cardoso A, Ryden P, Torto-Alalibo TA, Studholme DJ, Wang Y, Win J, Wood J, Clifton SW, Rogers J, den Ackerveken G, Jones JDG, McDowell JM, Beynon JL, Tyler BM. 2010. Signatures of adaptation to obligate biotrophy in the *Hyaloperonospora arabidopsidis* genome. Science 330:1549–1551.
9	Leonard G, Labarre A, Milner DS, Monier A, Soanes D, Wideman JG, Maguire F, Stevens S, Sain D, Grau-Bové X, Sebé-Pedrós A, Stajich JE, Paszkiewicz K, Brown MW, Hall N, Wickstead B, Richards TA. 2018. Comparative genomic analysis of the 'pseudofungus' *Hyphochytrium catenoides*. Open Biol 8.
10	Radakovits R, Jinkerson RE, Fuerstenberg SI, Tae H, Settlage RE, Boore JL, Posewitz MC. 2012. Draft genome sequence and genetic transformation of the oleaginous alga *Nannochloropsis gaditana*. Nat Commun 3:1–11.
11	Wang D, Ning K, Li J, Hu J, Han D, Wang H, Zeng X, Jing X, Zhou Q, Su X, Chang X, Wang A, Wang W, Jia J, Wei L, Xin Y, Qiao Y, Huang R, Chen J, Han B, Yoon K, Hill RT, Zohar Y, Chen F, Hu Q, Xu J. 2014. *Nannochloropsis* genomes reveal evolution of microalgal oleaginous traits. PLoS Genet 10:e1004094.
12	Feng C, Lamour KH, Bluhm BH, Sharma S, Shrestha S, Dhillon BDS, Correll JC. 2018. Genome sequences of three races of *Peronospora effusa*: a resource for studying the evolution of the spinach downy mildew pathogen. Mol Plant-Microbe Interact MPMI 31:1230–1231.
13	Thines M, Sharma R, Rodenburg SYA, Gogleva A, Judelson HS, Xia X, van den Hoogen J, Kitner M, Klein J, Neilen M, de Ridder D, Seidl MF, den Ackerveken GV, Govers F, Schornack S, Studholme DJ. 2019. The genome of *Peronospora belbahrii* reveals high heterozygosity, a low number of canonical effectors and CT-rich promoters. bioRxiv 721027.
14	Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A, Maheswari U, Martens C, Maumus F, Otillar RP, Rayko E, Salamov A, Vandepoele K, Beszteri B, Gruber A, Heijde M, Katinka M, Mock T, Valentin K, Verret F, Berges JA, Brownlee C, Cadoret J-P, Chiovitti A, Choi CJ, Coesel S, Martino AD, Detter JC, Durkin C, Falciatore A, Fournet J, Haruta M, Huysman MJJ, Jenkins BD, Jiroutova K, Jorgensen RE, Joubert Y, Kaplan A, Kröger N, Kroth PG, Roche JL, Lindquist E, Lommer M, Martin–Jézéquel V, Lopez PJ, Lucas S, Mangogna M, McGinnis K, Medlin LK, Montsant A, Secq M-PO, Napoli C, Obornik M, Parker MS, Petit J-L, Porcel BM, Poulsen N, Robison M, Rychlewski L, Rynearson TA, Schmutz J, Shapiro H, Siaut M, Stanley M, Sussman MR, Taylor AR, Vardi A, Dassow P von, Vyverman W, Willis A, Wyrwicz LS, Rokhsar DS, Weissenbach J, Armbrust EV, Green BR, Peer YV de, Grigoriev IV. 2008. The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. Nature 456:239–244.
15	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.
16	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 AA, 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 IV, 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.

Key	Citation
17	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.
18	McGowan J, Fitzpatrick DA. 2017. Genomic, network, and phylogenetic analysis of the oomycete effector arsenal. mSphere 2:e00408-17.
19	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.
20	Haas BJ, Kamoun S, Zody MC, Jiang RHY, Handsaker RE, Cano LM, Grabherr M, Kodira CD, Raffaele S, Torto-Alalibo TA, Bozkurt TO, Ah-Fong AM, Alvarado L, Anderson VL, Armstrong MR, Avrova AO, Baxter L, Beynon JL, 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, Grunwald NJ, Horn K, Horner NR, Hu CH, Huitema E, Jeong DH, Jones AME, Jones JDG, Jones RW, Karlsson EK, Kunjeti SG, Lamour KH, 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 WE, 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.
21	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.
22	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.
23	Ali SS, Shao J, Lary DJ, Kronmiller BA, Shen D, Strem MD, Amoako-Attah I, Akrofi AY, Begoude BAD, ten Hoopen GM, Coulibaly K, Kebe BI, Melnick RL, Guiltinan MJ, Tyler BM, Meinhardt LW, Bailey BA. 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.
24	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.
25	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 MK, 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 IV, Rokhsar DS, Boore JL. 2006. *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. Science 313:1261–1266.
26	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:e75072.
27	Dussert Y, Gouzy J, Richart-Cervera S, Mazet ID, Delière L, Couture C, Legrand L, Piron M-C, Mestre P, Delmotte F. 2016. Draft genome sequence of *Plasmopara viticola*, the grapevine downy mildew pathogen. Genome Announc 4.
28	Sharma R, Xia X, Cano LM, Evangelisti E, Kemen E, Judelson HS, 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.

Key	Citation
29	Basu S, Patil S, Mapleson D, Russo MT, Vitale L, Fevola C, Maumus F, Casotti R, Mock T, Caccamo M, Montresor M, Sanges R, Ferrante MI. 2017. Finding a partner in the ocean: molecular and evolutionary bases of the response to sexual cues in a planktonic diatom. New Phytol 215:140–156.
30	Uzuhashi S, Endoh R, Manabe R, Ohkuma M. 2017. Draft genome sequences of the oomycete *Pilasporangium apinafurcum* strains jcm 30513 and jcm 30514, formerly classified as pythium apinafurcum. Genome Announc 5.
31	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.
32	Berger H, Yacoub A, Gerbore J, Grizard D, Rey P, Sessitsch A, Compant S. 2016. Draft genome sequence of biocontrol agent *Pythium oligandrum* strain P037, an oomycota. Genome Announc 4.
33	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 CMM, Gaulin E, Govers F, Grenville-Briggs LJ, Horner NR, Hostetler J, Jiang RHY, Johnson J, Krajaejun T, Lin H, Meijer HJG, Moore B, Morris PF, Phuntumart 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.
34	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 PF, 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.
35	Lommer M, Specht M, Roy A-S, Kraemer L, Andreson R, Gutowska MA, Wolf J, Bergner SV, Schilhabel MB, Klostermeier UC, Beiko RG, Rosenstiel P, Hippler M, LaRoche J. 2012. Genome and low-iron response of an oceanic diatom adapted to chronic iron limitation. Genome Biol 13:R66.
36	Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH, Zhou S, Allen AE, Apt KE, Bechner M, Brzezinski MA, Chaal BK, Chiovitti A, Davis AK, Demarest MS, Detter JC, Glavina T, Goodstein D, Hadi MZ, Hellsten U, Hildebrand M, Jenkins BD, Jurka J, Kapitonov VV, Kröger N, Lau WWY, Lane TW, Larimer FW, Lippmeier JC, Lucas S, Medina M, Montsant A, Obornik M, Parker MS, Palenik B, Pazour GJ, Richardson PM, Rynearson TA, Saito MA, Schwartz DC, Thamatrakoln K, Valentin K, Vardi A, Wilkerson FP, Rokhsar DS. 2004. The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. Science 306:79–86.

TABLE S3 | Pan and core sizes for orthogroups, metabolic enzymes, and reactions for different groups of species based on lineage and lifestyle.

	Lineages					**Lifestyles**			
	Stramenopiles	Oomycetes	*Saprolegniales*	*Pythiales*	*Peronosporales*	necrotrophs	hemibiotrophs	biotrophs	animal
Number of species (n)	54	42	8	9	24	8	19	6	6
	Orthogroups								
Pan	28463	18705	12585	10356	12878	12421	12577	8779	13170
Core	413	1057	6173	4836	1678	4807	2310	3600	5573
Ratio core/pan	0.01	0.06	0.49	0.47	0.13	0.39	0.18	0.41	0.42
Core size from curve at x=n		1057	3566±578	3351±531	1760±244	3566±578	2103±320	4089±643	4089±643
P(X>x)			3.21E-06	2.59E-03	6.31E-01	1.59E-02	2.59E-01	7.76E-01	1.05E-02
	Metabolic Enzymes								
Pan	1112	857	733	755	753	766	746	683	769
Core	69	146	549	445	222	446	289	389	517
Ratio core/pan	0.06	0.17	0.75	0.59	0.29	0.58	0.39	0.57	0.67
Core size from curve at x=n		146	397±46	381±46	228±27	397±46	267±34	437±49	437±49
P(X>x)			4.64E-04	7.91E-02	5.88E-01	1.42E-01	2.57E-01	8.36E-01	5.01E-02
	Reactions								
Pan	1939	1572	1368	1423	1407	1469	1405	1277	1418
Core	215	400	1104	881	547	897	683	811	996
Ratio core/pan	0.11	0.25	0.81	0.62	0.39	0.61	0.49	0.64	0.70
Core size from curve at x=n		400	824±66	793±67	548±42	824±66	609±51	888±70	888±70
P(X>x)			1.07E-05	9.53E-02	5.13E-01	1.35E-01	7.31E-02	8.63E-01	6.13E-02
hypergeometric test: core ratio of metabolic genes vs. overall core ratio	9.15E-25	1.22E-34	3.61E-49	7.00E-13	4.83E-35	7.84E-30	8.56E-42	8.52E-19	6.28E-47

TABLE S5 | Graph properties of metabolic networks and seed compounds with seed scores.

species	nodes	edges	seeds	average degree	average outdegree	average indegree	average betweenness	components
Thalassiosira oceanica	1159	1480	247	2.55	1.28	1.28	1153.13	189
Thalassiosira pseudonana	1221	1631	241	2.67	1.34	1.34	1739.42	182
Pseudo-nitzschia multistriata	1094	1420	207	2.60	1.30	1.30	1191.44	158
Fragilariopsis cylindrus	1220	1631	231	2.67	1.34	1.34	1216.17	180
Phaeodactylum tricornutum	1221	1687	221	2.76	1.38	1.38	1531.29	162
Fistulifera solaris	1244	1683	240	2.71	1.35	1.35	1670.03	177
Aureococcus anophagefferens	1246	1616	241	2.59	1.30	1.30	811.20	188
Ectocarpus siliculosus	1224	1622	245	2.65	1.33	1.33	1788.25	185
Nannochloropsis oceanica	1312	1767	266	2.69	1.35	1.35	1676.67	191
Nannochloropsis gaditana	1324	1773	267	2.68	1.34	1.34	1421.51	200
Hondaea fermentalgiana	1250	1723	259	2.76	1.38	1.38	1844.75	171
Hyphochytrium catenoides	1297	1742	275	2.69	1.34	1.34	2134.67	200
Aphanomyces euteiches	1287	1806	257	2.81	1.40	1.40	2166.47	168
Aphanomyces invadans	1257	1765	255	2.81	1.40	1.40	2033.27	166
Aphanomyces astaci	1230	1721	258	2.80	1.40	1.40	1901.25	157
Aphanomyces stellatus	1302	1812	262	2.78	1.39	1.39	2026.90	172
Thraustotheca clavata	1288	1776	261	2.76	1.38	1.38	1682.13	176
Achlya hypogyna	1305	1832	258	2.81	1.40	1.40	2245.35	174
Saprolegnia parasitica	1287	1813	254	2.82	1.41	1.41	1966.62	171
Saprolegnia diclina	1291	1812	255	2.81	1.40	1.40	2139.75	171
Albugo laibachii	1076	1504	216	2.80	1.40	1.40	1788.82	151
Pythium insidiosum	1198	1655	242	2.76	1.38	1.38	1827.09	174
Pythium oligandrum	1297	1813	270	2.80	1.40	1.40	2084.31	186
Pythium arrhenomanes	1202	1646	245	2.74	1.37	1.37	1846.30	169
Pythium aphanidermatum	1250	1730	256	2.77	1.38	1.38	1746.09	182
Pythium apinafurcum	1300	1848	252	2.84	1.42	1.42	2048.66	172
Pythium ultimum	1239	1740	253	2.81	1.40	1.40	1803.71	170
Pythium irregulare	1279	1808	256	2.83	1.41	1.41	1765.37	176
Pythium iwayamai	1142	1586	228	2.78	1.39	1.39	1463.16	153
Phytopythium vexans	1291	1772	275	2.75	1.37	1.37	1806.21	193
Phytophthora kernoviae	1261	1744	259	2.77	1.38	1.38	1923.95	184
Phytophthora cinnamomi	1280	1769	268	2.76	1.38	1.38	1547.28	186
Phytophthora fragariae	1247	1726	253	2.77	1.38	1.38	1908.28	182

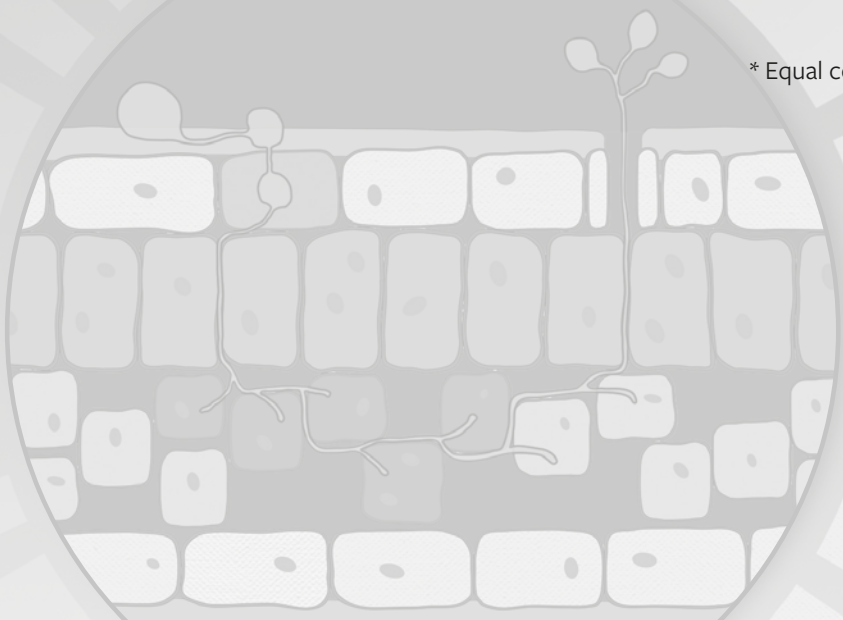
species	nodes	edges	seeds	average degree	average outdegree	average indegree	average betweenness	components
Phytophthora sojae	1287	1798	262	2.79	1.40	1.40	1975.53	184
Phytophthora pisi	1276	1765	266	2.77	1.38	1.38	1789.70	185
Phytophthora pinifolia	1264	1739	267	2.75	1.38	1.38	1673.65	188
Phytophthora cryptogea	1223	1685	255	2.76	1.38	1.38	1740.67	179
Phytophthora ramorum	1203	1669	254	2.77	1.39	1.39	1732.25	174
Phytophthora lateralis	1145	1503	249	2.63	1.31	1.31	999.60	180
Phytophthora agathidicida	1301	1833	261	2.82	1.41	1.41	1996.28	181
Phytophthora taxon totara	1303	1839	257	2.82	1.41	1.41	2077.57	183
Hyaloperonospora arabidopsidis	1095	1557	218	2.84	1.42	1.42	1797.96	140
Peronospora belbahrii	1135	1594	224	2.81	1.40	1.40	1913.98	156
Peronospora effusa	1123	1601	225	2.85	1.43	1.43	1973.27	154
Phytophthora palmivora	1219	1683	260	2.76	1.38	1.38	1627.77	179
Phytophthora megakarya	1267	1735	270	2.74	1.37	1.37	1762.26	195
Phytophthora multivora	1299	1827	258	2.81	1.41	1.41	2081.83	184
Phytophthora capsici	1257	1723	260	2.74	1.37	1.37	1686.96	176
Phytophthora pluvialis	1305	1828	261	2.80	1.40	1.40	2064.81	187
Plasmopara halstedii	1178	1656	232	2.81	1.41	1.41	1867.29	162
Plasmopara viticola	1168	1653	234	2.83	1.42	1.42	1858.31	151
Phytophthora infestans	1255	1754	258	2.80	1.40	1.40	1960.77	178
Phytophthora parasitica	1300	1806	261	2.78	1.39	1.39	1872.44	186
Phytophthora nicotianae	1282	1782	258	2.78	1.39	1.39	1793.28	179

Chapter 3

Genome-wide characterization of *Phytophthora infestans* metabolism: a systems biology approach

Sander Y.A. Rodenburg,
Michael F. Seidl,
Dick de Ridder*,
Francine Govers*

* Equal contribution



This chapter was published in Molecular Plant Pathology 19(6) 1403-1413 (2018)
doi: 10.1111/mpp.12623

Abstract

Genome-scale metabolic models (GEMs) provide a functional view of the complex network of biochemical reactions in the living cell. Initially mainly applied to reconstruct the metabolism of model organisms, the availability of increasingly sophisticated reconstruction methods and more extensive biochemical databases now make it possible to reconstruct GEMs for less characterized organisms as well, and have the potential to unravel the metabolism in pathogen-host systems. Here we present a GEM for the oomycete plant pathogen *Phytophthora infestans* as a first step towards an integrative model with its host. We predict the biochemical reactions in different cellular compartments and investigate the gene-protein-reaction associations in this model to get an impression of the biochemical capabilities of *P. infestans*. Furthermore, we generate life stage-specific models to place the transcriptomic changes of genes encoding metabolic enzymes into a functional context. In sporangia and zoospores there is an overall downregulation, most strikingly reflected in the fatty acid biosynthesis pathway. To investigate the robustness of the GEM, we simulate gene deletions to predict which enzymes are essential for in vitro growth. This model is an essential first step towards an understanding of *P. infestans* and its interactions with plants as a system, which will help to formulate new hypotheses on infection mechanisms and disease prevention.

Introduction

The growth and functioning of any living cell is governed by a complex, interconnected set of biochemical reactions, comprehensively referred to as its metabolism (J. Nielsen, 2017). It is essential for cells to consume and break down nutrients taken from the environment, and to use the resulting basic building blocks to construct the molecules needed for life (nucleic acids, amino acids, lipids etc.) and for survival (secondary metabolites). However, the many molecules in this system and the many parameters that govern the biochemical reactions make metabolism hard to study. Systems biology was introduced as a method to study a biological system as a whole by capturing its behaviour in a mathematical abstraction, i.e. a model (Ideker *et al.*, 2001). A model can provide insights into the response of a biological system to certain perturbations or stimuli (Bordbar *et al.*, 2014). A widely studied class of models is that of genome-scale metabolic models (GEMs), which simulate and predict the metabolic behaviour of a cell (Lewis *et al.*, 2012), such as the nutrients it can assimilate and the molecules it can synthesize.

The foundation of a GEM is the set of biochemical reactions that may occur in a cell, often catalyzed by enzymes. Hence, the identification of enzyme encoding genes in the genome of an organism can help to reconstruct an overview of its biochemical capabilities (O'Brien *et al.*, 2015; Yilmaz and Walhout, 2016). In a metabolic model, every reaction is considered as a conversion of substrate metabolites into product metabolites that takes place at a specific rate. The stoichiometry represents the balance of metabolites within the reaction. In steady state, i.e. a situation in which the net metabolite concentrations do not change, the reaction rates are called fluxes. A class of methods called constraint-based modelling can be used to simulate the distribution of these fluxes in certain conditions (Orth *et al.*, 2010). A well-known constraint-based method is flux balance analysis (FBA), which calculates the optimal set of flux values for the entire GEM to attain a specific metabolic objective. Typically, this metabolic objective is maximization of biomass production, a synonym for growth, but can also entail different objectives, for instance, minimization of energy consumption or redox potential (García Sánchez and Torres Sáez, 2014).

To date several semi-automated GEM reconstruction methods and protocols have been proposed (Agren *et al.*, 2013; Karp *et al.*, 2009; Schellenberger *et al.*, 2011; Thiele *et al.*, 2014; Thiele and Palsson, 2010), and the development of central databases for metabolic pathways and models has made biochemical information widely available (Caspi *et al.*, 2014; Kanehisa *et al.*, 2015; King *et al.*, 2016). While initially GEM reconstruction was mainly limited to microbes (prokaryotes and simple eukaryotes), the available resources now allow for reconstruction of genome-scale metabolic models for complex organisms such as mammals and higher plants (Dharmawardhana *et al.*, 2013; Thiele *et al.*, 2013; Yuan *et al.*, 2016). Such models have also already been applied to understand the metabolic interactions between pathogen and host (Duan *et al.*, 2013; Huthmacher *et al.*, 2010; Peyraud *et al.*, 2016). This can provide new hypotheses about a pathogen's infection strategy and may suggest novel control targets (Chavali *et al.*, 2012; Sharma *et al.*, 2017).

Phytophthora infestans is the causal agent of the devastating disease late blight on tomato and potato, posing an important threat to global food production. It belongs to the oomycetes, a class in the eukaryotic Stramenopile lineage that comprises many plant and animal pathogens. *P. infestans* is considered as one of the model species for oomycetes (Haas *et al.*, 2009). In the asexual life cycle of *P. infestans* different stages can be distinguished (Judelson, 2017). When the mycelium starts to sporulate it forms sporangia that are dispersed by wind and water. Sporangia either germinate directly starting new infections or develop into zoosporangia that release zoospores. The latter encyst upon plant contact and germinate, thereby forming an appressorium at the tip from which a penetration peg emerges that mediates entry into the epidermal cells of the host plant. Cell wall degrading enzymes are secreted that may facilitate the penetration process (Brouwer *et al.*, 2014; Meijer *et al.*, 2014). After penetration, hyphae colonize the mesophyll where they grow intracellularly and form haustoria inside the host cells (Whisson *et al.*, 2016). These feeding structures provide a large contact area with the host cytosol, enabling efficient exchange of molecules, to mediate further infection. Apart from the pathogen-host interactions at the protein level, it can be anticipated that an unknown combination of metabolites is taken up from the plant by the pathogen as nutrients.

P. infestans is able to assimilate a wide range of compounds (Hohl, 1991). *In vitro*, *P. infestans* is for example able to grow on pea, rye or Henninger medium which contains an undetermined mixture of various nutrients, such as amino acids, organic acids and lipids (Griffiths *et al.*, 2003; Meijer *et al.*, 2014). Many of the Peronosporales, the lineage that comprises the *Phytophthora* genus, are sterol and thiamine auxotrophs, which implies that these compounds must be acquired from the host (Dahlin *et al.*, 2017; Gaulin *et al.*, 2010; Judelson, 2012). Although sterols are highly beneficial for mycelial growth, they are not essential (Hohl, 1991). Conversely, thiamine is essential for growth. The nutrients that are taken up by the pathogen are converted into biomass and secondary metabolites. P. infestans forms various long-chain polyunsaturated fatty acids, predominantly arachidonic- and eicosapentaenoic acid (Griffiths *et al.*, 2003; Sun *et al.*, 2013). The oomycete cell wall is composed of various sugar polymers, mainly 1,3- and 1,6-β-glucans and cellulose (Grenville-Briggs *et al.*, 2008). Notably, both the long-chain polyunsaturated fatty acids and the cell wall glucans can elicit plant immune responses (Robinson and Bostock, 2015), but it is likely that during infection such responses are suppressed by secreted effector proteins.

Large transcriptional changes of genes encoding metabolic enzymes were observed during the asexual lifecycle of *P. infestans* (Ah-Fong *et al.*, 2017a), suggesting profound changes at the metabolic level. Notably, metabolic enzymes in general were downregulated in the sporangia and zoospores, and many metabolic processes (e.g. biosynthesis of various amino acids) were upregulated in cysts, and during mycelial growth (Ah-Fong *et al.*, 2017a; Grenville-Briggs *et al.*, 2005). Moreover, elevated expression in planta of various nutrient transporter genes suggest a rich influx of nutrients during infection (Abrahamian *et al.*, 2016). Transcriptome studies have analysed the metabolism of P. infestans from a regulatory point of view. However, these studies do not consider post-transcriptional regulation and metabolic reaction fluxes. A genome-scale

metabolic model can provide an overview of the *P. infestans* metabolism, and at the same time predict the functioning of the primary metabolism as a system. Here we propose a first genome-scale metabolic model for *P. infestans*.

Results & Discussion

Draft model reconstruction

We identified all putative enzymes encoded in the *Phytophthora infestans* genome (Haas *et al.*, 2009) by matching all predicted protein sequences to hidden Markov models (HMMs), trained on groups of orthologous proteins from the KEGG Orthology (KO) database (Agren *et al.*, 2013; Kanehisa *et al.*, 2015). This is a particularly suitable method for detection of distant orthologs, since conserved domains have a strong influence on the alignment score and thus this method is sensitive to conserved catalytic domains (Pearson, 2013). Roughly 32% (5,856) of the 18,140 predicted *P. infestans* proteins matched a KO group, yet not every KO group represents a metabolic enzyme catalyzing a biochemical reaction. In total, 1,408 *P. infestans* genes were associated with 1,569 different biochemical reactions, involving 1,663 different metabolites (**Table S1**).

P. infestans is able to assimilate a range of nitrogen compounds, preferably amino acids but also inorganic forms such as nitrate (Hohl, 1991). As a carbon source, *P. infestans* prefers glucose or sucrose, but can also utilize many mono- and disaccharides (Judelson, 2017). Early experiments determined that *P. infestans* can utilize a range of organic sulphur and phosphorus compounds, although more optimal growth rates were observed with inorganic sulphate and phosphate sources (Fothergill and Child, 1964). We added uptake reactions to the model for the minimal synthetic growth medium from literature (Hohl, 1991), the simplest nutrient combination shown to yield *in vitro* growth: glucose, ammonia, phosphate, sulphate and thiamine. Next, we composed a pool of biomass precursor metabolites that must be produced to sustain life: all nucleotides, all 20 L-type amino acids, energy carriers (ATP, GTP), and the cofactors Coenzyme-A, NADH, NADPH and $FADH_2$, that are generally essential for a eukaryotic cell (J. Nielsen, 2017). The exact relative abundance of biomass components has never been quantified for *P. infestans*, therefore the aforementioned biomass metabolites were added to the model as substrates of a single artificial biomass reaction with equal stoichiometry. Additionally, for the phospholipids and fatty acids detected in *P. infestans* (Griffiths *et al.*, 2003b) excretion reactions were included. The known cell wall components 1,3- and 1,6-β-glucan and cellulose are all polysaccharides for which glucose is the precursor metabolite.

We used flux balance analysis (FBA) to calculate the flux through each reaction, optimizing for biomass production (Orth *et al.*, 2010). To predict quantitative fluxes using FBA, it is required to provide: an accurate biomass composition, maintenance ATP requirements, growth rates and species-specific reaction constraints (Thiele and Palsson, 2010). While the lack of detailed

data on *P. infestans* metabolism currently impairs reliable quantitative flux predictions, we can nevertheless deploy FBA to interrogate the model for its connectivity and topology.

Metabolic enzymes are located in various organelles, causing specific metabolic processes to take place in different parts of the cell. For example, the TCA cycle typically occurs in the mitochondria (Zimorski *et al.*, 2017). There is an extensive exchange of metabolites between subcellular compartments (Wanders *et al.*, 2016). Obviously, the compartmentalization influences the connectivity of the reactions and the global behaviour of the model. Based on localization predictions by LocTree 3 (Goldberg *et al.*, 2014), we expanded the model by dividing *P. infestans* proteins over seven subcellular compartments (**Figure 1**). Reactions in the model were assigned to a particular compartment if at least one of the associated enzymes was predicted to localize there. LocTree has been trained on general eukaryotic sequences, which could influence the accuracy of our enzyme localization predictions. However, previous analyses using similar localization predictors showed that proteins predicted to co-localize are often also co-expressed in *P. infestans* (Seidl *et al.*, 2013). The cytosol contained 1138 reactions while the mitochondria contained 359, which is approximately 15% of the total number of reactions in the model (**Table 1**). Of these 359, 160 (45%) were shared with the cytosol (**Figure S1**). Notably, these shared reactions are part of various metabolic pathways, but a relatively large number (42) is linked to the fatty acid biosynthesis (FAB) pathway. The elongation of fatty acids can be governed by a single fatty acid synthase enzyme (EC 2.3.1.86). *P. infestans* has three gene copies for this enzyme, one of which is predicted to encode a mitochondrial isoform (PITG_18025). It was reported that many eukaryotes have a highly conserved, independent mitochondrial FAB pathway that is crucial for development (Hiltunen *et al.*, 2009; Kastaniotis *et al.*, 2017). *Phytophthora* spp. are thought to store energy in fatty acid molecules to facilitate movement of zoospore flagella (Judelson, 2017).

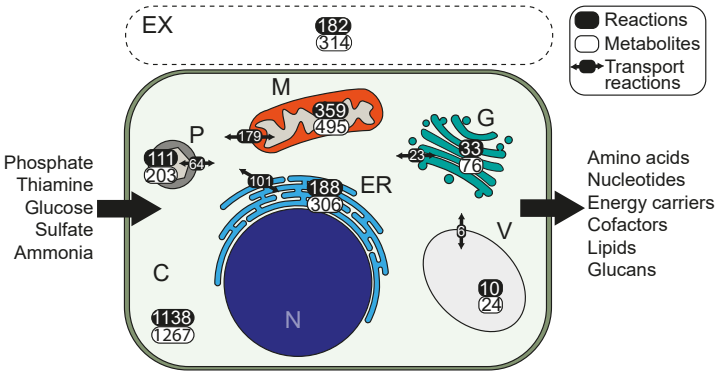


FIGURE 1 | Schematic representation of a *Phytophthora infestans* cell with the number of reactions and metabolites per subcellular compartment and the number of transport reactions deduced from the model presented in this study. In this model the nucleus (N) is not included as a separate subcellular compartment. C: cytosol, M: mitochondrion, P: peroxisome, G: Golgi complex, V: vacuole, EX: extracellular space, ER: endoplasmic reticulum.

In our model, the TCA cycle shares five reactions with the cytosol. One of these is catalyzed by malate dehydrogenase (MDH, EC 1.1.1.37). *P. infestans* has two genes encoding MDH, one encoding an isoform of MDH shown to be active in mitochondria in *P. infestans* and the other encoding a cytoplasmic isoform (López-Calcagno *et al.*, 2009). Other mitochondrial reactions are part of various metabolic pathways, including fatty acid biosynthesis, fatty acid degradation (β-oxidation) and even three glycolytic reactions, involving seven enzymes. The mitochondrial localization of these latter enzymes is likely a remnant of a secondary endosymbiosis event (Judelson, 2017).

Model correction enables flux simulation

After initial reconstruction of the model, 153 invalid reactions (e.g. polymer reactions, see **Methods**) and 107 associated genes were removed from the model. The reconstructed metabolic model of *P. infestans* was initially unable to simulate growth (flux towards all biomass components), due to missing reactions or invalid reaction directionality constraints. This can be the result of an incomplete genome sequence or missannotations. We therefore performed a model gap-filling optimization to find the minimal set of reactions in KEGG that must be added to the model to correct this (**Table S2**). This method proposed 16 additional reactions, and highlighted three reactions that must be reversed to allow for production of all biomass precursors. Next, eight extra drain reactions were added to the model to satisfy the steady-state constraint. Notably, no gap-filling solutions were found for production of the fatty acids eicosapentaenoic acid (EPA) and behenic acid (A.K.A. docosanoate), both of which are produced by *P. infestans* (Griffiths *et al.*, 2003b; Robinson and Bostock, 2015; Sun *et al.*, 2013). This is caused by the lack of fatty acid reactions in KEGG, leaving multiple fatty acid reactions unconnected to other reactions (see KEGG map 01040).

To simulate the metabolite exchange between subcellular compartments, the model must also include intracellular transport reactions. Although nutrient transporters in *P. infestans* have been studied (Abrahamian *et al.*, 2016; Grenville-Briggs *et al.*, 2010), hardly anything is known about the metabolites that are exchanged between the cytosol and subcellular compartments. The annotated substrates for transporter proteins are not specific, and are therefore hard to integrate into the metabolic model. Moreover, transporter substrates such as those from the Transporter Classification Database (Saier *et al.*, 2016) are not cross linked with other databases. To overcome these limitations, we performed an optimization to identify the most likely set of intracellular transport reactions to be added to the model to allow production of all possible metabolites. We determined what metabolites could ultimately be produced by the model, after which we selected the minimal set of transport reactions between the cytosol and any compartment to allow for this (**Figure 1**). The extracellular space (regarded as a subcellular compartment) was excluded from this optimization. The metabolism in this compartment is largely governed by cell wall degrading enzymes. Since it is not possible to distinguish the origin of the metabolites, the pathogen or the host plant, we had to exclude the extracellular space in these analyses.

After all correction steps, 928 of the 2,394 (39%) reactions in the model were able to carry flux based on the defined growth medium, 377 of which carried a nonzero flux when we calculated the optimal fluxes for maximal biomass production (**Table S1**). Of the 2,685 metabolites in the model, 809 could not be produced based on our defined growth medium, and may require additional nutrient uptake. By iteratively adding uptake reactions to the model for each of these metabolites, we can simulate if the import of a specific metabolite would allow the production of additional metabolites (**Table S2**). This reveals unresolved gaps in the model that could have a technical cause, but may also hint at biological properties. For example, episterol was proposed as a compound that would enable production of four other metabolites. This is striking since *Phytophthora* spp. lack sterol biosynthesis enzymes and depend on sterol acquisition from the host plant (Dahlin *et al.*, 2017). Another proposed metabolite is tyramine which would, upon import in the model, enable the production of six other metabolites. Tyramine is a product of decarboxylation of tyrosine, and based on the genome annotation *P. infestans* seems to lack the enzyme that catalyzes this reaction i.e. tyrosine decarboxylase (EC 4.1.1.25). However, a more precise examination of the genome sequence revealed an unannotated open reading frame (on supercontig 1.18, position 2365580 to 2367055) that likely encodes this enzyme.

The metabolic model connects genomic and metabolic properties

We compared the properties of the *P. infestans* GEM (designated *iSR1301*, **File S1**) to GEMs of other eukaryotic microbes (**Table 1**). The size of our *P. infestans* model, in terms of integrated reactions and genes, is in the same order of magnitude as that of a recent GEM of *Phaeodactylum tricornutum*, a closely related diatom (Levering *et al.*, 2016), although our model involves more metabolites. The sizes of the GEMs of the malaria parasite *Plasmodium falciparum* (Plata *et al.*, 2010) and the Leishmaniasis parasite *Leishmania donovani* (Sharma *et al.*, 2017) is much smaller, but the proportion of genes in the model is similar to that of the *P. infestans* model (~7% of total gene number of genes). Although these numbers might be smaller because of genome annotation quality and the level of model curation, it could also be due to the loss of primary metabolic pathways, for which these parasites rely on nutrient import from their hosts (Dean *et al.*, 2014; Gardner *et al.*, 2002). Despite the fact that *P. infestans* has a similar parasitic lifestyle, a pattern of pathway loss is not reflected in the size of our model.

The relation of a gene to an enzyme and its associated reactions is called the gene-protein-reaction (GPR) association (Machado *et al.*, 2016; Thiele and Palsson, 2010). A reaction can be associated with multiple enzymes (isozymes) and genes (paralogs). Conversely, one enzyme may present multiple catalytic domains, or it can have a broad substrate specificity, which associates it to multiple reactions. This "many-to-many-to-many" relationship holds information about redundancy of enzyme encoding genes in a genome, but also about gene essentiality, and the metabolic robustness of an organism to perturbations and fluctuations in nutrient availability (Belda *et al.*, 2012). In our model, 40.4% of the genes are associated with just a single reaction, and 44.9% of the reactions in the model are associated with a single gene, which makes the respective genes essential for specific metabolic tasks (**Figure S2**). In comparison, for

the *P. tricornutum* GEM these numbers are higher (68.6% and 54.5% respectively). The diatom model is presumably of higher quality, since most reactions are manually curated. However, it might also hint at less redundancy of metabolic enzymes.

TABLE 1 | Statistics of the *Phytophthora infestans* GEM *iSR1301* and GEMs of other eukaryotic microbes.

	Phytophthora infestans	***Phaeodactylum tricornutum***	***Plasmodium falciparum***	***Leishmania donovani***	***Saccharomyces cerevisiae***
Reactions	2394	2156	1001	1135	1882
Transport	373	308	233	358	N/A
Cytosolic	1138	942	503	363	N/A
Mitochondrial	359	409	49	197	N/A
Metabolites	2685	1704	616	1135	1454
Genes	1301	1025	366	604	901
% of total	7.17%	9.85%	6.91%	7.30%	13.64%
Model name	iSR1301	iLB1025	iTH366	iMS604	Yeast 7
Reference	This study	Levering *et al.* 2016	Plata *et al.* 2010	Sharma *et al.* 2017	Aung *et al.* 2013

Stage-specific models reflect reduced metabolic activity in sporangia and zoospores

It has been demonstrated that integration of transcriptomics data into a metabolic model has the potential to unveil condition- or tissue specific metabolic activity (Agren *et al.*, 2012; Becker and Palsson, 2008; Gatto *et al.*, 2014; Huthmacher *et al.*, 2010). We had access to transcriptome data of four asexual life stages, i.e. mycelium, sporangia, zoospores and germinating cysts (C. Schoina *et al.,* unpublished; **Chapter 5**), and deployed the iMAT algorithm (Shlomi *et al.*, 2008) to predict stage-specific metabolic models for these life stages. This algorithm considers binary gene expression, i.e. a gene can either be expressed or not. Subsequently, it finds the fluxes through the model, supported by the maximum number of expressed genes, independent of defined medium and biomass composition. This results in sub-models for which all included reactions can carry flux. However, not all underlying genes have to be expressed. In other words, the resulting stage-specific models are sets of reactions that correlate best to the expression of the underlying genes. These reactions are therefore most likely to be metabolically active. If a reaction is absent from a stage-specific model, it is either absent because the expression of the associated genes is low, or because upstream reactions are absent. Comparing the sets of reactions in each stage-specific model might reveal highly active life-stage specific metabolic activity. The distribution of stage-wise expression values for the genes in the model form a slimmer distribution (with slightly higher mean) than that of the total set of genes, indicating that genes in the model are more uniformly expressed (**Figure 3A**). To generate a sufficiently large contrast between the stage-specific models we set the

binary gene expression threshold at 7.04 TPM, the median of all expression values. Based on this threshold, genes were called expressed/not expressed, and life stage-specific models were calculated. Fewer genes were considered expressed in the sporangium and zoospore stages than in mycelium and germinating cyst stages (**Figure 3B**).

The stage-specific models for sporangium and zoospore stages contain fewer reactions in total (**Figure 2A**), concordant with the observed general downregulation of many metabolic pathways in these stages (Ah-Fong *et al.*, 2017a). The mycelium and germinating cysts models contain 1,021 and 1,017 reactions, respectively. Of these, 997 are shared, indicating that these models are highly similar. Although the majority of the reactions, i.e. a core set of 901 reactions, are shared between all four stage-specific models there are also obvious differences; 55 reactions are specifically absent from the zoospore model (hence present in the other three), 21 reactions are only absent from the sporangium model while 20 reactions are absent from the sporangium and zoospore models, but present in mycelium and germinating cyst models. A principal component analysis of stage-wise reaction presence/absence (**Figure 2B**) shows that the mycelium and germinating cyst models cluster relatively close, whereas the sporangium and the zoospore models are more isolated. In summary, our data reflects the regulatory changes that reroute the metabolism of *P. infestans* during each life stage, especially the transitions between mycelium/germinating cyst and sporangium/zoospore stages (Ah-Fong *et al.*, 2017a).

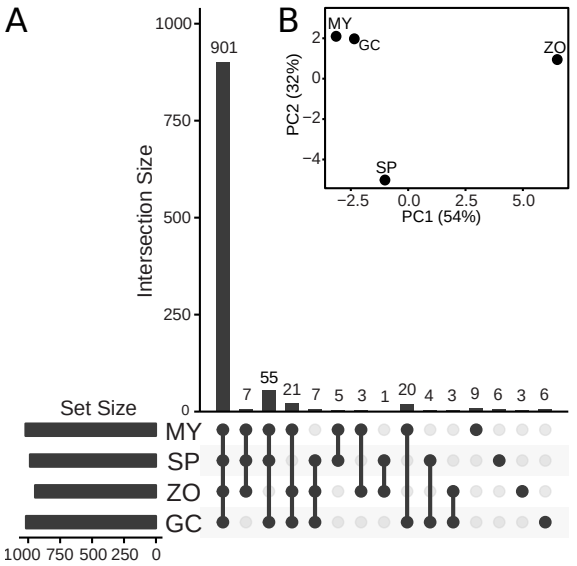


FIGURE 2 | Stage-specific models of *Phytophthora infestans* mycelium (MY), sporangia (SP), zoospores (ZO) and germinating cysts (GC) **A:** Overlap of reaction content between the four stage-specific models. The bars on the bottom-left show the total numbers of reactions in each stage-specific model. The connected bullets indicate the models that are compared, and the bars in the graph represent the number of reactions (intersection size, Y-axis) that overlap between the stage-specific models. **B:** PCA on the stage-wise presence/absence (1/0) of a reaction in the stage-specific models.

To further interpret the presence/absence of reactions in the stage-specific models, we looked at the associated metabolic pathways (**Figure S4**). For instance, the mycelium model contains two unique reactions of the "Vitamin B6 metabolism" pathway (KEGG R00173 and R00174), which represent the interconversion of pyridoxal (vitamin B6) to pyridoxal phosphate, an important cofactor for a large number of reactions, especially for the synthesis of amino acids (Percudani and Peracchi, 2003). The nitrogen metabolism pathway is represented by a core set of nine reactions, but the zoospore model lacks two reactions compared to mycelium and sporangia. Interestingly, these are reactions that contribute to glutamine and glutamate synthesis. Recently, the upregulation of nitrate transporters in zoospores has been reported (Ah-Fong *et al.*, 2017a), which suggests an active nitrogen flux during this life stage. However, a reduced concentration of all amino acids was found in zoospores compared to other life stages (Grenville-Briggs *et al.*, 2005). As pointed out earlier the expression of enzymes in the nitrogen metabolism pathway is highly dynamic and depends on available nutrients (Abrahamian *et al.*, 2016). Possibly the nitrogen imported during the zoospore stage is stored and converted to amino acids at later life stages.

We observe the largest contrast of stage-wise reaction presence/absence in the fatty acid biosynthesis (FAB) pathway (**Figure 3**). A set of 10 reactions is present in the mycelium and germinating cyst models, and absent in the sporangium and zoospore models. Eight reactions are specifically absent in the zoospore model, but three other reactions are specifically present. The latter are all mediated by two cytosolic fatty acid synthases (PITG_10922 and PITG_10926), seemingly downregulated in other stages. Instead, the mitochondrial fatty acid synthase (PITG_18025) seems active in the mycelium and germinating cyst stages. It is likely that fatty acids are synthesized during hyphal stages, since zoospores are thought to use stored fatty acids as nutrient source (Grant *et al.*, 1988; Yousef *et al.*, 2012). These data emphasize that fatty acids likely have an important role in *Phytophthora* zoospores. The three fatty acid synthase enzymes in *P. infestans* play a major role in the FAB process. Intriguingly, there could be a switch between cytosolic and mitochondrial FAB in zoospores. An unanticipated finding, both reported by Ah-Fong and colleagues (2017) and based on our model, is that fatty acid degradation (β-oxidation) is not pronounced in the zoospore stage, despite the predicted role of fatty acids in zoospore motility (Judelson, 2017).

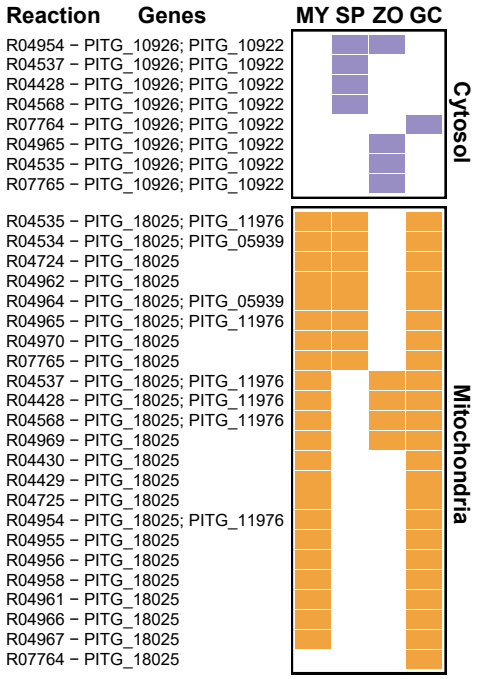


FIGURE 3 | Fatty acid biosynthesis reactions in the stage-specific models of *Phytophthora infestans* mycelium (MY), sporangia (SP), zoospores (ZO) and germinating cysts (GC). Presence/absence of a KEGG reaction (indicated by its ID followed by associated gene IDs) is shown by filled/empty tiles, respectively while mitochondrial and cytosolic reactions are shown in orange and purple, respectively.

Gene deletion simulations propose metabolic vulnerabilities

We investigated what effect gene deletions could have on the primary metabolism of *P. infestans*. By removing single genes from the model, one or more of the associated reactions in the model may be disabled. If such reactions are essential for production of any of the biomass precursors, these deletions disable growth i.e. the mathematical solution of the model becomes infeasible (O'Brien *et al.*, 2015), making such genes interesting candidates for further study. We performed single gene deletion (SGD) simulations of all genes in the model, which suggested 72 genes that would disable growth by disabling production of one of the essential biomass precursors (**Table S2**). These genes were associated with 285 reactions in various metabolic pathways (**Figure 4**). The pathways "phenylalanine, tyrosine and tryptophan biosynthesis" (17 out of 26 reactions vulnerable to SGD) and "valine, leucine and isoleucine" (11/17) are by far the most vulnerable pathways. Notably, the fatty acid degradation pathway is also delicate (17/111). In contrast, the most robust pathways are "tyrosine metabolism" (1/39), and "amino sugar and nucleotide sugar metabolism" (1/27).

There are numerous examples of the application of this method to suggest drug targets in pathogens (Hartman *et al.*, 2014; Kaltdorf *et al.*, 2016; Plata *et al.*, 2010; Sharma *et al.*, 2017;

Yizhak *et al.*, 2013), thus suggesting that identified enzymes in *P. infestans* represent interesting candidates for further study. To confirm that these enzymes are essential for viability, ideally the encoding genes have to be deleted or targeted by site-directed mutagenesis. However, *P. infestans* is a diploid or polyploid organism and making gene knock-outs is not (yet) a straightforward procedure. Gene silencing though is feasible and has been applied to study the function of several genes involved in pathogenesis or signal transduction. The first functional study dealing with genes involved in primary metabolism was published only recently. Abrahamian and colleagues (2016) silenced the genes encoding nitrate and nitrite reductase and their results suggested a role of these genes in virulence. Although both genes are included in our model, they were not marked as essential based on the SGD simulations suggesting alternative routes for nitrogen metabolism. Our simulations are of course based on *in vitro* growth conditions, while in its natural habitat *P. infestans* mainly resides *in planta*. It is thought that *P. infestans* imports larger, organic nutrients such as amino acids during infection (Abrahamian *et al.*, 2016).

Concluding remarks

Here we present, to our knowledge, the first genome-scale metabolic model for the oomycete *Phytophthora infestans*, reconstructed mostly *in silico*, based on reactions found in KEGG. The aim of this study was not to provide a fully quantitative model, but rather to provide a broad overview of cellular metabolism, related to its genome. We optimized the model to be able to convert a minimal pool of nutrients into a set of minimal biomass precursors established from literature. Our model contributes to an understanding of the metabolism of *P. infestans*. However, even after gap filling, the fatty acids EPA and behenic acid could not be generated from the model, due to missing reactions in KEGG. This underscores the limitations of using solely KEGG as a resource, which may be overcome by manual refinement of fatty acid biosynthesis pathways from different databases, such as MetaCyc or BRENDA (Caspi *et al.*, 2014; Placzek *et al.*, 2017). In fact, the BRENDA database holds information on an omega-3 desaturase able to convert arachidonic acid into EPA (EC 1.14.19.25). This exact enzyme of *P. infestans* was recently proven capable of catalyzing this reaction in yeast (Yilmaz *et al.*, 2017). We also revealed one unannotated tyrosine decarboxylase gene in the *P. infestans* genome, emphasizing the fact that the genome annotation of the reference genome published in 2009 by Haas et al. needs to be revisited. On the other hand, it demonstrates this model can be used to aid the discovery of unannotated genes. Obviously the model will improve when a more accurate genome annotation becomes available, and future versions of the model should help us to address the shortcomings encountered in this study. Additionally, an experimentally assessed biomass composition, transporter integration, inclusion of species-specific reaction constraints and growth rates may be used to improve the accuracy of this model. The absence of these data restrains us from making quantitative predictions, such as growth rates or influence of different nutrients.

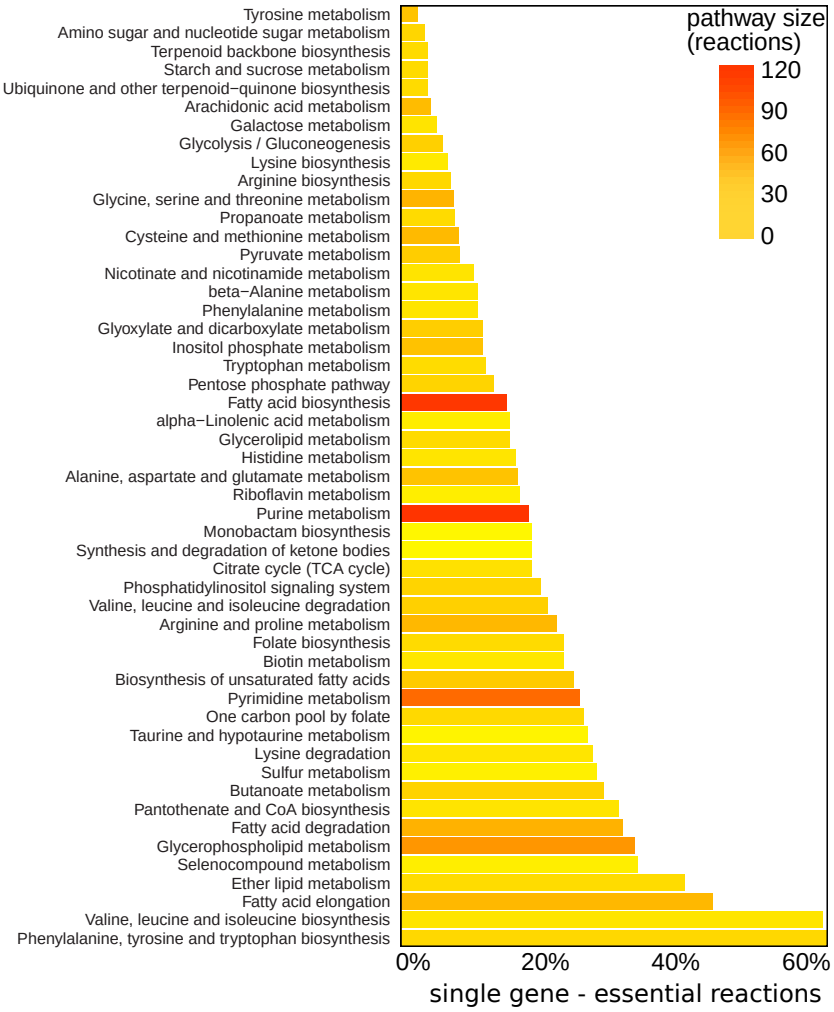


FIGURE 4 | Gene deletion simulations in the *Phytophthora infestans* model. Percentage of essential reactions in each KEGG pathway that can be knocked out by a single gene deletion, thereby disabling production of at least one biomass precursor (i.e. essential). The colors of the bars scale with the absolute numbers of reactions found in our model in a particular KEGG pathway (inset top right).

The life stage-specific models provide a direct functional context for the transcriptome data, by predicting the behavior of *P. infestans* metabolism under influence of stage-wise gene expression and are an alternative for the enrichment methods typically employed in metabolic pathway analyses. Approaching transcriptomic data from a functional point of view may emphasize certain features that are otherwise easily overlooked (e.g. the fatty acid biosynthesis). By building this model we can identify genes that have an essential role when converting simple nutrients to the building blocks of life. The metabolic model we

reconstructed provides a scaffold for future genome-wide systems biology approaches to characterize the metabolism of *P. infestans*, and is an essential first step towards an integrative model of *P. infestans*–host interactions.

Methods

Draft reconstruction

We reimplemented the *getModelFromKEGG getKEGGModelForOrganism* method from the RAVEN Toolbox (Agren *et al.*, 2013) to improve performance and to incorporate minor adaptations. Briefly, Hidden Markov Models (HMMs) were trained on orthologous enzyme sequences derived from the KEGG Orthology (KO) database, Release 2015-11-23. We constructed multiple sequence alignments (MSA) of all eukaryotic protein sequences in every KEGG orthologous group using MAFFT version 7.273 (Katoh and Standley, 2013), using the "localpair" mode for local alignments. For performance reasons, the number of sequences used in an MSA was capped at 100, in which case we selected a random subset of sequences in the KO group. If fewer than 20 eukaryotic sequences were included in a KO group we constructed MSAs for prokaryotic sequences in respective KO group, maintaining the same rules. We used hmmbuild from the HMMER package version 3.1b (Eddy, 1998) to train HMMs on the MSAs. By using *hmmsearch*, we matched the HMMs to the protein sequences of *P. infestans* strain T30-4 (Haas *et al.*, 2009), downloaded on the 25th of July, 2015 from the BROAD Institute website (https://www.broadinstitute.org), currently hosted at NCBI (bioproject 17665). Default parameters were maintained for hmmbuild and hmmsearch, and an E-value threshold of 10^{-20} was applied for hmmsearch. Similar to the *getKEGGModelForOrganism* function of the RAVEN Toolbox (Agren *et al.*, 2013), we performed two pruning steps, but we applied slightly stricter thresholds. First, any protein match to a KO group was removed if $\frac{log(E)}{log\ (E_{bestKO})} < 0.9$, where E represents the E-value of respective protein to a KO group, and E_{bestKO} represents the E-value of the best-matching KO group for that protein. In other words, protein hits are often removed if they have a better match to another KO group. Second, any protein match to a KO group was removed if $\frac{log(E)}{log(E_{bestProt})} < 0.5$, where E again represents the E-value of respective protein match to a KO group, and $E_{bestProt}$ represents the lowest E-value of any protein to this KO group. This reduces the number of matches per KO group to reduce the number of false positives, since there is clearly a better matching protein. Subsequently, we retrieved all KO annotations of *P. infestans* from KEGG (organism ID "pif"). The combined set of matched KO groups was used to retrieve all associated reactions and metabolites from KEGG. Consequently, each reaction in the model was associated with a number of *P. infestans* genes. We hypothesized that each of the genes associated with a reaction is able to catalyze it. We did not consider enzyme complexes, which together would fulfil a single enzymatic task. Reactions were removed automatically if their stoichiometry was undefined (e.g. "1,3-beta-D-

Glucan(n) + UDP-D-glucose <=> 1,3-beta-D-Glucan(n+1) + UDP"), and if the same metabolite ID occurred at both sides of the reaction arrow, which implies a polymer reaction (e.g. "UTP + RNA <=> Diphosphate + RNA"). Additionally, reactions that were associated to metabolites containing the substrings "acceptor", "donor", "tRNA", "enzyme", "aglycon" and "fatty acid" were removed.

Model correction

We predicted the subcellular location of *P. infestans* proteins using LocTree 3 (Goldberg *et al.*, 2014), and subsequently distributed the associated reactions of the model over the cellular compartments. We selected seven compartments for our model: cytosol, extracellular space, mitochondria, endoplasmic reticulum, Golgi complex, peroxisome and vacuole. Reactions associated to enzymes with transmembrane predictions (plasma or intracellular membranes) were assigned to respective compartments on both sides of the membrane, since it is unclear where the catalytic domain is localized. Proteins assigned to any other than our seven compartments, were assigned to the cytosol. Next, all proteins that were predicted to be secreted by Raffaele and colleagues (2010) were assigned to the extracellular space compartment.

The initially reconstructed model was exported to SBML and Microsoft Excel format using CobraPy v0.4.1 (Ebrahim *et al.*, 2013). For the next steps we imported the model into MATLAB (R2015b) using the RAVEN Toolbox v1.8 (Agren *et al.*, 2013). We used Gurobi v7.0.1 (http://www.gurobi.com/) to solve the (mixed-integer) linear programs.

Nutrient uptake and biomass reactions were added to the model by applying the *fillGaps* function from the RAVEN Toolbox to propose gap solutions from KEGG, by constraining biomass production to a positive flux. This method implements the SMILEY algorithm (Rolfsson *et al.*, 2011), including all reactions from a universal set of reactions (in this case KEGG), and subsequently minimizing the flux through these. Prior to this, we temporarily removed the extracellular space compartment to prevent gap solutions here, and added all possible transport and excretion reactions to the model. During gap filling, we allowed the net production of metabolites, whereafter we added drain reactions (allowing excretion) for unbalanced metabolites, to enable steady-state solutions of the model at this point. To predict the set of transport reactions between compartments, we assessed which metabolites can ultimately be produced by the model. Then, we constrained a positive flux for the production of these metabolites, and we removed the transport reactions that did not carry flux.

After these correction steps, we used the function *solveLP* from the RAVEN Toolbox to solve the linear programs of flux balance analysis (FBA), and to get the flux distribution for maximal biomass production. We applied the function *checkProduction* to detect metabolite gaps in the model. This method checks which metabolites are not producible (blocked) from the model, and then iteratively adds uptake reactions for these metabolites, to see whether the uptake of these metabolites unblocks metabolites elsewhere in the model.

Stage-specific models

We used RNA sequencing data (C. Schoina *et al.*, unpublished; **Chapter 5**) to quantify gene expression in four *in vitro* life stages of *P. infestans*, i.e. mycelium, sporangia, zoospores and germinating cysts. Gene expression of *P. infestans* strain T30-4 (Haas *et al.*, 2009) was quantified using Kallisto v0.42.4 (Bray *et al.*, 2016), which expresses mRNA abundance in transcripts-per-million (TPM). This unit represents the number of reads aligned to transcript sequences, normalized for transcript length and sequencing depth, scaled by a million. We determined a binary expression threshold to define if a gene is expressed or not, for which we used the median of all expression values over the four life stages. We decomposed the biomass and other excreted compounds (lipids etc.) into separate excretion reactions, and used the iMAT algorithm (implemented in the function *createTissueSpecificModel*) incorporated in the COBRA Toolbox (Schellenberger *et al.*, 2011; Shlomi *et al.*, 2008) to derive stage-specific models.

Gene knockout simulations

We used the function *findGeneDeletions* from the RAVEN Toolbox to predict the genes that would, upon deletion, disable biomass flux. This method first selects reactions that are supported by a single gene, and then iteratively constrains the flux of these reactions to zero. A gene is marked as essential if the linear program of FBA becomes infeasible after deletion, i.e. when no flux through the biomass reaction is possible.

Acknowledgements

This work was partly supported by the Food-for-Thought campaign from the Wageningen University Fund and by The Netherlands Organization for Scientific Research in the framework of a VENI grant (M.F.S.).

Supplemental Information

TABLE S2A | Reactions added to the model as gap-filling solutions.

	Reaction ID	Compartment
Added	R00383	ER
	R00665	Cytosol
	R00728	Mitochondria
	R00915	Cytosol
	R00965	Cytosol
	R02251	Cytosol
	R02325	Cytosol
	R03013	Cytosol
	R03028	Mitochondria
	R07280	Cytosol
	R08176	Cytosol
	R09375	Cytosol
	R09675	Cytosol
	R10170	Cytosol
	R10404	Mitochondria
	R10706	Cytosol
Reversed	R01213	Cytosol
	R01933	Mitochondria
	R03348	Cytosol

TABLE S2B | Drain/sink Reactions added to the model.

Reaction ID	metabolite
EXC_DRAIN_C00042	Succinate
EXC_DRAIN_C00054	Adenosine 3',5'-bisphosphate
EXC_DRAIN_C00060	Carboxylate
EXC_DRAIN_C00080	H+
EXC_DRAIN_C00229	Acyl-carrier protein
EXC_DRAIN_C00996	Ferricytochrome b5
EXC_DRAIN_C01134	Pantetheine 4'-phosphate
EXC_DRAIN_C01267	3-(Imidazol-4-yl)-2-oxopropyl phosphate
EXC_DRAIN_C03373	Aminoimidazole ribotide
EXC_DRAIN_C03688	Apo-[acyl-carrier-protein]

TABLE S2C | Candidate metabolites to be added to the model, and how many reactions their addition would unblock. Only those unblocking >5 reactions are shown, full table can be found online.

Metabolite	Connects
(S)-3-Hydroxy-3-methylglutaryl-CoA[c]	12
S-Adenosyl-L-homocysteine[c]	9
1-(beta-D-Galactosyl)-2-(2-hydroxyacyl)sphingosine[c]	9
5-Methyltetrahydrofolate[c]	8
Retinyl palmitate[c]	8
Estrone[c]	7
1D-myo-Inositol 1,3,4-trisphosphate[c]	7
L-Selenocysteine[c]	7
m7G(5')pppR-mRNA[c]	7
7,8-Dihydroneopterin 3'-triphosphate[c]	7
Gibberellin A12[c]	7
Tyramine[c]	6

TABLE S2D: Essential genes in the model.

Essential genes			
PITG_18374	PITG_05374	PITG_17925	PITG_02221
PITG_18185	PITG_02925	PITG_15045	PITG_11013
PITG_09954	PITG_15449	PITG_06969	PITG_01411
PITG_00127	PITG_02503	PITG_03410	PITG_05245
PITG_03620	PITG_01245	PITG_17262	PITG_14699
PITG_08445	PITG_01769	PITG_01742	PITG_09271
PITG_01752	PITG_17786	PITG_03530	PITG_16024
PITG_04441	PITG_17983	PITG_17003	PITG_05520
PITG_08061	PITG_06604	PITG_22928	PITG_14831
PITG_09848	PITG_13399	PITG_13402	PITG_07027
PITG_04010	PITG_01049	PITG_04601	PITG_12232
PITG_02087	PITG_09566	PITG_03404	PITG_09726
PITG_20559	PITG_02750	PITG_00765	PITG_01103
PITG_00221	PITG_09536	PITG_01783	PITG_14966
PITG_10595	PITG_02594	PITG_00034	PITG_10026
PITG_17032	PITG_01235	PITG_19174	PITG_01700
PITG_05318	PITG_18048	PITG_06749	PITG_01913
PITG_06352	PITG_16102	PITG_14634	

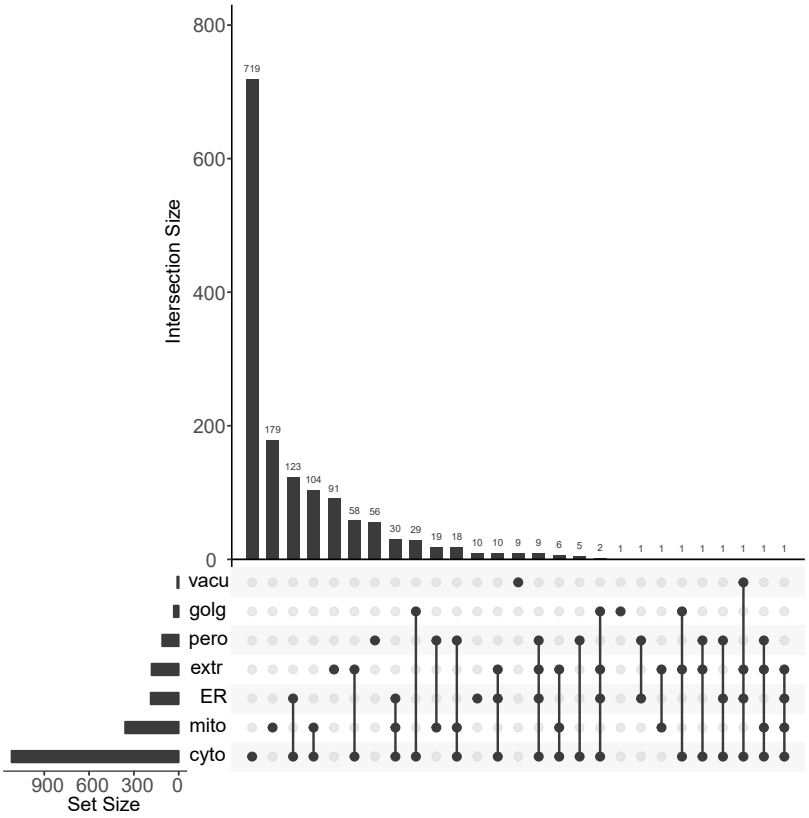


FIGURE S1 | Reactions in the *Phytophthora infestans* model per subcellular compartment and overlap in reaction content between different subcellular compartments. These include cytosol (cyto), mitochondrion (mito), extracellular space (extr), endoplasmic reticulum (ER), peroxisome (pero), Golgi complex (golg), and vacuole (vacu). The bars on the bottom-left show the total numbers of reactions in each subcellular compartment. The connected bullets indicate the compartments that are compared, and the bars in the graph represent the number of reactions (intersection size, Y-axis) that overlap between the compartments.

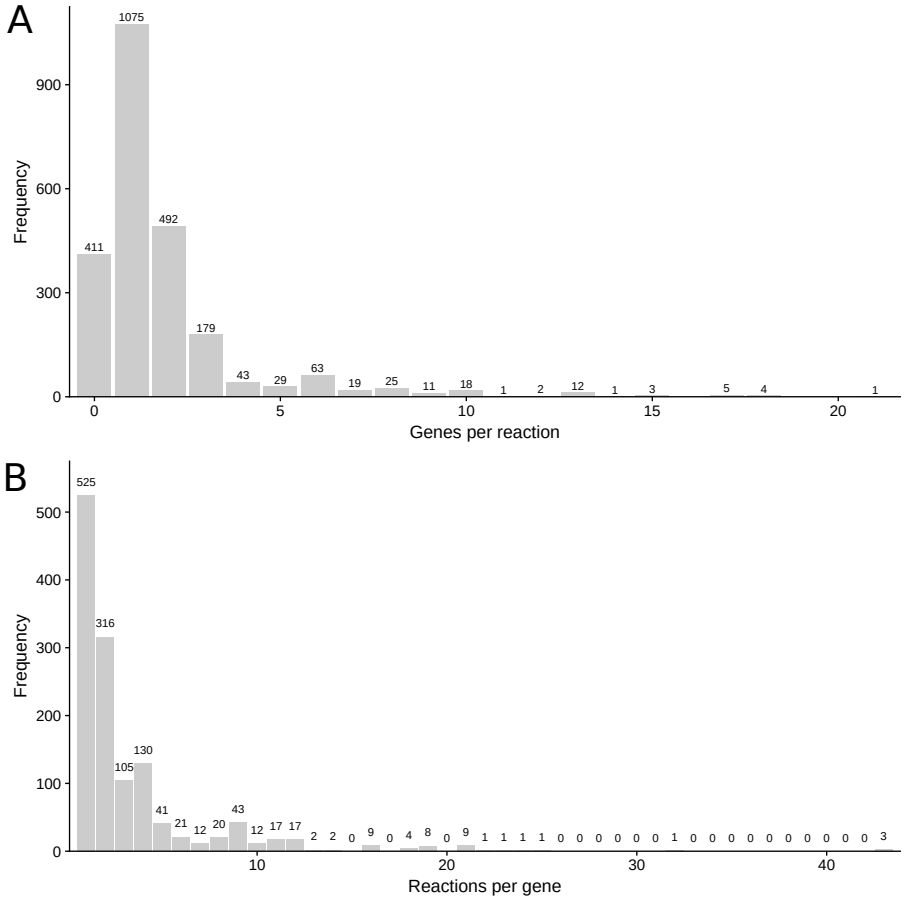


FIGURE S2 | Frequencies of (**A**) gene numbers per reaction and (**B**) reaction numbers per gene in the *Phytophthora infestans* model.

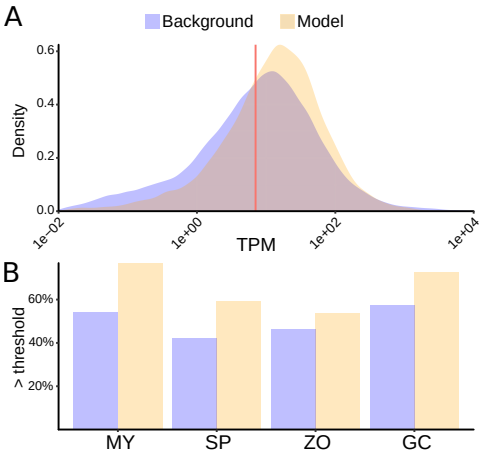


FIGURE S3 | Transcriptome data of *Phytophthora infestans* in relation to stage-specific metabolic models. **A:** Distributions of transcripts-per-million (TPM) expression values of all *P. infestans* genes (background) and the genes in the model, combined from four life stages. The red line indicates the TPM threshold set to distinguish expressed/non-expressed genes in the model. **B:** The percentages of all *P. infestans* genes (background), and the genes in the model for which gene expression in each life stage exceeds the TPM threshold. MY: mycelium; SP: sporangia; ZO: zoospores; GC: germinating cysts.

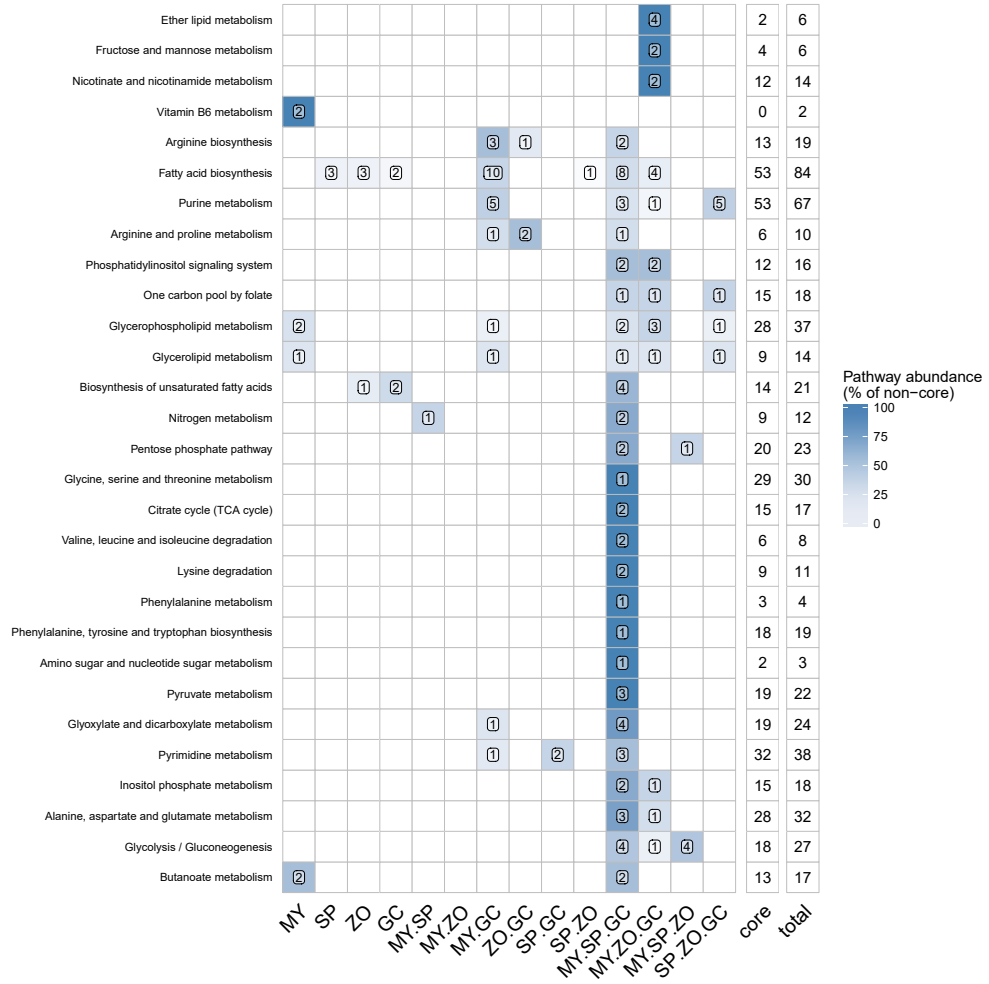


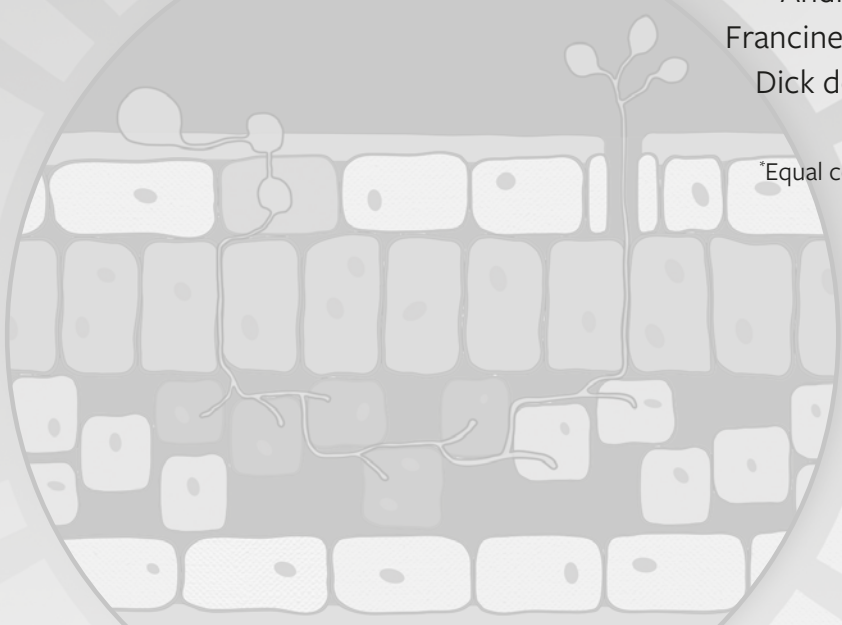
FIGURE S4 | Numbers of reactions per KEGG pathway that are shared between the *Phytophthora infestans* life stage specific models of mycelium (MY), sporangia (SP), zoospores (ZO) and germinating cysts (GC). The colors of the tiles scale to the relative frequencies of all non-core reactions (i.e. the reactions that are absent in at least one stage-specific model). The numbers in the two rightmost columns represent the core set of reactions (shared by all stage-specific models) and the total set of reactions for respective pathway (core + non-core).

Chapter 4

Metabolic model of the *Phytophthora infestans* - tomato interaction reveals metabolic switches during host colonization

Sander Y.A. Rodenburg,
Michael F. Seidl,
Howard S. Judelson,
Andrea L. Vu,
Francine Govers*,
Dick de Ridder*

*Equal contribution



This chapter was published in mBio 10:e00454-19 (2019)
https://doi.org/10.1128/mBio.00454-19

Abstract

The oomycete pathogen *Phytophthora infestans* causes potato and tomato late blight, a disease that is a serious threat to agriculture. *P. infestans* is a hemibiotrophic pathogen and during infection it scavenges nutrients from living host cells for its own proliferation. To date, the nutrient flux from host to pathogen during infection has hardly been studied and the interlinked metabolism of pathogen and host remains poorly understood. Here, we reconstructed an integrated metabolic model of *P. infestans* and tomato (*Solanum lycopersicum*) by integrating two previously published models for both species. We used this integrated model to simulate metabolic fluxes from host to pathogen and explored the topology of the model to study dependencies of the metabolism of *P. infestans* on that of tomato. This showed for example, that *P. infestans,* a thiamine auxotroph, depends on certain metabolic reactions of the tomato thiamine biosynthesis. We also exploited dual-transcriptome data of a time course of a full late blight infection cycle on tomato leaves and integrated the expression of metabolic enzymes in the model. This revealed profound changes of pathogen-host metabolism during infection. As infection progresses, *P. infestans* performs less *de novo* synthesis of metabolites and scavenges more metabolites from tomato. This integrated metabolic model for the *P. infestans*-tomato interaction provides a framework to integrate data and generate hypotheses about *in planta* nutrition of *P. infestans* throughout its infection cycle.

Importance

Late blight disease caused by the oomycete pathogen *Phytophthora infestans,* leads to extensive yield losses in tomato and potato cultivation worldwide. To effectively control this pathogen, a thorough understanding of the mechanisms shaping the interaction with its hosts is paramount. While considerable work has focused at exploring host defense mechanisms and identifying *P. infestans* proteins contributing to virulence and pathogenicity, the nutritional strategies of the pathogen are mostly unresolved. Genome-scale metabolic models (GEMs) can be used to simulate metabolic fluxes and help in unravelling the complex nature of metabolism. We integrated a GEM of tomato with a GEM of *P. infestans*, to simulate the metabolic fluxes that occur during infection. This yields insights in the nutrients that *P. infestans* obtains during different phases of the infection cycle and helps in generating hypotheses about nutrition *in planta*.

Introduction

Plants and pathogens maintain a complex relationship that generally involves the secretion of effector proteins by the pathogen to manipulate plant cell processes and the scavenging of nutrients from the host by the pathogen to support its growth and proliferation (Rovenich *et al.*, 2014). While increasing knowledge is gained on secreted effectors proteins that facilitate host colonization (Whisson *et al.*, 2016), understanding of pathogen nutrition remains underexplored. A class of organisms comprising important plant and animal pathogens are the oomycetes (Derevnina, Petre, *et al.*, 2016). These share morphological characteristics with fungi, yet belong to the Stramenopiles, a eukaryotic lineage that besides oomycetes also includes diatoms and brown algae (McCarthy and Fitzpatrick, 2017). The most well-known oomycete is *Phytophthora infestans,* the causal agent of late blight disease on tomato and potato, which leads to significant yield losses worldwide (Kamoun *et al.*, 2015). *P. infestans* is challenging to control. The pathogen has a highly flexible genome facilitating rapid adaptation to control strategies, be it resistant cultivars or chemical agents (Fry, 2016), and its profuse sporulation causes *P. infestans* to spread extremely fast (Leesutthiphonchai *et al.*, 2018). Hence, there is a continuous quest for novel, more durable control strategies.

P. infestans sporangia are aerially dispersed and after landing on a plant surface these can release flagellate zoospores (Judelson, 2017). These encyst and germinate to form a germ tube with an appressorium to penetrate epidermal cells of the plant. From there, *P. infestans* colonizes the mesophyll; hyphae grow in the apoplast while forming intracellular feeding structures, called haustoria. Both the apoplastic hyphae and haustoria provide close contact with the plant, facilitating the exchange of effectors and nutrients (Judelson and Ah-Fong, 2018). Oomycetes are considered osmotrophs that extracellularly catabolize complex host polymers such as proteins, sugars, and fatty acids using an arsenal of secreted enzymes, followed by the import of degraded nutrients into the cell (Richards and Talbot, 2013). Moreover, *P. infestans* is a hemibiotrophic pathogen that requires viable host cells during the initial stages of the infection cycle, the so-called biotrophic stage with minimal symptoms. Typically after three to six days, this is followed by a necrotrophic stage during which the lesion becomes necrotic and new sporangia emerge (Van West *et al.*, 1998; Zuluaga *et al.*, 2016). This implies that the physiology of the host tissue changes throughout the infection cycle, and consequently also the nutrients available for the pathogen (Ah-Fong *et al.*, 2017b). Conceivably, *P. infestans* fine-tunes its metabolism to available nutrients, for example by regulating the expression of enzyme-encoding genes through catabolite repression and/or substrate induction (Divon and Fluhr, 2007; Judelson *et al.*, 2009b).

The *P. infestans* metabolism is remarkably dynamic. Transcriptome-based studies revealed significant differences in transcript abundance of enzyme-encoding genes throughout the asexual and sexual lifecycles and during plant infection (Abrahamian *et al.*, 2016; Abrahamian *et al.*, 2017; Ah-Fong, Kim, *et al.*, 2017; Niu *et al.*, 2018). While these studies provide detailed insight into the potential dynamics of metabolic enzymes, insight gained into the overall

characteristics of the cell metabolism is very limited. Importantly, cell metabolism is not a static framework of reactions and pathways but adapts to different environments to allow for uptake of metabolites or the production of required compounds. These dynamics are facilitated by the regulation of the rates of individual reactions within the pathways. However, as many intrinsic (e.g. enzyme activity) or extrinsic (e.g. pH or temperature) factors influence these reaction rates, studying cell metabolism remains challenging.

Genome-scale metabolic models (GEMs) can help to understand cell metabolism. GEMs represent cell metabolism as a cell-scale network of biochemical reactions, typically distributed over several cellular compartments, that connects the uptake of nutrients to the production of biomass precursors (J. Nielsen, 2017; O'Brien *et al.*, 2015). Assuming metabolism to be in steady-state allows the derivation of possible rates (called metabolic fluxes) for each reaction in the network (**File S1**). A GEM can be used to predict putative nutrients and essential enzymes/reactions in the cell, thereby generating hypotheses about the responses of the cell metabolism to perturbations (Chavali *et al.*, 2012). Therefore, metabolic models have great potential to aid in the development of novel control strategies against pathogens (Cesur *et al.*, 2018; Peyraud *et al.*, 2017). The metabolism of a pathogen is tightly interconnected with that of its host, and can be regarded as a single system (Olive and Sassetti, 2016). Moreover, the relationship of a pathogen with its host shapes its metabolism through evolution, leading to enzyme gene loss, which makes the pathogen dependent on its host (Spanu, 2012). An integrated host-pathogen metabolic model can yield insights into the system-wide metabolic fluxes that shape an infection. Thus far only few models exist that describe the joint metabolism of beneficial or pathogenic microbes and their hosts (Aller *et al.*, 2018; Ankrah *et al.*, 2018; Bordbar *et al.*, 2010; Huthmacher *et al.*, 2010; Magnúsdóttir *et al.*, 2016; Raghunathan *et al.*, 2009; Raghunathan *et al.*, 2010).

Recently, we published the first GEM for *P. infestans* (Rodenburg *et al.*, 2018a). We used this model to simulate the growth of *P. infestans* on minimal culture medium, and predicted the essential genes and corresponding reactions that are required to convert its nutrients into biomass components (Rodenburg *et al.*, 2018a). Here, we integrated the *P. infestans* GEM with a tomato GEM (Yuan *et al.*, 2016). Using these models, we developed a host-pathogen interaction metabolic model that allowed us to identify and study hallmarks of the *P. infestans*-tomato interaction. Exploiting gene expression data of late blight infections on tomato enabled us to dissect the changes in pathogen and host metabolism, providing novel insights into this host-pathogen interaction.

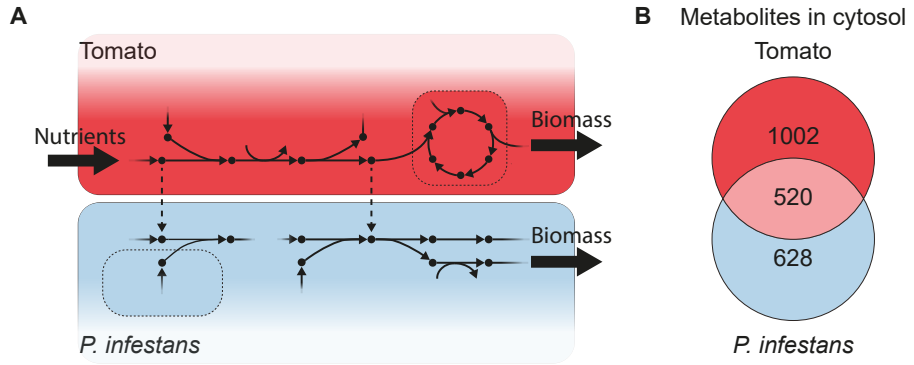


FIGURE 1 | The integrated *P. infestans*-tomato model. **A:** Schematic illustration. Dots are metabolites, arrows are reactions, dotted lines represent the host-pathogen transport reactions. **B:** The numbers of unique and shared cytosol metabolites.

Results & Discussion

A metabolic model for the tomato-*P. infestans* pathosystem

To reconstruct an integrated tomato-*P. infestans* metabolic model, we exploited a recently constructed tomato GEM (Yuan *et al.*, 2016). This tomato model comprised 2143 reactions and 3410 genes, whereas the previously constructed *P. infestans* GEM comprised 2394 reactions and 1301 genes, suggesting that *P. infestans* has less genetic redundancy for metabolic enzymes than tomato (**Table 1**). Our original *P. infestans* GEM was based on a framework reconstructed by annotation of enzyme orthologs in the *P. infestans* genome, linking those the biochemical reactions in KEGG (Kanehisa *et al.*, 2015; Rodenburg *et al.*, 2018a). Based on literature we added a minimal growth medium, a set of known biomass precursors and cellular compartments, and simulated *in vitro* growth. The tomato GEM we used was originally used to simulate photorespiratory fluxes under different conditions, and was reconstructed from LycoCyc (Caspi *et al.*, 2014). We manually improved the *P. infestans* and the tomato GEMs (see Methods). For example, according to recent insights into the mitochondrial localization of glycolytic enzymes (Abrahamian *et al.*, 2017), we curated the gene-reaction association of *P. infestans* in this pathway. In the tomato GEM we curated the thiamine biosynthesis, since *P. infestans* is a thiamine auxotroph (Judelson, 2017). The two GEMs were connected by so-called transport reactions, representing the nutrient flux from tomato to *P. infestans*. Although the transporter repertoire of *P. infestans* has been predicted (Abrahamian *et al.*, 2016), we could not attribute specific substrates to transporters. Most functional annotations lack specificity, and even between oomycete species transporter substrate specificity can vary (Savory *et al.*, 2018). Therefore, we chose for an unbiased approach, connecting the two GEMs by addition of a hypothetical unidirectional transport reaction for each for the 520 metabolites that were

found shared between the cytosol compartments, each one representing the flux of a single metabolite from the tomato cytosol to the *P. infestans* cytosol (**Figure 1).** We realized that fluxes may occur in both directions (i.e. from host to pathogen and *vice versa*) for a pathogen to maintain homeostasis during infection, but the absent knowledge of this process only allowed us to model host-to-pathogen nutrient fluxes. The 520 shared cytosol metabolites for which transport reactions were added to the model predominantly include metabolites taking part of primary metabolic subsystems (from KEGG). These include the subsystems amino acid biosynthesis (*n*=75), carbon metabolism (*n*=61), purine- and pyrimidine biosynthesis (*n*=51 and *n*=37 resp.), and the subsystem 'ABC transporters' (*n*=42). The latter indicates that at least 42 metabolites are known ABC transporter substrates. Of the 520 introduced transport reactions, 179 could carry a flux, meaning that the respective metabolites can both be produced by tomato and assimilated by *P. infestans*. The integration of the two GEMs resulted in an integrated metabolic model of *P. infestans*-tomato interaction, comprising a total of 4695 reactions, 4303 metabolites and 4578 genes (**Table 1**).

Table 1: Properties of the genome-scale metabolic models of *P. infestans* and tomato (*S. lycopersicum*), and of the integrated metabolic model.

System	*P. infestans*[a]	*S. lycopersicum*[b]	*P. infestans* + *S. lycopersicum*[c,d]
Model name	iSR1301	iHY3410	iSR4578
Reactions	2394	2143	4695
Metabolites	2685	1998	4303
Genes	1301	3240	4578
Compartments	7	5	12

[a] Rodenburg *et al.* (2018a)
[b] Yuan *et al.* (2016)
[c] This study
[d] The numbers in this column are not the exact sum of the two independent models, since several adaptations were made (**Methods**).

Flux simulations pinpoint nutrients utilized by *P. infestans*

Given a metabolic model and a metabolic objective for the cell, optimal values for the fluxes through all reactions can be calculated to simulate production of *P. infestans* biomass precursors through the import of nutrients from the tomato cytosol (Orth *et al.*, 2010). An objective could be, for instance, to maximize the production of biomass or to produce biomass at minimal enzyme expense. To study the versatility of the model, we composed four scenarios based on different cellular objectives (**Table S2**). In Scenario I, we calculated the minimal set of nutrients *P. infestans* needs to import from tomato to form all its biomass precursors. This was already possible by importing three compounds: thiamine, L-gamma-glutamyl phosphate and cysteine. While this scenario is unlikely to occur *in planta*, it does illustrate that *P. infestans* has a comprehensive enzyme repertoire, enabling it to form its biomass precursors based on

a very small pool of nutrients. In Scenario II we calculated *P. infestans* biomass production using the minimum number of *P. infestans* reactions possible, to simulate growth with minimal enzyme cost for *P. infestans*. In this scenario, 48 nutrients were used, including ammonium, a known growth substrate (Hodgson, 1958). However, ammonium is an unfavored nitrogen source compared to amino acids (Abrahamian *et al.*, 2016). Among other imported nutrients in this scenario are the tricarboxylic acid (TCA) cycle intermediates malic and succinic acid, for which *P. infestans* has an experimentally verified transporter (Savory *et al.*, 2018). Organic acids such as succinic and malic acid strongly promote *P. infestans* growth *in vitro*, especially in combination with ammonium as nitrogen source (Hohl, 1991). Notably, 29 of the nutrients in this scenario are direct biomass precursors for *P. infestans*, including 19 amino acids (aspartate being the sole exception) (**Figure S1**). In Scenario III, we maximized the usage of *P. infestans* reactions, to simulate a scenario in which the pathogen maximally exploits its metabolism. This simulation predicted a pool of 29 nutrients, including the inorganic compounds nitrite and hydrogen sulfide. Interestingly, *Phytophthora* spp. retained the complete assimilation pathways for these compounds (Abrahamian *et al.*, 2016), in contrast to multiple obligate biotrophic oomycetes which lost multiple genes throughout evolution (Judelson, 2017). In Scenario IV we combined scenarios I and II: we calculated the minimal nutrient uptake combined with minimal usage of *P. infestans* reactions to produce biomass. Since both nutrient import and assimilation into biomass require energy, these processes likely require a tradeoff, and hence we anticipate that this is a more realistic scenario. Here, 38 nutrients were used, 16 of which were amino acids.

The four scenarios described above have opposite objectives (minimize/maximize the fluxes of *P. infestans*) and simulate extreme circumstances not likely found in nature. The true pool of imported nutrients will likely be a combination of the pools predicted in each individual scenario. A comparison of the nutrients imported in all four scenarios revealed sets of nutrients in common between scenarios. There are 14 nutrients found in three out of four scenarios, i.e. II, III and IV (**Figure S1**), including several amino acids, nucleotide precursors, and glycerol 3-phosphate as lipid precursor. Surprisingly, aspartate is not imported in any of the scenarios, even though it is a biomass precursor of *P. infestans* and serves as a precursor to a variety of other amino acids (Azevedo *et al.*, 2006; Jander and Joshi, 2009). An oomycete-specific form of aspartate aminotransferase (EC 2.6.1.1) was found that seems to play a key role in pathogenicity in *P. sojae*, possibly by balancing nitrogen and carbon metabolism and facilitating the interconversion between amino acids (Wang *et al.*, 2018). The presence of this enzyme in our model explains why aspartic acid import may not be necessary. Interestingly, thiamine is only imported in scenario III (maximizing the *P. infestans* fluxes), while thiamine pyrophosphate (TPP) is imported in scenarios II and IV (minimizing fluxes). It is assumed that many oomycetes import thiamine to form TPP (Judelson, 2017), an essential cofactor in carbohydrate metabolism (Chan *et al.*, 2013) and in TPP-responsive riboswitches (Mukherjee *et al.*, 2018). However, related Stramenopiles were found to also grow on thiamine alternatives (McRose *et al.*, 2014), and possibly *P. infestans* can import up- or downstream compounds instead. Taken together, these simulations suggest a range of nutrients from tomato that

can be effectively assimilated by *P. infestans*. Nutrients in common between the simulated scenarios suggest that these are likely more versatile than others and hence most useful for *P. infestans* to import.

Network analysis identifies dependencies of *P. infestans* to tomato metabolism

To identify dependencies of *P. infestans* on tomato, we investigated the topology of the integrated model by looking for essential reactions (i.e. reactions that are indispensable in the model to form all biomass precursors) and (inter-)dependencies of reactions (i.e. the flux of a particular reaction relies on the flux of another), also known as coupled reactions. Coupled reactions within a metabolic model can be identified using a method called Flux Coupling Analysis (FCA) (**File S1**) (Burgard *et al.*, 2004). FCA of the model identified 77 coupled *P. infestans*-tomato reaction pairs (**Figure 2**), involving 53 unique *P. infestans* reactions and 49 unique tomato reactions (**Table S3**). One of these *P. infestans*-tomato reaction couplings comprises the biomass reaction of TPP in *P. infestans*, which is coupled to 17 tomato reactions, illustrating that *P. infestans* TPP is dependent on the thiamine biosynthesis pathway in tomato. In addition, of the 4695 reactions in the model, 112 *P. infestans* reactions and 35 tomato reactions were found to be essential for *P. infestans* (**Figure 2A**). Some reactions are essential and coupled to many other reactions at the same time (appearing as hubs in the graph), implying that these play a central role in the model with potentially large biological implications. For instance, there is a tomato transport reaction of aspartate into the plastid compartment, supplying aspartate as an amino-group donor for the synthesis of a thiamine precursor (aminoimidazole ribotide) (Goyer, 2010; Nelson and Cox, 2017), thus indirectly making this tomato reaction essential for *P. infestans*. Similarly, ATP and phosphate transport between *P. infestans* cytosol and mitochondria is essential and seems to play an important role in the model. Phosphate and sulphate uptake by tomato are part of the defined growth medium, hence coupled to reactions in both species. FCA also revealed several clusters of tightly interconnected, coupled essential reactions (**Figure 2A**). One of these represents the *P. infestans* fatty acid biosynthesis. In the model this comprises a mostly linear pathway, and consequently most reactions are coupled (interdependent). Fatty acid biosynthesis in oomycetes is associated to fungicide resistance (Maridueña-Zavala *et al.*, 2017) and energy storage for sporangia (Judelson, 2017).

Couplings to the host-pathogen transport reactions in particular can provide information about the importance of transported metabolites; the more *P. infestans* reactions are coupled to a transport reaction, the more likely that the associated substrate is important for *P. infestans.* We selected the most frequently coupled transport reactions from the model and assessed their couplings to other tomato and *P. infestans* reactions (**Figure 2B**). The substrates of these transport reactions were associated to a diverse range of metabolic processes in *P. infestans*, i.e. pantothenate/CoA biosynthesis (2-dehydropantoate), *de novo* pyrimidine biosynthesis (S-dihydroorotate) (García-Bayona *et al.*, 2014), riboflavin metabolism (2,5-diamino-4-hydroxy-6-(5-phosphoribosylamino)pyrimidine), and inositol phosphate metabolism (myo-inositol 3-phosphate). Notably, three transporter substrates were associated with glycerophospholipid

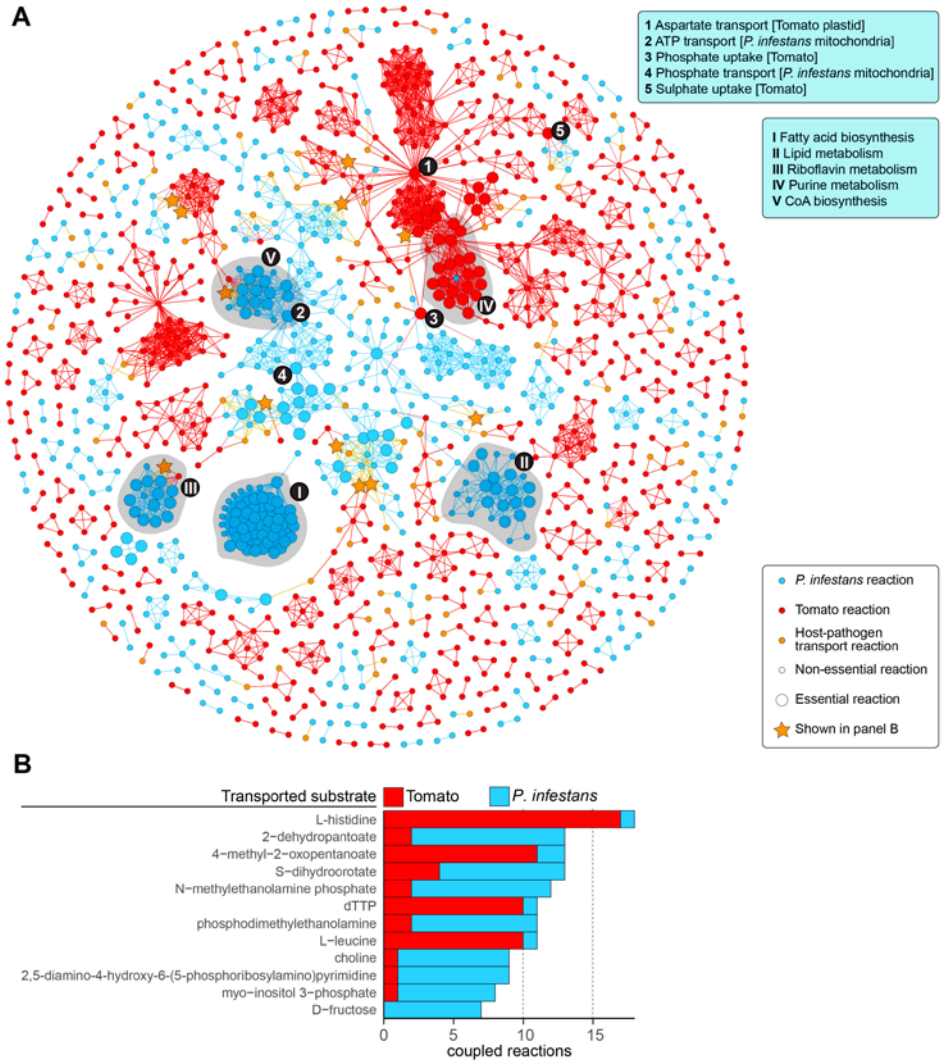


FIGURE 2 | Flux coupling between reactions in the *P. infestans*-tomato model. **A:** Graph showing the coupled reactions in the model. Nodes represent reactions in tomato (red) or *P. infestans* (blue) and host-pathogen transport (green), edges represent coupling between those reactions. Node size reflects essentiality for *P. infestans* biomass production. Stars represent transport reactions listed in panel B. Highly connected nodes (1-5) and clusters (I-V) are indicated and listed in the boxes on the right (see also main text). **B:** The nutrients associated to the 12 most frequently coupled host-pathogen transport reactions. The bars are stacked and indicate the number of coupled reactions per species.

metabolism (phosphodimethylethanolamine N–methylethanolamine phosphate and choline). It is conceivable *P. infestans* can take up glycerophospholipid precursors to facilitate the formation of membrane lipids (Griffiths *et al.*, 2003). Seven *P. infestans* reactions were coupled to the reaction importing fructose, an efficient growth substrate (Hodgson, 1958). Among the most frequently coupled host-pathogen transport reactions are also those transporting dTTP,

histidine and leucine. Since these are mainly coupled to tomato reactions, the import of these nutrients seems to depend on their biosynthesis pathways in tomato. FCA can identify coupled reactions in metabolic models that are not necessarily directly connected in the network and can thus derive functionally related modules of (pathogen-host) metabolism that may otherwise be overlooked.

Transcriptome-based submodels provide insight into the dynamics of *P. infestans*-tomato metabolism

To obtain insight into the metabolism at subsequent developmental stages of the *P. infestans*-tomato interaction, we integrated the model with dual-transcriptome data of *P. infestans* strain 1306 on tomato leaf infection (2 to 6 days post-inoculation, sampled every 4 hours). This strain did not exhibit a strong necrotic phenotype and shows strong profuse sporulation around 4 days post-inoculation (dpi). Previous studies on this *P. infestans* strain have shown marker genes for biotrophic growth peak at 2/3 dpi, those for necrotrophic growth peak at 4/5 dpi (Abrahamian *et al.*, 2016), and those for sporulation are increasingly expressed from 3 days on, and peak at 4 dpi. Expectedly, the amount of reads mapping to the *P. infestans* genome steadily increased over time (7-84%), the amount of reads mapping to the tomato genome decreased accordingly (87-8%) (**Figure S2**). The transcriptome data was used to generate time point-specific submodels, which are subsets of the full model according to the expression of the genes in the model. Since gene expression and metabolic activity are not directly related, we used the INIT algorithm (Agren *et al.*, 2012) that calculates a submodel with maximal agreement to the expression of genes in the model, such that biomass can be produced (see **Methods**). This resulted in 25 submodels containing 32-44% of the reactions of the full model (**Figure 3**), indicating that only a subset of reactions is required to form all defined biomass precursors. Over time, the total number of reactions per submodel decreases (**Figure 3**), mostly due to a reduction of *P. infestans* reactions. The number of tomato reactions is relatively stable across the submodels, while the number of transport reactions increases slightly over time. These results suggest that as the infection progresses, *P. infestans* relies less on its own metabolism but increasingly imports metabolites from the necrotic tomato lesion. Conceivably, once *P. infestans* switches to its necrotrophic lifestyle, nutrients can be more easily obtained from the decaying leaf tissue (Divon and Fluhr, 2007). To get an impression of the metabolic changes in the infection over time, we calculated all pairwise distances of the reaction content of the submodels (**Figure 3**). This revealed a gradient of similarity scores, clearly reflecting metabolic changes over time, possibly as a response to changing nutrient availability in the tomato leaf tissue. A relatively large distance can be observed between groups of submodels, suggesting two more profound switches in metabolism at roughly 3d/8h and 4d/12h, possibly reflecting a transition from the pathogens biotrophic growth to necrotrophic growth and sporulation (Abrahamian *et al.*, 2016; Judelson *et al.*, 2009a; van West *et al.*, 1998; Zuluaga *et al.*, 2016).

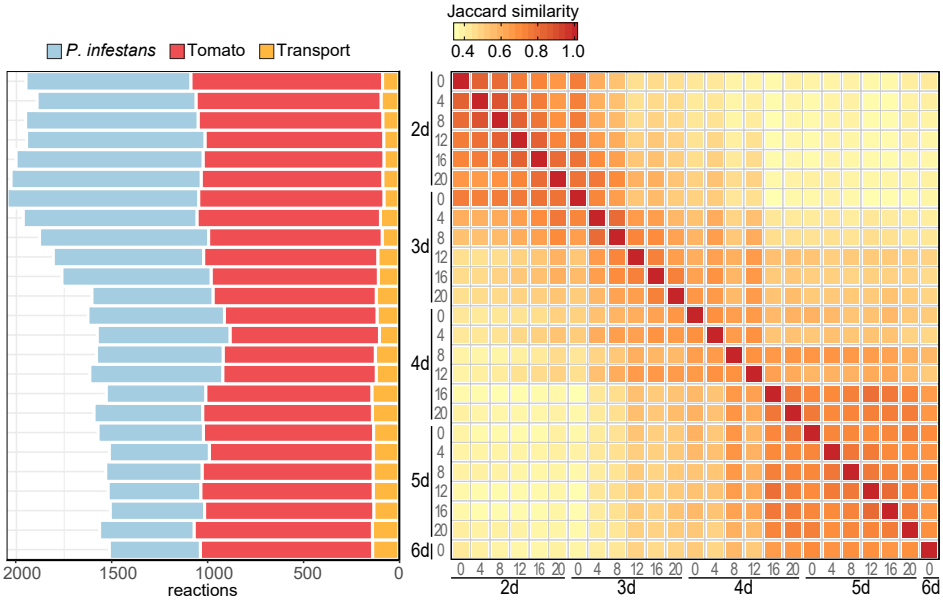


FIGURE 3 | Submodels based on dual-transcriptome data from a time course covering a full infection cycle of *P. infestans* on tomato leaf. Submodels are representative for the sampling time post inoculation shown in days (d) and hours on the Y-axis (middle) and X-axis (left panel). Left panel: Stacked bar chart indicating the number of reactions per species and number of host-pathogen transport reactions (not part of either species). Right panel: Jaccard similarity (intersection divided by union) between the reaction content of the submodels.

To get more insight into the metabolic processes of each submodel, we performed Fisher's exact tests on the KEGG pathways in the submodels. A variety of metabolic pathways were overrepresented in submodels compared to the full model (**Figure 4**). Consistent with the previously observed gradient of similarity scores (**Figure 3**), groups of overrepresented pathways can be distinguished before and after 3d/8h. Notably, in the early submodels prior to the first transition, *P. infestans* reactions are enriched for several primary metabolic pathways such as amino acid biosynthesis, glycolysis and the TCA cycle. This could be a response to sugars in the apoplast that are still produced by photosynthesizing tomato leaf cells during the biotrophic phase of infection (Chen, 2014; Fatima and Senthil-Kumar, 2015). In contrast, late submodels and in particular after the second transition (4d/12h) show that in tomato several amino acid biosynthesis pathways are enriched, suggesting that during early infection amino acids are mostly synthesized *de novo* by *P. infestans* but later scavenged from tomato. In addition, early submodels are enriched in folate-related pathways on both sides, suggesting that folate plays an important role in early infection, which is in line with the finding that *P. infestans* has an unusually large repertoire of putative folate-biopterin transporters (Ah-Fong *et al.*, 2017a). Arachidonic acid metabolism is enriched in mid-infection submodels, which is a characteristic fatty acid in *P. infestans* known to elicit plant defense, and to act as a signaling molecule in plants (Robinson and Bostock, 2015; Savchenko *et al.*, 2010). In summary, our

analyses reveal profound changes in metabolic processes during infection of tomato by *P. infestans* and suggest that *P. infestans* reduces its metabolism at the expense of tomato.

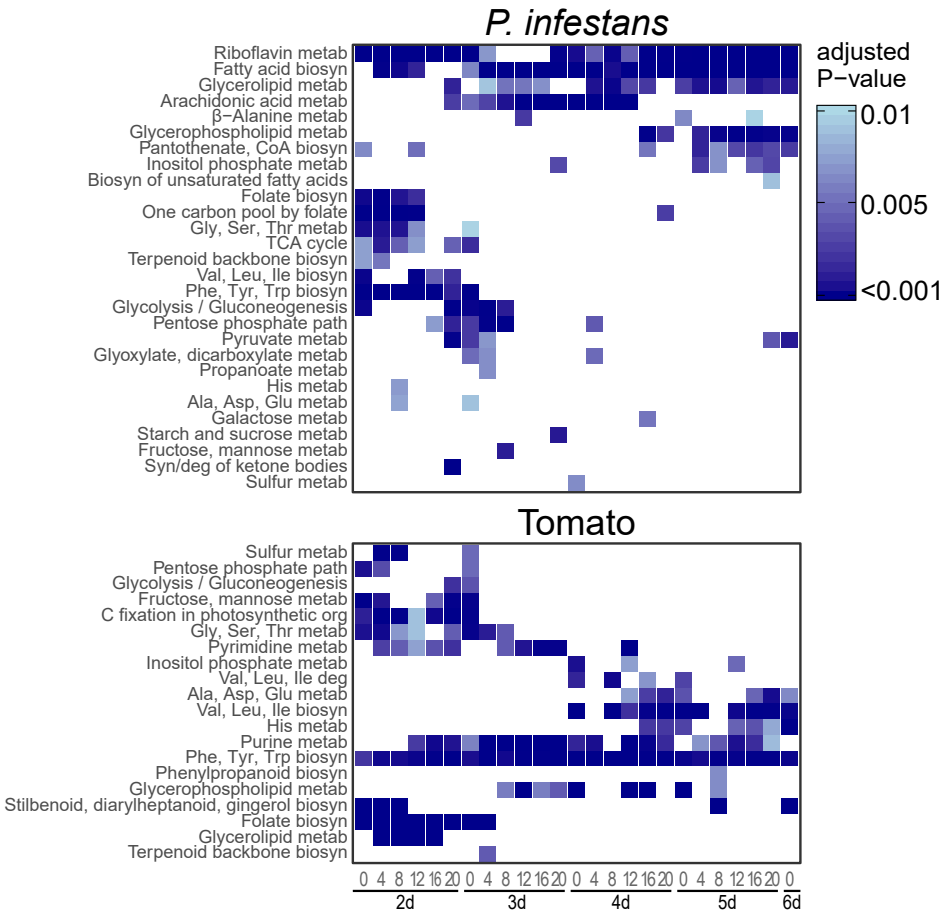


FIGURE 4 | Enriched KEGG pathways in transcriptome-based submodels from a time course covering a full infection cycle of *P. infestans* on tomato. Submodels are representative for the sampling time post inoculation shown on the X-axis in days (d) and hours. Enrichment is calculated based on the reaction content of each of the submodels compared to the full model. Colors scale to the adjusted P-value.

Import of specific nutrients becomes increasingly essential to *P. infestans*

The ability of a model to maintain its functionality (biomass production) under perturbations (e.g. simulated reaction deletions) is often referred to as robustness (Peyraud *et al.*, 2018). We can express the robustness of each submodel as the fraction of reaction deletions that do not disturb *P. infestans* biomass production (Peyraud *et al.*, 2018). This revealed that early infection submodels have higher robustness to reaction deletions than late infection submodels (**Figure 5**). This suggests that as infection progresses, *P. infestans* largely shuts

down accessory/alternative pathways to synthesize its biomass precursors, rendering the remaining reactions essential. The number of essential genes for *P. infestans* remained stable over time, with an average of 84±5 essential genes (**Table S1**).

To evaluate the importance of nutrient transport for *P. infestans* while colonizing tomato, we assessed the essentiality of the host-pathogen transport reactions in each submodel (Zhang *et al.*, 2018). A host-pathogen transport reaction is essential when its deletion disables biomass production, and partially essential when deletion in combination with deletions of other reactions disables biomass production. We found no essential nutrients for *P. infestans* when considering the full model, yet the submodels display varying patterns of nutrient essentiality (**Figure S3**). Early infection submodels have just a few essential nutrients, while mid and late infection submodels, after the first transition point, show various essential amino acid transport reactions. This corroborates our previous observations (**Figure 4**) where the first transition point marks the switch from *de novo* synthesis of amino acids in *P. infestans* to an increased uptake from tomato.

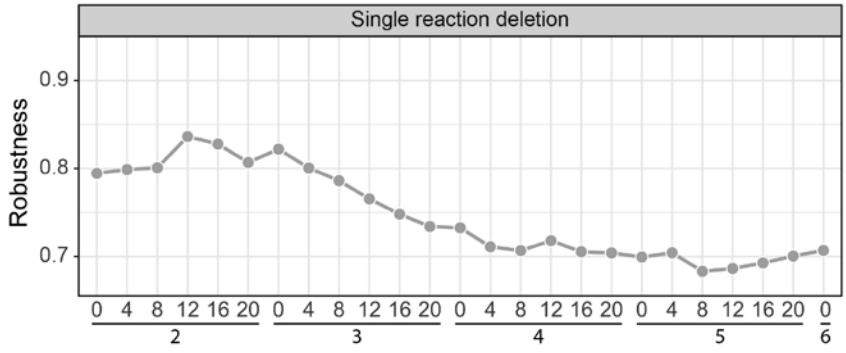


FIGURE 5 | Robustness of transcriptome-based submodels from a time course covering a full infection cycle of P. infestans on tomato. The Y-axis shows the robustness as the fraction of single reaction deletions that do not disable the biomass flux for P. infestans. Submodels are representative for the sampling time post inoculation shown on the X-axis in days and hours.

Metabolomics can be used to refine transcriptome-based submodels

To assess to what extent the transcriptome-based submodels are coherent with metabolome data, we utilized untargeted metabolome data of tomato leaves colonized by *P. infestans* at 2d/12h and 5d/12h post-inoculation. To relate the detected metabolites to metabolic fluxes, we hypothesized that metabolites strongly decreasing in abundance (log_2-fold change < –2) between these two time points are produced in the submodel of 2d/12h, and metabolites that increase (log_2-fold change > 2) are produced in the submodel of 5d/12h (Figure S4). Subsequently, we generated two additional submodels using INIT (Agren *et al.*, 2012) based on the transcriptome data at 2d/12h and 5d/12h, while enforcing the presence of the detected metabolites in the respective submodels. The submodels of the transcriptome-only (T) and

the metabolome-guided (T+M) submodels differed in 4% (2d/12h) and 11% (5d/12h) of reactions and 3% and 7% of metabolites, respectively (**Figure 6**). Overall, the addition of metabolomics data lead to a net increase of reactions and metabolites in particular at the late submodels, suggesting that the T submodels are slightly too conservative; there may be more active metabolism than suggested by mRNA levels of enzyme encoding genes. It should be emphasized that GEMs and metabolomics data only provide a partial description of the cell metabolism at the different infection time points. Discrepancies may arise since transcriptome-based submodels do not take account of any post-transcriptional regulation or post-translational modification of enzymes, and miss-annotations and missing/blocked reactions potentially increases uncertainty in submodels. Moreover, since detected metabolites are not necessarily continuously metabolized, steady-state fluxes are not directly linked to measured metabolite abundances (Jang *et al.*, 2018). Vice versa, metabolic fluxes do not necessarily yield detectable metabolites, for example when reaction products are immediately consumed in downstream reactions. For predicting metabolic fluxes more reliably, higher resolution metabolome data and sufficient data points over time are prerequisites which could provide sufficient data points in time to calculate metabolite coefficients (Kleessen *et al.*, 2015). Nonetheless, our initial results suggest that the integration of high-resolution metabolome data of tomato infection has the potential to refine stage-specific patterns that are embedded in the joint metabolism of the *P. infestans*–tomato interaction.

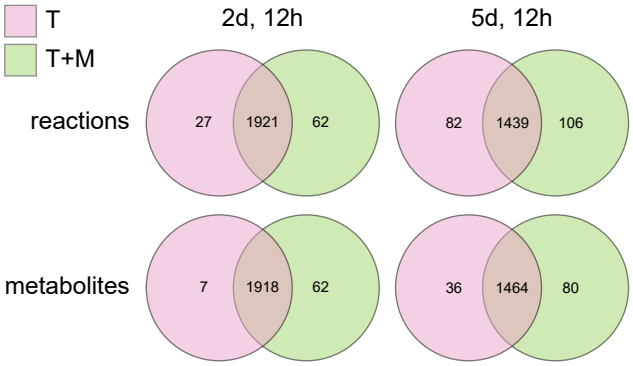


FIGURE 6 | Overlap of reaction and metabolite content of submodels, based on transcriptome data only (T) or transcriptome data and metabolome data (T+M). Data used to generate these submodels were obtained from *P. infestans*-infected tomato leaves at 2d/12h and 5d/12h post inoculation.

Conclusions

Pathogens scavenge metabolites from their hosts to support their growth and proliferation. Here, we took a systems biology approach and modelled the joint metabolism of tomato and *P. infestans* to generate hypotheses about their relationship. Our metabolic model of the *P. infestans*-tomato interaction represents one of the few integrated pathogen-host metabolic models published to date (Cesur *et al.*, 2018). Regarding pathogen and host as one entity

can yield hypotheses about the combined metabolism at a single metabolic equilibrium. The modeling allowed us to infer a conceivable pool of nutrients mainly consisting of amino acids, lipid precursors, and a TPP precursor, that is likely exploited by *P. infestans* while infecting tomato. The model also helped us to further characterize host-pathogen dependencies, such as the long known thiamine dependency of *Phytophthora* (Robbins, 1938). Modification of the thiamine biosynthesis pathway in tomato might be an interesting strategy for controlling late blight. Other interesting candidate pathways for pathogen intervention are its lipid metabolism, as our model predicts *P. infestans* takes up lipids as a membrane precursor, coherent with previous experiments (Griffiths *et al.*, 2003a). Alternatively, the fatty acid biosynthesis likely serves an important role to synthesize fuel reserves for its spores (Judelson, 2017), and our analyses showed that the fatty acid biosynthesis is a largely linear pathway with high interdependency of participating reactions (Rodenburg *et al.*, 2018a).

Clearly, many parameters that can further improve this model are still unknown. For example, the biomass composition of *P. infestans* growing *in planta,* and different metabolic processes in different zones of a lesion and during the infection cycle (Peyraud *et al.*, 2019). As such the model would benefit from more extensive and in depth metabolomics analyses (Galeano Garcia *et al.*, 2018). There is also a need to extend our knowledge on the potential role of transporters in this system, as these play a key role in infection and are potential control targets (Blume and Seeber, 2018; Meier *et al.*, 2018). This could be done for example, by ^{13}C-flux spectral analyses to monitor nutrient fluxes and to validate transporter functions in the pathogen (Beste *et al.*, 2013; Savory *et al.*, 2018). Altogether, this model provides insights into *P. infestans*-tomato metabolism and serves as a stepping stone for the design of novel control strategies for this devastating pathogen.

Materials & Methods

Model reconstruction

To improve the predictive capabilities of the previously constructed *P. infestans* iSR1301 GEM (Rodenburg *et al.*, 2018a), we performed several literature- and protocol-based curations (Thiele and Palsson, 2010). As starting point we used our published genome-scale metabolic model (GEM) of *P. infestans* iSR1301 (Rodenburg *et al.*, 2018a). We removed the sink- and gap-filling reactions that were previously added (Rodenburg *et al.*, 2018a). The sink reactions of the general metabolites 'fatty acid', 'holo-acyl-carrier protein', and 'apoprotein' have no mass or formula, but still play a role in flux of the fatty acid biosynthesis and were therefore retained. Recently it was found that some metabolic enzymes of the glycolysis and serine biosynthesis *P. infestans* occur in the mitochondria (Abrahamian *et al.*, 2017). We manually corrected the gene-reaction associations accordingly and corrected their localization in the appropriate compartment. The KEGG reaction identifiers of the *P. infestans* GEM were matched to MetaNetX identifiers (Moretti *et al.*, 2016), and the associated reaction formula was retrieved. MetaNetX indicates

for each reaction whether it is mass-balanced or not. Accordingly, all reactions that were not mass-balanced in the model were removed from the model. Conversion of arachidonic acid to eicosapentaenoic acid was manually added according to literature, and added to the biomass precursors (Robinson and Bostock, 2015; Yilmaz *et al.*, 2017). Thiamine diphosphate was added to the set of *P. infestans* biomass precursors, to represent the thiamine auxotrophy of *P. infestans*. Water and proton metabolites were removed from the model, as they do not fulfill a meaningful function in the *P. infestans* GEM, but do in the tomato GEM where these simulate proton-pumps (Yuan *et al.*, 2016). The thermodynamic constraints of other reactions were by default inferred from KEGG maps. Reaction thermodynamic constraints were manually adjusted according to standard model reconstruction protocol (Thiele and Palsson, 2010), such that ATP/GTP-consuming reactions were unidirectional, except for ATP synthase (EC 3.6.3.14), nucleotide diphosphate kinase (EC 2.7.4.6) and succinate coenzyme A synthase (EC 6.2.1.4).

To identify possible missing enzymes for *P. infestans* in the model, we annotated KEGG enzyme orthologs (KOs) in 19 oomycete proteomes which were downloaded from FungiDB (June 02, 2018) (Basenko *et al.*, 2018) (**Table S1**), as previously described (Rodenburg *et al.*, 2018a). We selected enzymes present in at least two oomycetes and not yet in the model. The encoding protein sequences were aligned to the *P. infestans* T30-4 genome sequence (Haas *et al.*, 2009), masked for its annotated open reading frames, using tblastn (v2.2.31+). Sequences with E-value <= $1e^{-50}$ and query coverage > 90% were submitted to the LocTree3 webserver (Goldberg *et al.*, 2014) to predict their subcellular localization, and the associated reactions were added to the model. Additionally, *P. infestans* gene models were manually corrected and re-annotated for enzyme orthologs (Rodenburg *et al.*, 2018a). To avoid spurious addition of reactions, any of the added reactions described above were removed if they were eventually found unable to carry flux in the model.

The tomato GEM was inferred from supplementary files of the associated publication (Yuan *et al.*, 2016). Since the tomato GEM *iHY3510* was built by a different lab using different reconstruction methodology, we tried to keep adaptations to a minimal level. Like the method applied to the *P. infestans* model, the reaction identifiers of the tomato model were matched to MetaNetX reaction identifiers for which the associated formula was retrieved. According to knowledge of thiamine biosynthesis in plants (Goyer, 2010), the thiamine biosynthesis was manually curated, such that thiamine is formed in the tomato cytosol from a plastidial thiazole and pyrimidine precursor (Guan *et al.*, 2014). The *P. infestans* and tomato GEMs were connected by unidirectional transport reactions for all metabolites shared between the tomato cytosol and the *P. infestans* cytosol compartments (**Table S1**).

Constraint-based modelling and optimization

Modelling was performed in MATLAB (R2017b) using the RAVEN (v2) (H. Wang *et al.*, 2018) and COBRA (v3) toolboxes (Heirendt *et al.*, 2017). Flux Coupling Analysis (FCA) was performed using F2C2 (v0.91) (Larhlimi *et al.*, 2012). Optimization was performed using Gurobi (v8.0) for

Flux Balance Analysis (FBA) and Integrative Network Inference for Tissues (INIT) (Agren *et al.*, 2012). GLPK (v2.8) was used for FCA. Essential genes/reactions and essentiality scores of the transport reactions were calculated using ESS (Zhang *et al.*, 2018), which implements Fast-SL (Pratapa *et al.*, 2015), constraining the total flux of *P. infestans* biomass production to > 5%. The global robustness statistic was calculated according to Peyraud et al. (2018) defined as the fraction of reaction deletions that did not render the biomass flux of *P. infestans* < 5%.

RNA sequencing and mass spectrometry

RNA and metabolites were isolated from tomato leaflets (cultivar New Yorker) inoculated with *P. infestans* isolate 1306. For the plant infection assays a zoospore suspension was prepared as described previously (Ah-Fong *et al.*, 2017a) and adjusted to a concentration of 5×10^4 per ml, applied to detached leaves placed on 1.5% water agar. The zoospores were sprayed on the leaves with a hand sprayer until run-off. The leaves were then incubated at 18°C under high humidity in plastic bags containing wet paper towels with a 12-hour light/dark cycle. Between two to six days post-inoculation leaves were harvested every four hours and flash-frozen until further use. After library construction, single-end 75 nt sequence reads were then obtained using an Illumina Nextseq 500, and quality was assured using FastQC (v0.11.8) (**Figure S2**) (Andrews, 2010). The reads were independently mapped to the *P. infestans* T30-4 (Haas *et al.*, 2009) and the tomato ITAG2.3 (Sato *et al.*, 2012) genomes using HiSat2 (v2.1.0) (Kim *et al.*, 2015), and mapping efficiencies were retrieved using samtools flagstat (v0.1.19) (Li *et al.*, 2009). Transcript abundance was quantified and normalized using cuffnorm (v2.2.1) (Trapnell *et al.*, 2010), which implements the DESeq2 normalization procedure (Robinson and Oshlack, 2010). This method divides the read counts by a factor that is calculated from the median of geometric means across samples, accounting for differences in sequencing depth and RNA composition.

For the metabolome analyses the flash-frozen leaves were lyophilized, weighed, and provided to Metabolon Inc. for further handling. After tissue grinding, proteins were removed by methanol precipitation. After eliminating the solvent, the samples were analyzed by reverse phase (RP) ultraperformance liquid chromatography (UPLC)-tandem mass spectrometry (MS/MS) using positive ion mode electrospray ionization (ESI), RP/UPLC-MS/MS with negative ion mode ESI, and hydrophilic interaction/UPLC-MS/MS with negative ion mode ESI. The area-under-the-curve method was used to quantify peaks. For each time point four biological replicates were analyzed. Differential abundance of metabolites was determined using t-tests, selecting metabolites with a $\log_2$ fold change of > 2 or < -2 with a Benjamini-Hochberg adjusted P-value of < 0.05 (Hochberg, 1995). We imputed missing values among replicates using the minimal value across other replicates.

Submodel generation

The transcriptome-based submodels were generated using the INIT algorithm (Agren *et al.*, 2012), which poses a mixed integer linear optimization problem (MILP) that aims to optimize

a global score based on reaction weights in the model. Reaction weights were calculated according to an adapted version of the formula described in Agren et al. (Agren *et al.*, 2012):

$$w_{ij} = 5log\left(\frac{E_{ij}}{\overline{x_i}+1}\right)$$

where w_{ij} is a weight for each reaction and condition j in the full model, E_{ij} is the transcript abundance value summarized per reaction (i.e. the maximum expression value of the genes associated to each reaction), and $\overline{x_i}$ is the mean expression of that reaction across samples. An inflation factor (+1) was added in the denominator to prevent inflated weights for extremely low expression values. According to Agren et al. (Agren *et al.*, 2012), we maintained the same minimum and maximum cutoff for $w-5 \leq w_{ij} \leq 10$. Reactions with an expression lower than the average (thus negative weight) will be less likely to be included in the submodels, and reactions expressed above average (thus a positive weight) will be more likely to be included. Transport reactions were assigned a weight of –0.1, based on the hypothesis that these should not be 'free' to include (weight 0), but should only have a minimal influence on the submodel solution. Reactions without associated genes were given a weight of 0. For the MILP optimization, the flux of all biomass precursors of both tomato and *P. infestans* was constrained to be > 1.

Data availability

Transcriptome data are deposited in the NCBI Short Read Archive under BioProject PRJNA516028.

4

Acknowledgements

This work was partly funded by the Food-for-Thought campaign from Wageningen University Fund, by a VENI grant (M.F.S.) from the Netherlands Organization for Scientific Research, and by awards to HSJ from the National Institute of Food and Agriculture of the United States Department of Agriculture and the National Science Foundation of the United States.

Supplemental Information

FILE S1

Basic principles of metabolic modelling

The reconstruction of genome-scale metabolic models (GEM) yields large networks of interconnected biochemical reactions, typically inferred from homology with known metabolic enzymes in biochemical databases such as KEGG (Kanehisa *et al.*, 2015), MetaCyc (Caspi *et al.*, 2014) or BRENDA (Placzek *et al.*, 2017). This may be complemented with general or species-specific biological knowledge of metabolism. Reactions in a GEM have a certain stoichiometry, i.e. the balance of metabolites that are consumed and produced in the reaction. For instance, in the reaction of acetyl-CoA C-acetyltransferase (EC 2.3.1.9):

1 CoA + 1 acetoacetyl-CoA ⇄ 2 acetyl-CoA

one CoA and one acetoacetyl-CoA are consumed and converted into two acetyl-CoA molecules or vice versa. A downstream reaction may again consume acetyl-CoA and convert it into another metabolite. In a GEM, the stoichiometry of each reaction is stored in a stoichiometric matrix $\boldsymbol{S}$, which is sparse and has dimensions m x n, where m is the number of metabolites and *n* is the number of reactions . The value of each entry S_{ij} is the stoichiometry of metabolite i in reaction j. When j is consumed in j, S_{ij} is given a negative value, when it is produced a positive value.

Each reaction in a GEM has a flux value v_j, which is defined as the rate of that reaction in steady-state. Steady-state is (often) the central assumption in a GEM, based on the idea that in the long run, metabolites in a cell cannot accumulate or deplete, since cells must maintain homeostasis to proliferate (J. Nielsen, 2017). In a GEM, the steady-state assumption implies that the net production and consumption of all metabolites in the model is zero. In other words, the net uptake of nutrient mass into the system must also be produced in the form of biomass.

In mathematical terms, the steady-state of a GEM is described as:

$$S \cdot v = 0$$

where $\boldsymbol{v}$ is the vector of all flux values. Generally, solving this system of equations yields a large number of solutions; that is, it does not result in a single, unique set of fluxes $\boldsymbol{v}$. Therefore, it is often additionally assumed that the cell attempts to optimize a certain objective, such as maximize growth, minimize energy consumption etc. (Lewis *et al.*, 2012). Given a GEM, one can then find the set of fluxes that optimize the selected objective. This can be posed as a linear optimization program:

$$\text{Maximize: } v_{biomass}$$
$$\text{Subject to: } S \cdot v = 0$$
$$lb_j \leq v_j \leq ub_j$$

Here $v_{biomass}$ specifies the flux of one or more biomass reactions that consume the precursor metabolites for all main cell components. The optimal flux values in the solution are limited by the steady-state constraint explained earlier, and thermodynamic constraints that specify for each flux upper- and lower bounds, lb and ub respectively. lb_j can be less than zero, indicating a reverse flux is allowed, representing bidirectional or reverse reactions. Thermodynamic constraints are usually applied for certain reactions by specifying that a reaction may only occur in one direction, thus setting lb or to zero.

Linear optimization can be performed by specialized software, such as Gurobi, CPLEX, Mosek or GLPK, that can optimize a GEM consisting of thousands of reactions in a matter of seconds.

FILE S1 | *Continued.*

Flux coupling

Flux coupling analysis (FCA) is a method that finds coupled fluxes in a GEM. Three basic types of flux coupling were developed (Burgard *et al.*, 2004; Larhlimi *et al.*, 2012). FCA finds for each pair of unblocked fluxes in the model v_a and v_b whether they are:

I) Directionally coupled: $v_a \neq 0 \rightarrow v_b \neq 0$. In general terms, any nonzero flux for v_a implies nonzero flux for v_b but not necessarily the reverse. This coupling can hold independently for the other direction: $v_b \neq 0 \rightarrow v_a \neq 0$.

II) Partially coupled: $v_a \neq 0 \leftrightarrow v_b \neq 0$, but not at a constant ratio $v_a / v_b \neq c$.

III) Fully coupled: $v_a \neq 0 \leftrightarrow v_b \neq 0$ and in a constant ratio $v_a / v_b = c$.

In any case, a coupling between two fluxes in a steady-state GEM indicates that flux of one reaction at least partially relies on another, indicating a functional relationship between the associated reactions.

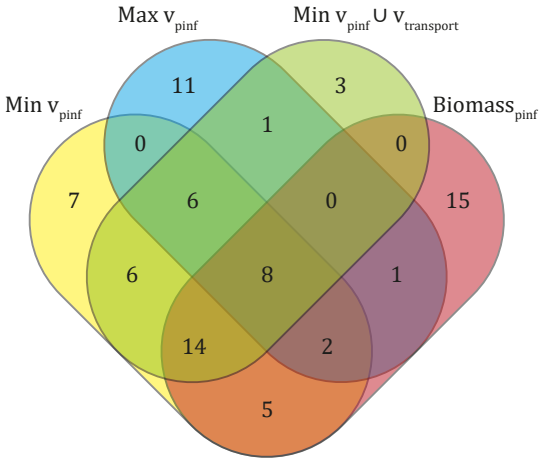


FIGURE S1 | Overlap in imported nutrients in three flux scenarios of different objective functions. $Biomass_{pinf}$: maximization of flux of *P. infestans* biomass precursors; Min v_{pinf}: scenario **II**, minimization of *P. infestans* flux; Max v_{pinf}: scenario **III**, maximization of *P. infestans* flux; Min v_{pinf} U $v_{transport}$: scenario **IV**, minimize both *P. infestans* and host-pathogen transport flux. The biomass precursors and imported nutrients in each scenario are listed in **Table S2**.

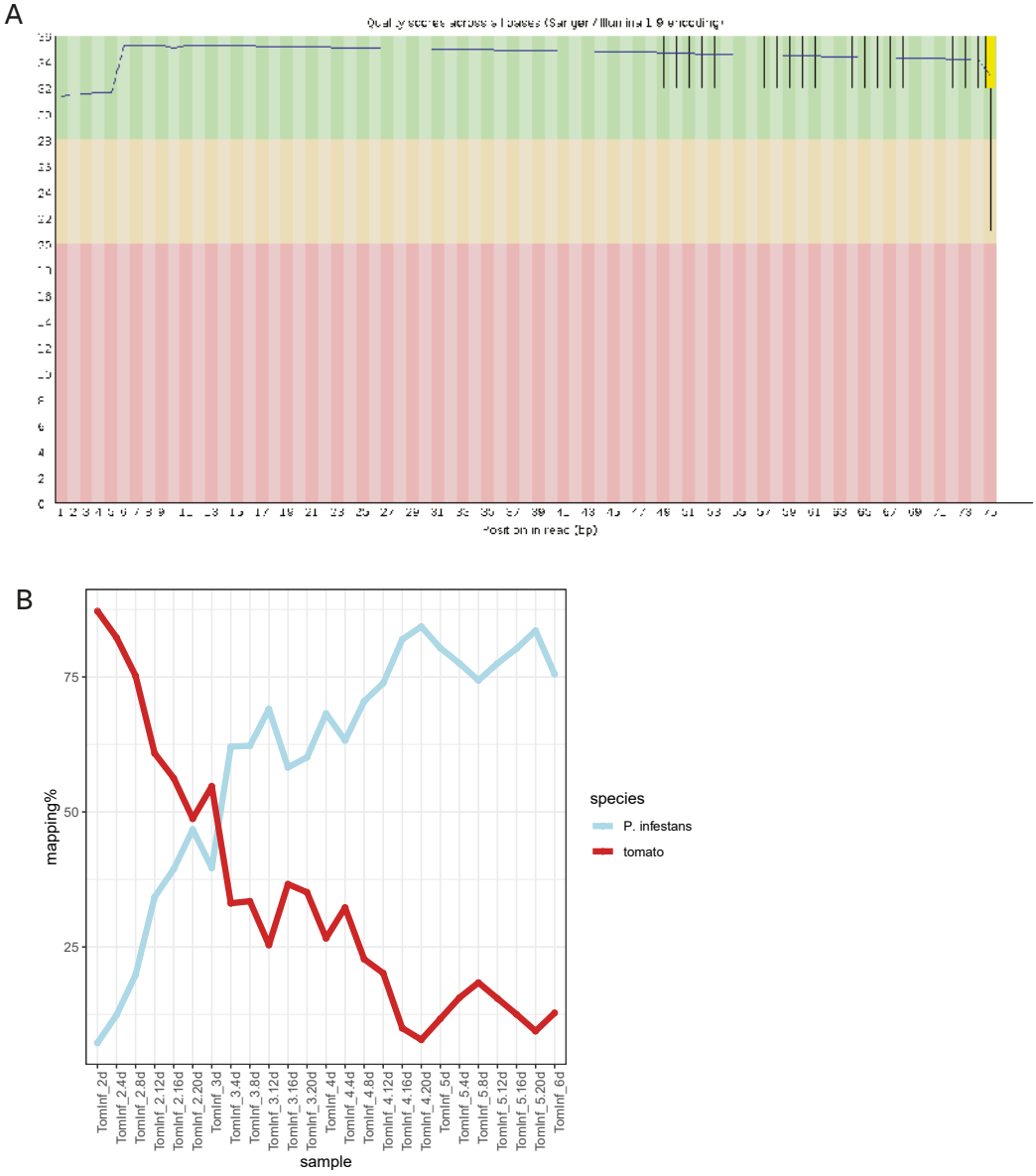


FIGURE S2 | Statistics of the RNA-Seq read mapping. **A:** Per-base-quality values for the Illumina reads (all samples combined), derived from the FastQC software **B:** Mapping efficiency per sample for *P. infestans* and tomato.

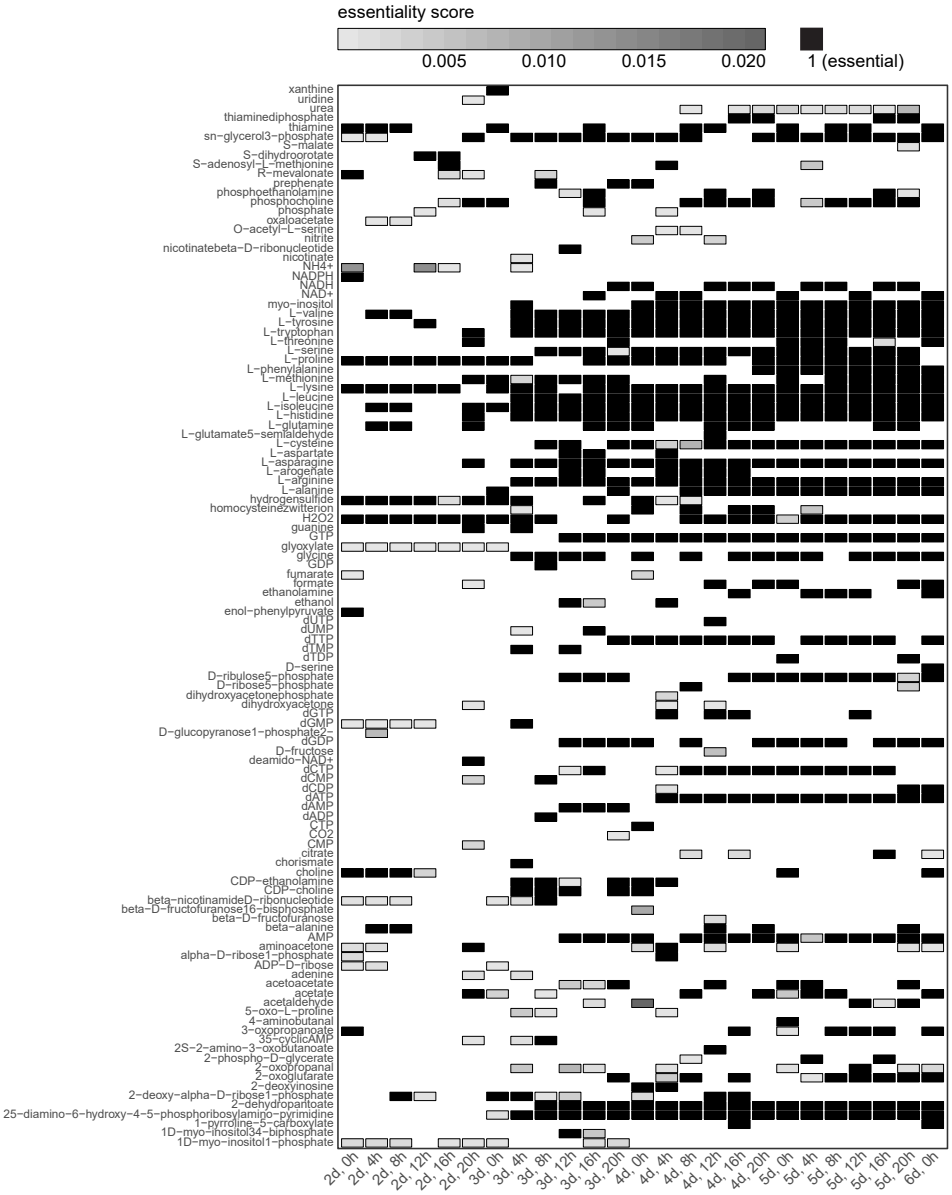


FIGURE S3 | Essentiality score (ESS) for each host-pathogen transport reaction per submodel. The X-axis represents transcriptome-based submodels, the Y-axis indicates the nutrient that is transported. The transport reactions of nutrients indicated in black are fully essential and of those indicated in various shades of gray are partially essential.

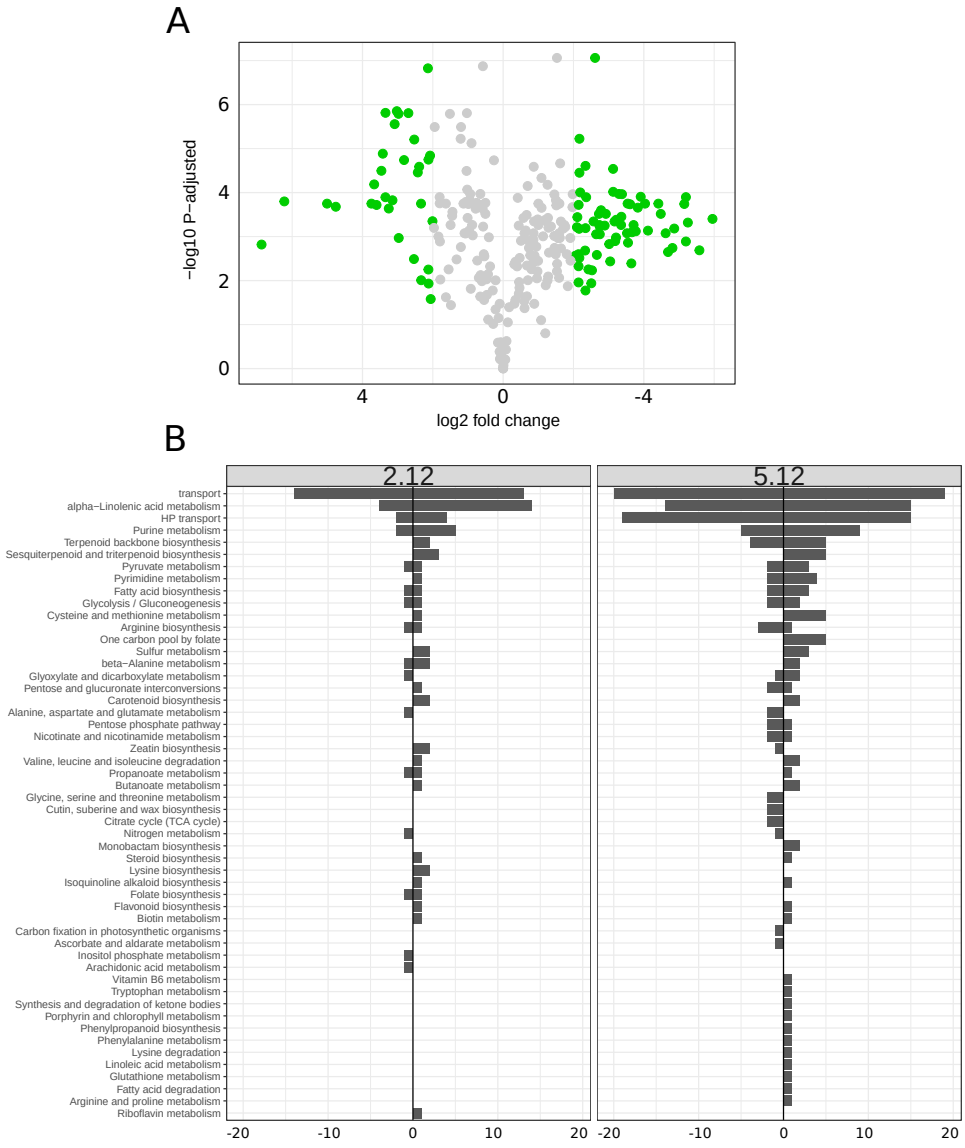


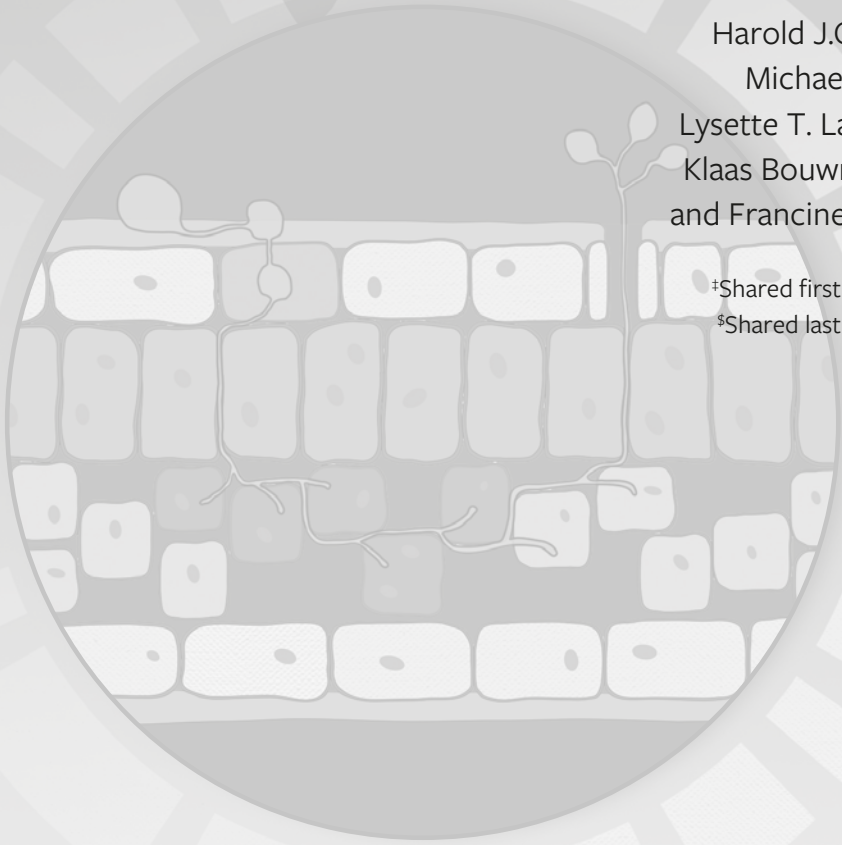
FIGURE S4 | Integration of metabolome data in the model. **A:** Volcano plot of the metabolome data, highlighting metabolites that show a significant $\log_2$ fold change of < -2 or > 2 between two time points post inoculation, 2d/12h and 5d/12h. X-axis represents the $\log_2$ fold change, Y-axis the significance with the $-\log_{10}$ P-value. **B:** Adding metabolome data leads to an increase or decrease of reactions in various pathways. The analysis was done on two submodels, 2d/12h (left) and 5d/12h (right). Negative bars indicate reactions that were missing in the metabolome-guided (T+M) submodels, compared to the transcriptome-only (T) submodels, and positive bars indicate reactions that were added.

Chapter 5

Mining oomycete proteomes for metalloproteases leads to identification of candidate virulence factors in *Phytophthora infestans*

Charikleia Schoina[‡],
Sander Y.A. Rodenburg[‡],
Harold J.G. Meijer,
Michael F. Seidl,
Lysette T. Lacambra,
Klaas Bouwmeester[$]
and Francine Govers[$]

[‡]Shared first authorship
[$]Shared last authorship



Abstract

Pathogens deploy a wide range of pathogenicity factors to modify host tissue or manipulate host defenses, among which a plethora of proteases. Metalloproteases (MPs) have been implicated in virulence in several animal and plant pathogens. Here we investigated the repertoire of MPs in 47 stramenopile species including 37 oomycetes, five diatoms and four brown algae. Screening of their proteomes using hidden Markov models (HMMs) trained for MP detection resulted in over 4000 MPs, with most species having between 65 and 100 putative MPs. Classification in clans and families according to the *MEROPS* database showed a highly diverse MP repertoire in each species. Analyses of domain composition, orthogroup clusters, and distribution and abundance within the oomycete and stramenopile lineage revealed a few oomycete specific MPs and others potentially related to lifestyle. In-depth analyses of MPs in the plant pathogen *Phytophthora infestans* revealed 91 MPs, divided over 21 families, including 25 MPs with a predicted signal peptide or signal anchor. Expression profiling showed different patterns of MP gene expression during pre-infection and infection stages. Of 27 MPs tested in plant assays, three inhibited and eight promoted lesion growth of *P. infestans*. To our knowledge, this is the first systematic inventory of MPs in oomycetes and the first study pinpointing MPs as potential pathogenicity factors in *Phytophthora*.

Introduction

Proteases are enzymes that catalyze the breakdown of proteins into smaller polypeptides or single amino acids and play important roles in numerous biochemical processes. They are crucial components of the molecular and cellular machinery in all organisms, not only within the organism itself but also in symbiotic interactions with other organisms. Pathogenic microbes, for example, often produce a variety of proteases that they use for degrading or modifying host tissue, or for disrupting the basic cellular machinery in the host to create suitable conditions for successful colonization (Bellincampi *et al.*, 2014; Miyoshi and Shinoda, 2000; Monod *et al.*, 2002).

Proteases, also referred to as proteinases, peptidases or proteolytic enzymes, can be divided in seven major groups according to their catalytic types: namely aspartic-, cysteine-, glutamic-, metallo-, asparagine lyase, serine and threonine proteases. In addition, there are proteases of mixed or unknown catalytic type. *MEROPS* is a peptidase database currently containing over 4000 entries that groups proteases into families based on amino acid sequence similarity. At a higher hierarchical level, homologous protein families are grouped into clans (Rawlings *et al.*, 2016; Rawlings *et al.*, 2018; Rawlings and Morton, 2008). As such, the *MEROPS* classification is an excellent tool for categorizing proteases resulting from genome annotations and for predicting potential functions of validated proteases.

Metalloproteases (MPs) function by virtue of a divalent metal cation positioned at their catalytic site. The majority is zinc-dependent and this is accommodated by two zinc-binding histidine residues in the HEXXH motif that is conserved in over 50% of the MPs (Rawlings and Barrett, 1995). In animals, MPs are implicated to play a role in receptor modification, by a process referred to as ectodomain shedding. These so-called sheddases include several members of the A Disintegrin And Metalloproteases (ADAMs) and Matrix MetalloProteases (MMPs) families (Hayashida *et al.*, 2010; Higashiyama *et al.*, 2011). They shed the extracellular part of membrane proteins resulting in a truncated receptor and an extracellular protein, leading to loss of function of the receptor or alteration of downstream signaling mediated by the receptor or the extracellular protein (Sanderson *et al.*, 2006).

Pathogenic microbes may exploit MPs as pathogenicity factors (Deu, 2017; Miyoshi and Shinoda, 2000). For example, leishmanolysin GP63, a zinc-binding MP in protozoan *Leishmania* species, was shown to have proteolytic activity on protein-tyrosine phosphatases of macrophages during infection and to enhance migration of the pathogen through the extracellular matrix (Gomez *et al.*, 2009; McGwire *et al.*, 2003). Deletion of leishmanolysins in *Leishmania major* resulted in reduced lesion formation in mice, thereby identifying GP63 as a virulence factor (Joshi *et al.*, 2002). In the mammalian pathogen *Vibrio cholerae*, the extracellular Zn-dependent MP hemagglutinin, also known as vibriolysin, is involved in the modification of toxins, degradation of mucus barriers, and disruption of host intestinal junctions (Benitez and Silva, 2016), and in a nematicidal strain of *Bacillus thuringiensis* the MP ColB was found to

play a role the colonization of nematodes (Peng *et al.*, 2015). The fungal pathogen *Aspergillus fumigatus* uses its MP Mep1p to cleave major complement proteins and pattern recognition molecules that function in defense. Mep1p thus facilitates early immune evasion by disarming defense in the human host (Shende *et al.*, 2018).

Also in plant pathogens MPs have been implicated in pathogenicity (Franceschetti *et al.*, 2017). For example, the rice blast fungus *Pyricularia oryzae* (*Magnaporthe oryzae*) has a MP-like protein that directly interacts with the rice resistance protein Pi-ta and is recognized as the avirulence factor AVR-Pita (Jia *et al.*, 2000; Jia *et al.*, 2016; Orbach *et al.*, 2000). More recently, comparative genome analyses of blast fungi revealed a novel *Pyricularia*-specific MP family that is expanded in several *Pyricularia* species and includes potential effector proteins (Gómez Luciano *et al.*, 2019). The fungus *Fusarium oxysporum* f. sp. *lycopersici* produces the MP Mep1 that acts in concert with the serine protease Sep1 to cleave chitinases produced by the host as part of the defense machinery. Both, *FoMep1* and *FoSep1* are required for full virulence of *F. oxysporum* on tomato (Jashni *et al.*, 2015). In the fire blight bacterium *Erwinia amylovora* (renamed *Pectobacterium*) the lack of PrtA, a secreted extracellular zinc-binding MP, as well as inhibition of PrtA activity with EDTA, results in reduced host-plant colonization (Zhang *et al.*, 1999). Similarly, *PrtA* mutants of *Burkholderia glumea,* a bacterial pathogen on rice, show significantly reduced virulence (Lelis *et al.*, 2019).

Here we focus on MPs in oomycetes, a diverse class of eukaryotic microbes that belong to the Stramenopiles, a lineage also comprising diatoms and brown algae. Most oomycetes described so far are plant pathogens among which *Phytophthora* spp., *Pythium* spp. and the downy mildews (Kamoun *et al.*, 2015). The latter are obligate biotrophs that entirely depend on their hosts to survive while *Pythium* spp. and *Phytophthora* spp. have a (hemi)biotrophic or necrotrophic lifestyle and can be cultured *in vitro*. Some *Pythium* species are pathogens of animals or parasitize on other microbes (mycoparasites) while *Saprolegnia* is a genus that mainly comprises animal pathogens (Jiang *et al.*, 2013).

The most (in)famous oomycete is *Phytophthora infestans*, the late blight pathogen that made its first appearance in Europe in the mid 19th century and ravaged the potato crop, culminating in a famine, in particular in Ireland. Even to this date *P. infestans* causes large economic losses in potato and tomato production worldwide. Farmers heavily invest in chemical spraying to protect the canopy against late blight (Haverkort *et al.*, 2008). *P. infestans* is a hemibiotrophic pathogen that depends on living host tissue during the first stages of infection. Therefore, avoidance of recognition by the host is important for successful penetration of host cells and subsequent colonization. Genome sequencing has revealed that *P. infestans* has over 500 genes encoding cytoplasmic effectors and around 200 encoding apoplastic effectors (Haas *et al.*, 2009), several of which are now known to function in virulence by suppressing host defense responses (Whisson *et al.*, 2016). Apart from these effector genes, *P. infestans* has many other putative pathogenicity genes including genes encoding hydrolases, lipases, proteases and protease inhibitors (Haas *et al.*, 2009). Examples are the carbohydrate-active enzymes (CAZymes) (Ospina-Giraldo *et al.*, 2010), phospholipase D's (Meijer *et al.*, 2011; Meijer

and Govers, 2006), and aspartic proteases (APs) (Kay *et al.*, 2011; Schoina *et al.*, 2019). The latter is a family with twelve members including one, PiAP5, that is a fusion of the AP domain and a G-protein coupled receptor domain, preceded by a signal peptide (SP) (Kay *et al.*, 2011). Such a peculiar combination of two subsequent protein domains is a typical example of a *Phytophthora*- or oomycete-specific bigram that is not found outside the oomycete lineage (van den Hoogen *et al.*, 2018; van den Hoogen and Govers, 2018). A comprehensive study of the protein domain organization in 67 eukaryotes revealed that especially *Phytophthora* spp. possess a relatively high proportion of multidomain proteins with unique bigrams, a feature that presumably provides novel functionality to proteins (Seidl *et al.*, 2011). In *P. infestans*, 259 genes coding for bigrams were found to be differentially regulated during infection, with the majority of those bigrams predicted to be secreted (Seidl *et al.*, 2011). Moreover, proteomic analyses of the *P. infestans* secretome revealed that besides effectors, several proteases were detected in the extracellular medium, including cysteine, aspartic and metalloproteases, confirming that these enzymes are secreted by the pathogen and could potentially play a role in the host-pathogen interplay (Meijer *et al.*, 2014).

While MPs have been shown to play various roles in cellular processes and in pathogenicity, they have hardly been studied in oomycetes. The aim of this study was to make an inventory of the full repertoire of MPs present in Stramenopiles, and to investigate the potential role of MPs in the virulence of *P. infestans*. With a bioinformatic search we identified over 4000 stramenopile MPs. We divided those in clans and families according to the *MEROPS* classification and studied their distribution and abundance within the oomycete and stramenopile lineage. Further analyses included orthogroup clustering aimed at identifying features potentially related to pathogenicity or lifestyle. We then zoomed in on the MP repertoire of *P. infestans* and consulted transcriptome data to determine the expression profiles of MP genes during the entire asexual life cycle, from sporulation on free-living hyphae to *in planta* growth. Finally, we selected *P. infestans* MPs predicted to be secreted and tested these for their ability to promote or inhibit the virulence. Taken together this research reveals the diverse repertoire of MPs in oomycetes and points to a few MPs that seem to play a role in pathogenicity of the late blight pathogen *P. infestans*.

Results

Oomycetes have a large diversity of metalloproteases

To gain insight into the metalloprotease (MP) repertoire and diversity in oomycetes, we used the *MEROPS* database as a starting point. This database divides proteases into families based on the amino acid similarity of the peptidase domain and groups homologous protein families into clans. The current release of the *MEROPS* database (release 12.1) recognizes 76 MP families, of which 70 families are grouped into a total of 16 clans whereas six are not assigned to a clan (Rawlings *et al.*, 2018). Families are coded by a prefix M followed by a number while

clans have a two-letter code, i.e., M followed by a second capital letter.

Putative MPs in oomycetes and other stramenopile species were identified using a bioinformatics procedure (**Figure 1**). We used all peptidase sequences retrieved from the *MEROPS* database to train profile hidden Markov models (HMMs) per *MEROPS* (sub)family. The HMMs were used to scan the predicted proteomes of 46 Stramenopiles, including 37 oomycetes, five diatoms and four algae (**Table S1**). Matching protein domains were realigned and used to train stramenopile-specific MP HMMs. In addition, Pfam peptidase domain HMMs were retrieved (Finn *et al.*, 2016) and all proteins matching these Pfam and/or *MEROPS* HMMs were considered putative MPs. A total of 4,178 MPs was identified in the entire set of stramenopile proteins ranging from 265 to 50 MPs per species (**Figure 2; Table S1**). On average, 91 MPs occur per species which corresponds to 0.60% of the total proteome, with the majority of the species (70%) having between 65 and 100 putative MPs (**Figure S1**). Outliers with vastly higher numbers are the crayfish pathogen *Aphanomyces astaci* (*n*=265, 1.39% of its proteome) and the diatom *Fistulifera solaris* (*n*=190, 0.94%), followed by the plant pathogen *Phytophthora parasitica* (*n*=134, 0.48%) and the animal and human pathogen *Pythium insidiosum* (*n*=127, 0.66%). Species with slightly lower MP numbers than average are the white rust *Albugo candida* (*n*=50, 0.38%), the alga *Nannochloropsis gaditana* (*n*=60, 0.55%) and a few plant pathogenic species in the genera *Phytophthora and Plasmopara*. In summary, we observe differences among species in the absolute number of MPs, also when accounting for differences in proteome size.

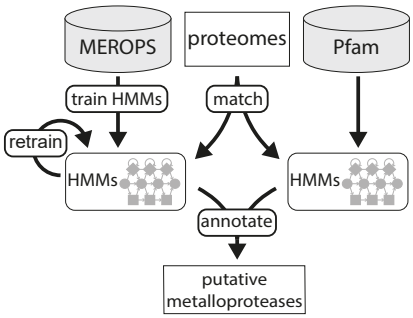


FIGURE 1 | Pipeline for the identification of putative metalloproteases in Stramenopiles.

The MPs identified in the 46 stramenopile species can be assigned to 37 families (out of 76 families in *MEROPS*) and 11 clans (out of 16) (**Figure 2**). Of the 37 families, more than half (21 or 57%) belong to the MA clan with the characteristic catalytic HEXXH motif. The other 10 clans are represented by one up to four families while families M79 and M82 are among the six families not assigned to a clan. When focusing on the oomycetes the overall presence/absence pattern of the MP families is similar for all species with a few remarkable exceptions. For instance, the M64 family seems to be specific for Pythiales and the two *Aphanomyces* species, whereas M97 is present in nearly all Pythiales species and most Peronosporales species but

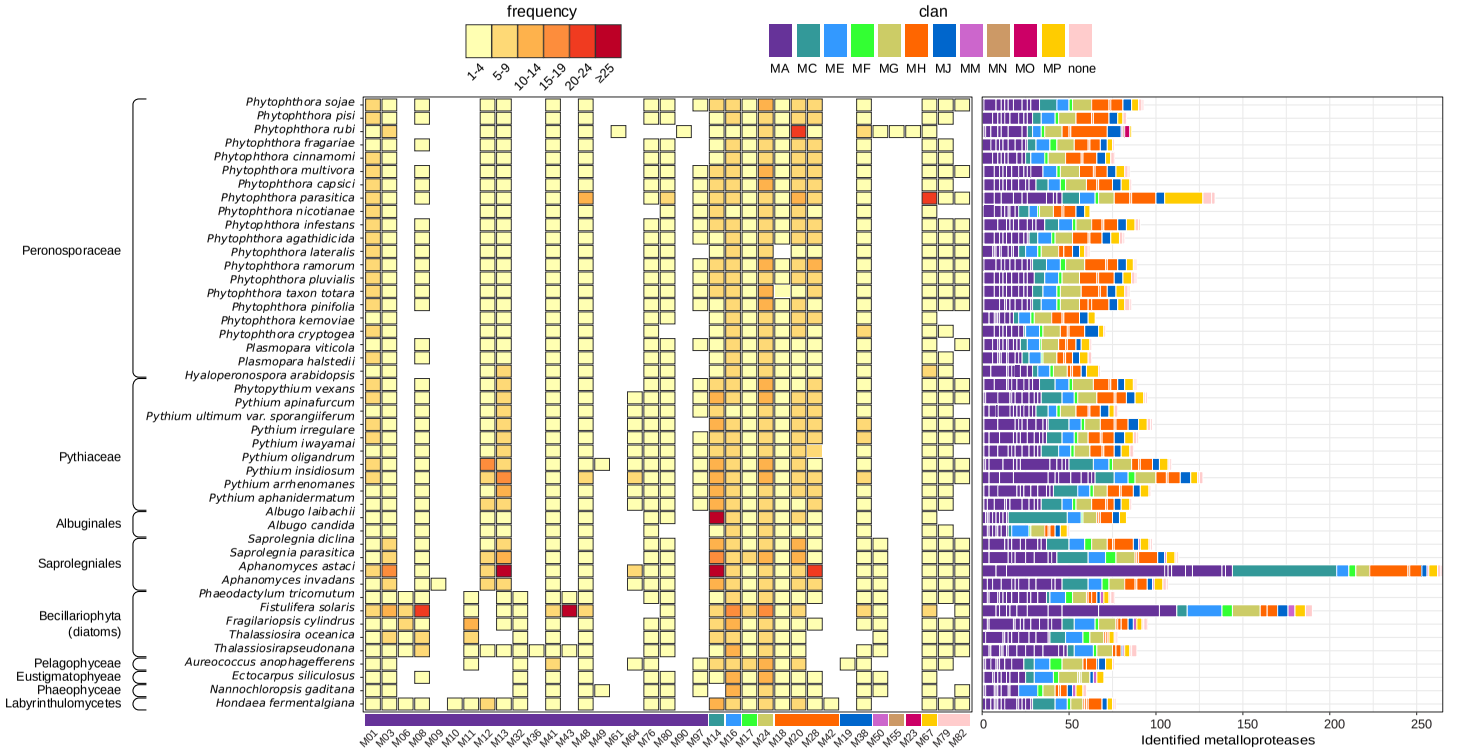


FIGURE 2 | Metalloproteases in Stramenopiles. MP families (lower bar) identified in 46 Stramenopiles, including 37 oomycetes, and divided over various taxonomic lineages (left). Approximate numbers of MP family members in each species are shown by frequency (center) and in bar charts (right). In line with the *MEROPS* database, the MP families are categorized in clans that are color-coded (top right).

is absent in the Albuginales and Saprolegniales. In contrast, M50 is a family that, within the oomycetes, is restricted to the Saprolegniales with one outlier. *Albugo laibachii* and *A. astaci* both have an exceptionally high number of MPs categorized as M14 while *P. parasitica* has a relatively large repertoire of M48, M80 and M67 family members. *Phytophthora rubi* seems to have a relatively high number of MPs in the M20 family and is the outlier for M50. One should be cautious though with this species. At first sight it seems to have four unique MP families (i.e., M61, M90, M55 and M23), but closer inspection revealed 100% identity to proteins of the bacterium *Pandoraea sputorum* suggesting potential contamination of the genome assembly (**Table S1**)**.**

When comparing oomycetes with non-oomycete species we observed a similar diversity in MP families. However, three families show up that are specific for the diatoms and algae included in this study (i.e. M06, M11 and M32) and one that is specific only for the diatoms (i.e. M43). Moreover, the M50 family is clearly not limited to the saprolegniales but more widely distributed within the Stramenopiles. The few families that were only found in a single species (M09, M19, M36 and M42) or limited to two species (M49) were subjected to a closer inspection. Unlike the situation in *P. rubi,* there were no indications for contaminations in the genome assemblies. However, to what extent these MPs are truly species-specific is questionable given the limited number of diatom, algae and *Aphanomyces* species included in this study.

Clustering of stramenopile metalloproteases in 85 orthogroups

To investigate the sequence similarity between the MPs in the dataset, we aligned and clustered the 4,178 MP sequences resulting in 85 orthogroups, which are defined as clusters of sequence-similar MPs (**Figure 3; Table S1**). The orthogroups contain on average 49 MPs, with a minimum of two MPs per orthogroup and a maximum of 287 suggesting that some MP families are rather large and often have multiple copies per species. The average number of species per orthogroup is 28. Two orthogroups contain MPs of just one species and thus are species-specific, while five orthogroups span all 46 species and thus contain one or more MPs from each species. Another nine orthogroups contain MPs derived from 45 out of a total of 46 species. Anticipating that a poorly assembled or miss-annotated genome can readily lead to the absence of just a single species in an orthogroup, we define the 14 orthogroups containing MPs from at least 45 species as the core MP repertoire of the Stramenopiles included in this study. This core comprises MPs from ten distinct *MEROPS* families, including five orthogroups from the M24 family (aminopeptidases). Overall, each of the 85 orthogroups contains a single *MEROPS* family but, as is the case with M24, some families are divided over multiple orthogroups. The more orthogroups are linked to a single family, the more sequence divergence there is within this family. The families M16 (MPP β-subunit) and M24 are each divided over 8 orthogroups. The largest MP family present in the dataset is M13 with 287 members divided over 42 species. They all cluster in one orthogroup, the largest one (#0), implying that there are only relatively small differences between the sequences. The M13 family mainly consists of single-domain proteins that are homologous to the human peptidase neprilysin, an MP that

cleaves peptides and is known to inactivate several peptide hormones (Rawlings and Salvesen, 2013). MP13 is present in all 37 oomycetes analyzed in this study, in most species in up to 5 copies. Strikingly, this MP is relatively more abundant in animal pathogens including *A. astaci* (91 copies), *Pythium insidiosum* (20), and *Saprolegnia parasitica* (11). Of the 4,187 MPs, 24 were not included in any orthogroup, among which the MPs from *P. rubi* that were mentioned earlier as likely contaminants.

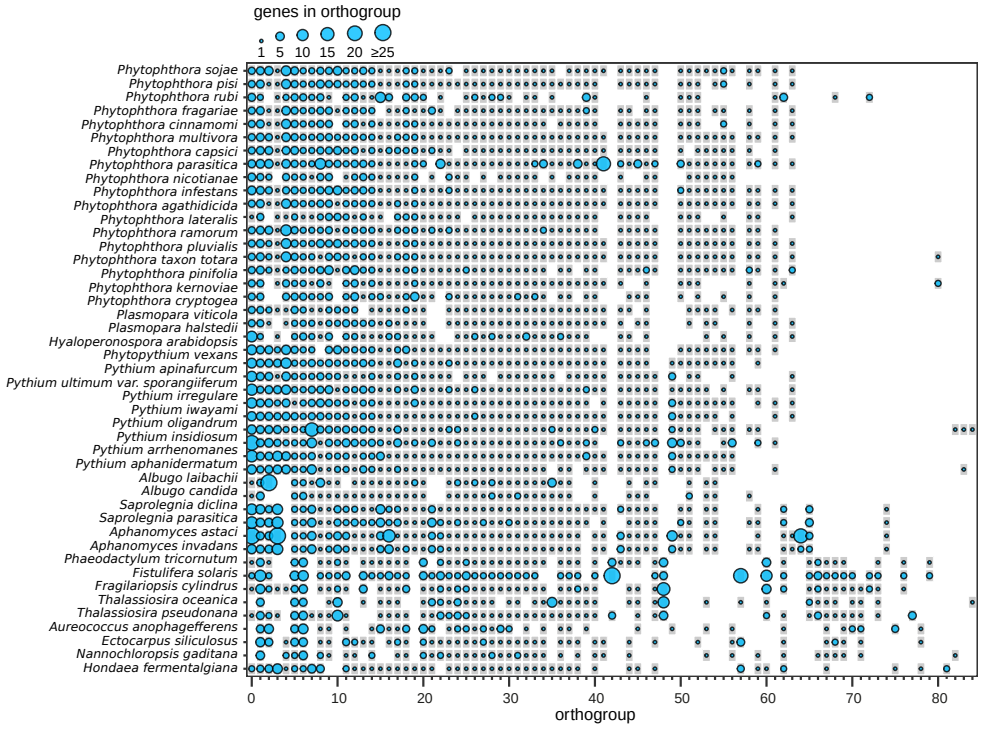


FIGURE 3 | Clustering of stramenopile MPs in orthogroups. Orthogroup distribution per species (left) and the approximate number of MPs per orthogroup in each species depicted by the sizes of the blue circles. Othogroup numbering starts with '0' (OG0000000) for the largest group with 287 MPs. The smallest orthogroups are OG0000082, OG0000083 and OG0000084 with just two MPs. For the gene IDs clustered in each orthogroup see Table S1.

Oomycete MPs have diverse domain architectures and N-terminal targeting sequences

To gain insight into the structural diversity of the Stramenopile MPs, we analysed their domain composition and more specially searched for the occurrence of MPs with distinct domain combinations or bigrams. These analyses showed that 38% of the 4,178 stramenopile MPs in our dataset contain at least one additional domain next to the protease domain. In those multidomain MPs we detected 217 different Pfam domains in addition to the MP domains (**Figure S2**; **Table S1**), and more than 250 bigrams, of which a vast majority occurs in only

one or a few species (**Figure S3**; **Table S1**). There is a small proportion of bigrams that is more common (around 10%) and present in over 80% of the 46 analysed Stramenopiles. The three most ubiquitous bigrams, occurring in nearly all species, consist of (I) the 'CAAX-prenyl protease N terminal, five transmembrane helices' domain (Pfam ID PF16491) and a M48 peptidase domain; (II) the 'ATPase associated with various cellular activities (AAA)' domain (PF00004) and a M41 peptidase; and (III) the 'mitochondrial processing peptidase (MPP) β-subunit' domain of family M16B combined with a 'Middle or third domain of peptidase_M16' (PF16187).

The average number of MPs with distinct domain combinations per stramenopile species is 34 and no clear differences were observed between taxonomic lineages (**Figure S4**). The highest number was identified in *Phytophthora lateralis* (58 bigrams) and the lowest in *Hyaloperonospora arabidopsidis* (21 bigrams). All *Phytophthora* and *Pythium* species have a relatively large number of M28 aminopeptidases with two flanking domains, a N-terminal 'PA (protein associated) domain' (PF02225) and a C-terminal 'transferrin receptor-like dimerization domain' (PF04253). The human ortholog of these proteins is the transferrin receptor (TfR), a cell surface receptor that assists in iron uptake into cells through a cycle of exo- and endocytosis of the iron transport protein transferrin (Lawrence, 1999). Ten bigrams were found exclusively in *Phytophthora* spp. The most frequent one, present in seven species is the 'mitochondrial processing peptidase (MPP) β-subunit' domain of family M16B combined with an N-terminal 'glycosyl hydrolase 10' domain (PF00331). An additional *Phytophthora*-specific bigram is a 'peptidase dimerisation' domain (PF07687) combined with either a 'Peptidase family M20/M25/M40' domain or a M20F domain. In *Pythium* species we found three lineage-specific bigrams. This included a M67C peptidase flanked by a PhoD-like phosphatase domain (PF09423) and a high number of 'MORN (membrane occupation and recognition nexus) repeats' (PF02493), 13 in the *Pythium aphanidermatum* MP and 17 in the *Pythium arrhenomanes* MP.

Metalloproteases can play a role in different compartments in the cell, for example in mitochondria or plastids, but they can also be embedded in membranes or secreted to the extracellular environment (Majsec *et al.*, 2017). To gain insight into their subcellular distribution we assessed all MPs for N-terminal targeting sequences using HECTAR, a method that was specifically developed for predicting the subcellular localization of proteins in Stramenopiles (Gschloessl *et al.*, 2008). Notably, in *A. astaci* and the diatom *F. solaris* - the two species in this study with the largest MP catalog - a strikingly large proportion of the MPs was predicted to have a signal peptide; 52% and 53% respectively, while the average among species is only 16% (**Figure S5**). The SP containing proteins were almost exclusively single-domain MPs belonging to several families (**Figure S6**). Additionally, few *Phytophthora* species seem to have slightly more MPs with a signal anchor, whereas other species have (almost) none. Conceivably, most of the diatoms and algae included in this study possess MPs predicted to be targeted to plastids. Remarkably, also some of the oomycetes have MPs with plastid targeting signals, possibly remnants indicative of photosynthetic endosymbionts that were lost during evolution (Wang *et al.*, 2017).

In summary, our analyses reveal remarkable diversity in stramenopile MPs not only with respect to the subcellular targeting but also in terms of domain composition. Many species have MPs containing unique bigrams next to MPs with common bigrams (**Figure S3**). Unique bigrams might have specific functions related to the lifestyle of a particular species. Testing this hypothesis warrants a more in-depth study of the MP repertoire at the single species level.

The MP repertoire in *Phytophthora infestans*

For further analyses at the single species level we choose to focus on *P. infestans*. The MP mining (**Figure 1**) resulted in the identification of 91 *P. infestans* genes encoding putative MPs (**Table S2**), equal to the average number per Stramenopile species. By carefully examining the predicted gene models in the genome browser, and inspecting the alignments with ESTs and RNA-Seq data, the majority of the 91 predicted gene models were found to be correct. The remaining ones were manually corrected. Based on the predicted peptidase domains we categorized the 91 MPs in 21 MP families and eight clans. More than one-third belongs to clan MA (36 MPs divided over ten families) and 15 group in clan MH (divided over three families; M18, M20 and M28). Three MPs from two families (M78 and M82) are not assigned to a clan (U: unassigned) and the remaining 37 are divided over six clans with only one MP family per clan but always more than one MP per family. Family sizes range from two (M17-clan MF) to nine (M24-clan MG) with intermediates of five (M38-clan MJ and M67-clan MP) and eight (M14-clan MC and M16-clan ME) (**Figure 2; Table S2**).

Consistent with the observations described above for the entire set of Stramenopile MPs, the *P. infestans* MPs are diverse with respect to their protease domains and domain architecture, even among MPs of the same family. An example is the M01 family with eight members, all having accessory domains. Of the 91 *P. infestans* MPs, 54 MPs only have the protease domain and no other protein domain, such as all M13 and M14 family members. The other 37 (41%) contain at least one accessory domain next to the protease domain. This can be another MP domain, as is the case in the M16 family with five out of eight members having one or two additional M16-like peptidase domains, or accessory domains not directly implicated in MP activity. The diversity in domain architecture among related MPs is nicely demonstrated in the M24 family in which two out of nine members have an additional M24-like domain ('aminopeptidase P, N-terminal domain'; PF05195), another two have a C-terminal 'zf-MYND-like zinc finger, mRNA binding domain' (PF15801), one has a C-terminal 'Creatinase/Prolidase N-terminal domain' (PF01321) and one (PITG_08392) even has three accessory domains. The three remaining M24 members are single domain proteins (**Figure 4; Table S2**).

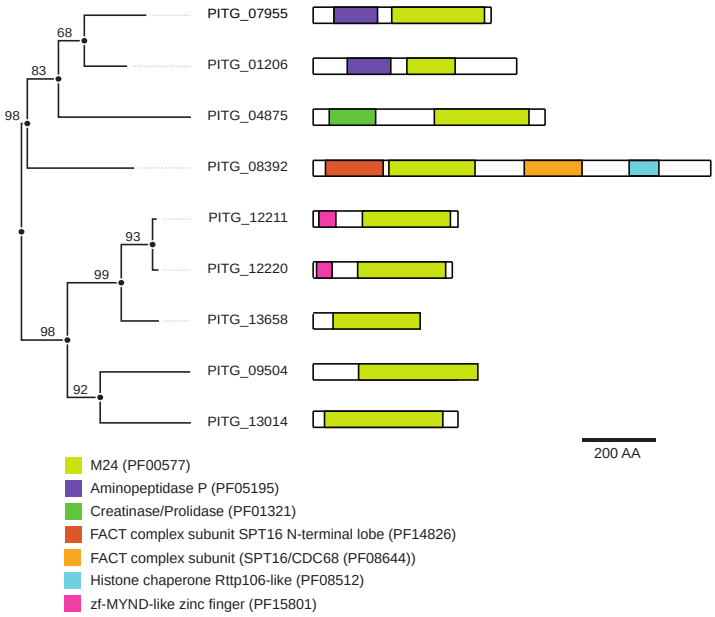


FIGURE 4 | Phylogenetic tree and domain composition of members of the M24 family in *Phytophthora infestans*. Multiple sequence alignment of M24 MP domains (lime colored) was performed with MAFFT v7 using the Mafft-homologs option and a gap opening penalty of 2.0. The phylogenetic tree was constructed with W-IQ-TREE using the WAG model and 1000 bootstraps. Bootstrap values are indicated at nodes.

Previously it was shown that *Phytophthora* spp. have proteins with unique, oomycete specific or even *Phytophthora* specific bigrams (van den Hoogen and Govers, 2018; Meijer and Govers, 2006; Seidl *et al.*, 2011), and also in the current study the analyses of all Stramenopile MPs revealed a few bigrams that might be lineage specific (**Figure S3; Table S1**). For *P. infestans*, most MPs with a similar domain architecture are present in metazoans, plants, fungi and/or bacteria, some of the multidomain MPs seem to be unique. Two examples are family M20 MPs (PITG_00577 and PITG_06978) in which the MP domain is preceded by a 'Amidohydro_1' domain (PF01979) or a 'CENP-B N-terminal DNA-binding domain' (PF04218), respectively. Proteins with these domain architectures are limited to oomycete species in the *Pythiaceae* and *Peronosporaceae* families. Another example is a family M67 MP (PITG_16722) that has a N-terminal 'PWWP' domain (PF00855). MPs with this domain architecture are also found in *Albugo* spp. and *Saprolegnia* spp. but not in non-oomycete lineages.

Expression profiling of *Phytophthora infestans* MP genes

To determine the expression profiles of the MP genes during the *P. infestans* life cycle we exploited transcriptome data from four *in vitro* life stages, i.e. zoospores, germinating cysts, mycelium and sporangia, and three infection stages, i.e. early, mid and late infection. The mean transcripts-per-million mapped reads (TPM) values were highly variable; eight genes

stand out with a mean TPM value above 500 pointing to a relatively high expression in one or more stages (**Figure 5**). Genes were empirically clustered into eight clusters with distinct expression profiles. More than one-third of the MP genes shows the highest transcript levels in the late infection stage (cluster 3 with 35 genes). Four clusters, i.e. clusters 1, 5, 6 and 8, show a clear peak in germinating cysts pointing to a relatively high expression in this pre-infection stage. In cluster 8 (eight genes) this coincides with a relatively high expression in sporangia and an overall lower expression in all other stages resulting in two peaks in the overall pattern. Cluster 6 also shows two peaks but in contrast to cluster 8, the genes in cluster 6 (17 genes) show lower expression in sporangia and higher expression in zoospores.

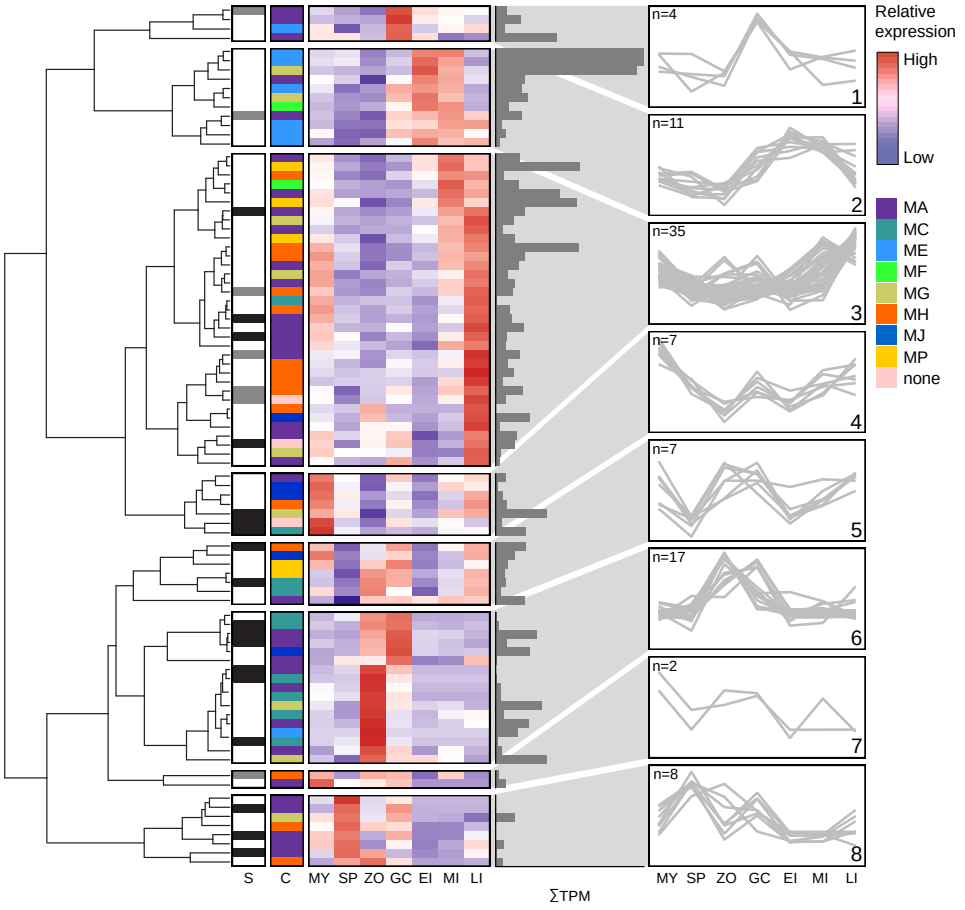


FIGURE 5 | Clustering of MP genes based on expression patterns during the *P. infestans* life cycle and during *in planta* growth. The dendrogram on the left shows the hierarchical clustering of the expression patterns. The bars in column S mark the MPs with a predicted signal peptide (dark gray) or signal anchor (light gray). The color-coded bars in column C indicate the clan in *MEROPS* comprising the MP. The heatmap shows the relative expression of each gene based on the stage-wise, z-score transformed expression values in mycelium (MY), sporangia (SP), zoospores (ZP), germinating cysts (GC), and early, mid and late infection stage (EI, MI, LI, respectively). The adjacent bar plot shows the mean TPM value for each gene. The line plots on the right display the expression profiles of the eight clusters and the number (*n*) of genes in each cluster.

Also, the seven genes in cluster 5 have higher transcript levels in zoospores and germinating cysts when compared to sporangia but here expression is also induced in mycelium and in the late infection stage. The expression of the four genes in cluster 1 peaks in germinating cysts and no other *in vitro* life stage or infection stage. Similar to the four cluster 1 genes the eleven genes in cluster 2 are clearly upregulated in germinating cysts but here the higher expression is maintained or even further boosted during early and mid infection stages. This cluster comprises the three genes with the highest average TPM value pointing to relatively high transcript levels.

The cluster analyses did not reveal any apparent correlation between expression pattern and specific features of the MPs. Each cluster contains MPs from various clans and families, and also MP genes that are predicted to encode potentially secreted MPs (designated sMPs) are randomly distributed over the clusters (**Figure 5**). Several of these sMP genes show relatively high expression levels in pre-infection stages such as (zoo)spores and germinating cysts, or during early, mid or late infection stages, expression profiles reminiscent of a role in pathogenicity.

MPs affecting lesion growth of *P. infestans*

To investigate a potential role for *P. infestans* MPs in causing late blight disease we selected a subset of the 91 MPs and tested their capacity to modulate the virulence of *P. infestans*. The selection was based on the presence of a N-terminal SP or SA in the predicted protein which initially resulted in a subset of 27 sMPs. A subsequent analysis using HECTAR (Gschloessl *et al.*, 2008) revealed three additional sMPs but discarded five of the initial 27 as likely not secreted (**Figure 6**). The vast majority are single domain proteins with only the protease domain preceded by a SP or SA. *In planta* activity was tested by monitoring lesion growth of *P. infestans* on *Nicotiana benthamiana* leaves in which full-length cDNAs of the MP genes were transiently expressed by means of agroinfiltration. Measurements of lesion sizes at five days post inoculation showed significantly larger lesions in nine cases pointing to a positive effect of these MPs on *P. infestans* growth. In three cases we recorded smaller lesions suggesting an inhibitory effect of these three MPs on growth.

The twelve MPs affecting lesion growth belong to only six families out of the 21 identified in *P. infestans,* and 8 out of the 12 belong to clan MA families. All members of family M12 and M79 (three and two, respectively) promoted lesion growth while two out of three M3 family members caused growth inhibition. Of the four tested M08 family members, only the one that has an EGF-like domain (PF07974) C-terminal of the protease domain showed a growth promoting effect (PITG_08874). It should however be noted that our initial analysis on this sequence predicted a N-terminal SA but HECTAR did not confirm this. The same discrepancy in predicted subcellular localization holds for one of the M03 MPs inhibiting growth (PITG_02711) and one of the M48 MPs (PITG_11607) promoting growth. The other M48 MP tested *in planta* (PITG_12517) and also showing a growth promoting effect, does have a SA but unlike PITG_11607 it is not a single domain protein. It comprises one of the three most ubiquitous bigrams, occurring in nearly all stramenopile species namely a M48 peptidase domain followed by the 'CAAX-prenyl protease N terminal, five transmembrane helices' domain (PF16491).

In the M14 family, opposing effects were found with one M14 MP (i.e., PITG_00756) hampering, and the other one (PITG_11126) promoting lesion growth. Both are single domain proteins but the latter is much larger in size and has signatures of transmembrane regions. Moreover, they are assigned to different subfamilies, M14B and M14C, respectively, and belong to different orthogroups (**Table S2**). *P. infestans* has eight M14 MPs, four of which are predicted to be secreted including three tested *in planta*. The third tested M14 MP (M14B; PITG_06850) showed no significant effect.

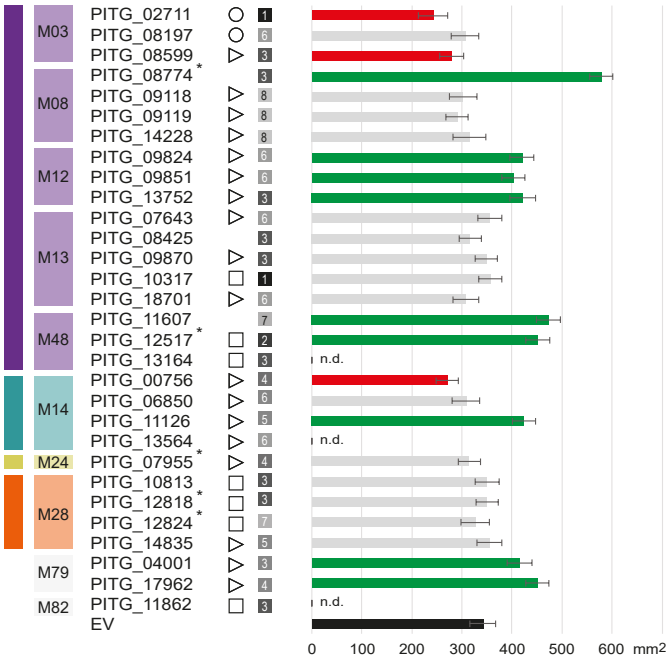


FIGURE 6 | *Phytophthora infestans* genes encoding putative secreted MPs and their capacity to promote or inhibit lesion growth. P. infestans genes encoding MPs that are predicted to be secreted based on the presence of a signal peptide (▷) or signal anchor (□). Five MPs listed here were initially predicted to have a SP or SA but this was not confirmed in a re-analysis. Two of these possess a mitochondrion transit peptide (○). Family and expression profiles are indicated in colours and gray squares with cluster numbers, respectively, with cross reference to Figure 5. Five MPs (marked by *) have additional domains. PITG_08774 has a 'EGF-like domain' (PF07974), PITG_12517 has a 'CAAX prenyl protease N-terminal, five membrane helices' (PF16491), PITG_07955 has an 'Aminopeptidase P, N-terminal domain' (PF16491), and PITG_12818 and PITG_12824) both have a 'PA domain' (PF02225) and a 'Transferrin receptor-like dimerisation domain' (PF04253). The bars on the right show the lesion sizes on *Nicotiana benthamiana* leaves transiently expressing MP genes. One day after agroinfiltration, inoculations were performed with zoospores of *P. infestans* strain 14-3-GFP. Lesions were measured at 5 days post inoculation. Bars represent the average lesion sizes from three biological repeats (n=20). The bar colors indicate smaller (red) or larger (green) lesions or no effect (gray) when compared to empty vector control (black). Three MPs (n.d.) were not tested in planta. Error bars indicate SD and bar colors indicate significantly smaller (red) or significantly larger (green) lesions or no effect (gray) when compared to empty vector (EV) control (black) (ANOVA, $p<0.05$).

There is no significant association between expression profile and *in planta* activity. The twelve MP genes that show *in planta* activity are divided over seven of the eight expression clusters with a slight preference for cluster 3. This is not surprising considering that cluster 3 comprises more than one third of all MP genes. It is reassuring though that the cluster 3 expression profile, i.e. upregulation in mid and late infection stages, complies with a putative function in virulence. In MP families with two members showing *in planta* activity, the expression profiles vary for each of the two members and often show complementarity. For example, one M03 MP peaks in germinating cysts (cluster 1) and the other in mid and late infection stages (cluster 3). In the M12 family with three members showing *in planta* activity, one is in cluster 3 and two in the complementary cluster 6 with relatively high expression in zoospores and germinating cysts.

Taken together, these results strongly suggest a role in virulence for a selected set of MPs. As yet there is no data confirming proteolytic activity of these particular MPs in *P. infestans* and also the question whether proteolytic activity as such is required for virulence remains to be answered.

Discussion

Although several studies highlight a direct or indirect role of MPs in microbial pathogenicity (Fernandes *et al.*, 2019; Joshi *et al.*, 2002; Sanz-Martín *et al.*, 2016), the roles of MPs in pathogenic oomycetes have not yet been explored. Here we aimed to identify and characterize the MP repertoire in Stramenopiles, with a specific focus on oomycetes, and in particular the late blight pathogen *P. infestans*.

Mining of stramenopile proteomes showed a diverse catalog of MPs in each species with a core of 19 MP families present in over 82% of the 46 Stramenopiles analyzed and with some families showing striking species-specific expansions. This is for example the case in *A. astaci*, a pathogen on crustaceans that has several large MP families, mostly comprising single-domain MPs with a SP. These observations are coherent with recent reports showing that the secretome of *A. astaci* is enriched for small secreted proteins with peptidase activities, many of which are indeed MPs (Gaulin *et al.*, 2018). One of the expanded families in *A. astaci* as well as in other animal oomycete pathogens, is M13, a family comprising peptidases known as neprilysins, zinc-metalloenzymes that play a regulatory role in peptide signaling in mammalians (Bland *et al.*, 2008). In bacterial pathogens it was shown that neprilysins have a role a virulence, for example PepO in the human pathogen *Streptococcus pyogenes* (Brouwer *et al.*, 2018) and PtrA in pathogen *Erwinia amylovora* (Zhang *et al.*, 1999). In our plant assays however, we found no indications for a virulence function for any of the five *P. infestans* M13 MPs, suggesting that in oomycetes M13 is relevant for animal pathogens but less so for plant pathogens.

While the majority (62%) of the 4,178 stramenopile MPs are single-domain proteins, a substantial number are multi-domain proteins with accessory domains that might be required for example, for dimerization, substrate binding, or targeting to membranes or subcellular compartments.

The remarkable versatility in the accessory domains implies a broad functionality of MPs. In our analyses we specifically focused on the domain composition of these multidomain MPs and searched for unique bigrams. Most oomycete proteins with unique bigrams seem to be involved in signal transduction (van den Hoogen and Govers, 2018), but there are also examples of a structural proteins, i.e. myosin (Seidl *et al.*, 2011), and an aspartic protease (van den Hoogen *et al.*, 2018). In general, diversity among MP family members and in between MP families and clans is well documented (Rawlings and Barrett, 1995) and in most cases the domain composition of the multi-domain Stramenopile MPs complies with that of MPs in other organisms. Still, we identified three MPs in *P. infestans* that have oomycete specific bigrams. One is a M67 MP with a N-terminal 'PWWP' domain, named after a conserved motif that binds to methylated histone-4 proteins. The other two are M20 MPs that are even more specific as they are limited to two sub-lineages within the oomycetes namely Pythiaceae and Peronosporaceae. One of the M20 MPs has a 'CENP-B N-terminal DNA-binding domain' that, similar to the 'PWWP' domain, points to a function in the nucleus. There is ample evidence that *Phytophthora* species exploit epigenetic modifications for tailoring their virulence (Wang *et al.*, 2020) and it would be worth investigating whether or not the oomycete specific MPs play a role in that process.

In the entire set of Stramenopile MPs we identified a few other intriguing bigrams. One is present in a M67 MP that is found in just two species, *Pythium aphanidermatum* and *Pythium arrhenomanes*, and not in any of the *Phytophthora* species. In this multi-domain protein the MP domain is followed by a PhoD-like phosphatase and a high number of MORN repeats that function as protein-protein interaction modules (Li *et al.*, 2019). MORN1, a protein in the parasite *Toxoplasma gondii* with 14 MORN repeats, appears to play a role in cell division and to act as a linker protein between membranes and the cytoskeleton (Gubbels, 2006).

One of the three most widespread bigrams that were found in this study is a M48 MP combined with a second M48 domain known as CAAX-prenyl-M48 peptidase (PITG_12517). *In planta* assays, revealed that the presence of this MP in leaves leads to an increase in lesion size suggesting a role in virulence of *P. infestans*. Whether or not both M48 domains in the bigram are essential for its virulence function is questionable especially since a second single domain M48 MP (PITG_11607) also showed a growth promoting effect. Furthermore, it is unclear where the M48 MPs are localized. Although the predicted SA might facilitate secretion, it is unlikely that the CAAX-prenyl-M48 part with seven transmembrane spanning regions is secreted. Conceivably, the first M48 domain could be cleaved off and secreted while the CAAX-prenyl-M48 part remains in the ER to exert other functions. CAAX-prenyl-M48 MPs are widespread in eukaryotes and play a role in posttranslational modification of proteins in particular in cleavage of the CAAX moiety of farnesylated proteins. A well-studied CAAX-prenyl–M48 MP is Ste24 in *Saccharomyces cerevisae*, that has a dual role in mating pheromone maturation, namely CAAX cleavage and removal of the N-terminal extension (Pryor *et al.*, 2013; Rawlings and Salvesen, 2013). Knocking-out an ortholog of Ste24 in the parasite *Leishmania donovani* resulted in reduced infectivity and growth (Bhardwaj *et al.*, 2017).

The second *P. infestans* MP that comprises a bigram and showed a growth promoting effect in the *in planta* assays is the M8 MP PITG_08774, that has an additional EGF-like domain. Since none of the four single domain M8 MPs showed a growth promoting effect, it is conceivable that the EGF-like domain is relevant for the virulence function of this M8 MP. It is known that in mammalians the extracellular epidermal growth factor (EGF) binds to the EGF receptor thereby stimulating growth and development. One could speculate that the EGF-like domain in the PITG_08774 protein acts as ligand of a receptor on the host cell membrane, to bring the M8 MP in the proximity of its substrate. Proteolysis of the substrate could then result in suppression of defense. Another example of a M8 MP acting in virulence is leishmanolysin GP63 in *Leishmania* parasites but unlike the PITG_08774 protein this is a single domain MP (Gomez *et al.*, 2009; Hallé *et al.*, 2009).

The other *P. infestans* MPs that showed growth promoting or inhibiting effects in plant assays are all single domain MPs divided over four families, i.e., M03, M12, M14 and M79. Two of the three *P. infestans* M03 MPs showed an inhibitory effect on lesion growth. All three have a SP while in the entire set of Stramenopile MPs only 22% (43/193) of the M03 MPs are predicted to be secreted. Inhibition of lesion growth is remarkable, especially because the expression seems to be relatively high in germinating cysts and/or mid and late infection stages. Presumably there is a mechanism in place that inhibits the activity of these MPs. One could imagine that the host induces expression of the M03 encoding genes as part of its defence machinery but that *P. infestans* in turn produces an effector to suppress M03 activity. To our knowledge this is the first report showing that a M03 MP plays a role in a host-pathogen interaction. The same holds for M79, a family with relatively few known members. Similar to M48 MPs, M79 MPs are endopeptidases that release the CAAX moiety from farnesylated proteins (Rawlings and Salvesen, 2013) and both M48 and M79 MPs show the capacity to promote lesion growth in our plant assays.

M12 MPs, also known as astacins, have been intensively studied due to their implications in Alzheimer's disease and breast cancer (Edwards *et al.*, 2008; Hartmann *et al.*, 2013). They can function as sheddases and as such modify the activity of membrane spanning receptors. This ectodomain shedding has also been observed in plant pattern recognition receptors acting in immunity against pathogens (Petutschnig *et al.*, 2014). As yet, the proteases involved have not been identified but it is conceivable that such proteases are part of the pathogen's weaponry when attacking plants. *P. infestans* has three astacin-like MPs that might have such a role. All three have a SP and showed a consistent growth promoting effect *in planta*. Over 80% of the stramenopile M12 MPs is predicted to be secreted and one of the *P. infestans* M12 MPs (i.e. PITG_09851) was indeed detected in the extracellular proteome (Meijer *et al.*, 2014). The increased expression of PITG_09851 in zoospores and germinated cysts suggests that astacin is present in the pre-infection stage when encountering the plant surface. The last MP family to consider in relation to virulence is M14, with two out of three tested members affecting lesion growth albeit opposite. The one inhibiting lesion growth is a M14C type, the one promoting lesion growth is a M14B type. M14 MPs are carboxypeptidases that hydrolyze single C-terminal

amino acids from polypeptide chains with each subtype having a specificity for certain amino acids. Possibly the released amino acids act as signal molecules triggering certain responses, either in the host or in the pathogen. Four of the eight *P. infestans* M14 MPs are secreted, while the other four presumably function intracellular. Alternatively, modification of the substrate by removal of one amino acid at the C-terminus could also trigger a cascade of new events affecting the host-pathogen interaction. As yet, literature searches did not reveal examples in which M14 MPs have been implicated in microbial pathogenicity.

This study presents an overview of the MP repertoire in Stramenopiles and provides new insights into the immense diversity of MPs in individual species as well the dynamics of genes and gene families related to lifestyle or taxonomic lineage. The lack of MPs or MP families could be due to gene losses, with the acquisition of novel unique MPs through gene gains as counterbalance. Previous studies showed that the evolutionary history of oomycetes is shaped by massive gene gains, duplications and losses and likely this has been a drive for speciation (Seidl *et al.*, 2012). The more in-depth analyses were limited to just one species, *P. infestans*, and were preceded by manual curation of the predicted gene models. During the analyses we encountered the pitfalls of poorly assembled and annotated genomes, and of unexpected patterns in the genome assemblies (Denton *et al.*, 2014). For comparative analyses of gene families one should strive for high quality genome assembly and annotation, and it should be emphasized that manual curation of predicted gene models is paramount for an accurate ortholog inference and family classification of genes and proteins (Vaattovaara *et al.*, 2019).

This study allowed us to pinpoint twelve MPs that affect the virulence of *P. infestans*. These data provide a foundation for further studies on the biochemistry and function of these MPs in *Phytophthora* and in other pathogenic oomycetes. By making use of the broad knowledge gained on MPs in a wide range of organisms including microbial pathogens, it should be feasible to unravel the mechanisms underlying the growth promoting or inhibiting activity of these MPs and to identify compounds that interfere with MP activity. Exploitation of this type of compounds is currently pursued for controlling alveolate parasites such as *Plasmodium* and *Toxoplasma* species (Deu, 2017; Escotte-Binet *et al.*, 2018) and might also be applicable for controlling oomycete diseases.

Experimental procedures

Annotation of putative metalloproteases

Complete proteomes of 37 oomycete species were retrieved from https://github.com/oomycetes/oomycetes.github.io (McGowan and Fitzpatrick, 2017). This included 14 *de novo* generated proteomes and 23 downloaded (**Table S1**). The most recent complete proteomes for four algae species (*Aureococcus anaphagefferens*, *Ectocarpus siliculosus*, *Hondaea fermentalgiana*, *Nannochloropsis gaditana*,) and five diatoms (*Fistulifera solaris*, *Fragilariopsis*

cylindrus, *Phaedactylum tricornutum, Thalassiosira pseudonana*, *Thalassiosira oceanica*) were downloaded from Uniprot (**Table S1**). Since not all proteomes were annotated with equal detail, we used only the longest isoform per gene. To start the annotation (**Figure 1**) the *MEROPS* database (release 12.0) (Rawlings *et al.*, 2018) was downloaded for extracting aligned peptidase sequences per (sub)family. These were used to train profile HMMs with hmmbuild (HMMER package v3.1b2; http://hmmer.org), generating one HMM per *MEROPS* (sub)family. We matched these HMMs to the proteomes and selected matches maintaining an empirical E-value threshold of $1e^{-10}$. All matching subsequences were extracted and used to train Stramenopile-specific HMMs. Again, the complete proteomes were matched to these HMMs, this time maintaining a lower E-value threshold of $1e^{-25}$. For both iterations of HMM searches, all proteins were assigned to their respective best matching *MEROPS* family. Apart from using solely *MEROPS* as a data source for MP searches, we extracted domain HMMs from the Pfam database (release 31.0) (Finn *et al.*, 2016). The HMMs of all peptidase domains that were cross-referenced with *MEROPS* were extracted, and HMM searches were performed using the complete proteomes (E-value $1e^{-20}$). Proteins that either matched a *MEROPS* MP HMM, or a Pfam MP HMM were considered putative MPs. Classification was done according to *MEROPS* families and clans.

Sequence analyses

All putative MP protein sequences were subjected to feature annotation using Interproscan (version 5.30-69.0) (Jones *et al.*, 2014). The Interproscan pipeline includes, among others, the predictors SignalP version 4.1 (H. Nielsen, 2017) and Phobius version 1.01 (Käll *et al.*, 2004) for signal peptide predictions, and Pfam version 31.0 for domain annotation. We used HECTAR (Gschloessl *et al.*, 2008) at the Galaxy Webserver of Station Biologique de Roscoff (https://webtools.sb-roscoff.fr/) to assign the MPs to five different categories of subcellular targeting: i.e. signal peptide (SP), signal anchor (SA), chloroplast or mitochondrion transit peptides, and other. To annotate protein domains, we exploited the HMM search results of Pfam and *MEROPS*. Notably, HMMER only reports significant local alignments, yet the full peptidase domains may be larger. To derive the coordinates of the full peptidase domains, we merged consecutive local alignments in the HMM alignments. All non-intersecting Pfam domains reported by Interproscan were considered accessory domains. To assess the sequence similarity between all identified putative MPs, we constructed clusters of ortholog sequences using Orthofinder version 2.2.6 (Emms and Kelly, 2015).

The gene models of all putative MPs for *P. infestans* were manually curated using the WebApollo genome browser (Lee *et al.*, 2013), loaded with the *P. infestans* T30-4 genome sequence (Haas *et al.*, 2009). RNA sequencing reads of various strains and life stages of *P. infestans* (see below) were aligned using HiSat2 (v2.1.0) spliced aligner (Kim *et al.*, 2015) and alignments were used as a guide to determine the correct exons. *P. infestans* MPs were blasted against the NCBI non-redundant database using the BLAST web server to check for their occurrence in other species. Significant hits (E-value $<1e^{-5}$ and 50% identity) were selected for manual in-depth domain composition analyses, using the web servers of InterPro and Pfam.

Gene expression analyses

P. infestans strain T20-2 was grown on rye sucrose medium at 18°C and 10-days old mycelium was collected for RNA extraction. Sporangia were isolated from 10-days old mycelium growing on rye agar medium flooded with 10 ml water. To isolate zoospores, 10-days old mycelium was flooded with ice-cold water for 3 hours. Zoospores were collected after filtration with a 50 µm mesh. Germinating cysts were obtained from collected zoospores that were left to germinate at room temperature for 3 hours. RNA was extracted using NucleoSpin RNA II extraction kit (Macherey-Nagel) according to manufacturer's instructions. RNA sequencing was performed using an Illumina HiSeq™ 2000, producing 90 bp paired-end reads. For *in planta* expression data, we downloaded *P. infestans* RNA-Seq libraries from NCBI BioProject PRJNA361417 (Ah-Fong *et al.*, 2017b) representing the transcription in infected tuber during early, mid and late infection. RNA abundance in all samples was quantified using Kallisto version 0.43.1 (Bray *et al.*, 2016), resulting in transcripts-per-million values (TPM). These were Z-score transformed and hierarchically clustered using Pearson correlation distance and average linkage. The optimal number of clusters was determined using the gap statistic implemented in the R package NbClust (Charrad *et al.*, 2014).

Transient expression in *Nicotiana benthamiana* and infection assays

Full length coding sequences of MP genes were amplified from cDNA of *P. infestans* strain 88069 using primers listed in **Table S2**. The amplified fragments were introduced by TOPO cloning in plasmid pENTR/D-TOPO and subsequently in the binary vector pGWB5 by LR recombination. The resulting constructs were transformed into *Agrobacterium tumefaciens* strain Agl1 and transformants were used for agroinfiltration in *N. benthamiana* leaves. Agroinfiltration, inoculation with *P. infestans* zoospores from strain 14-3-GFP and measurements of lesion sizes were performed as described previously (Meijer *et al.*, 2019).

Acknowledgements

We thank Natalie Verbeek-de Kruif for technical support with the cloning and plant assays.

Supplemental Information

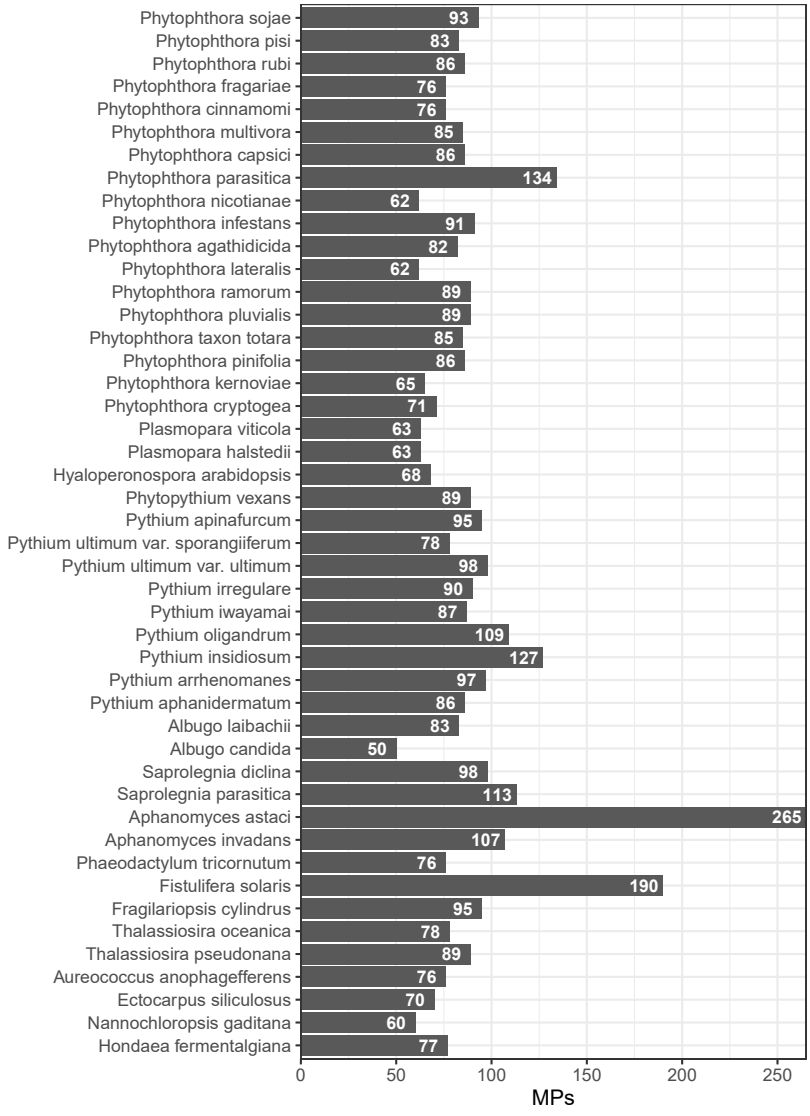


FIGURE S1 | The number of MPs per stramenopile species.

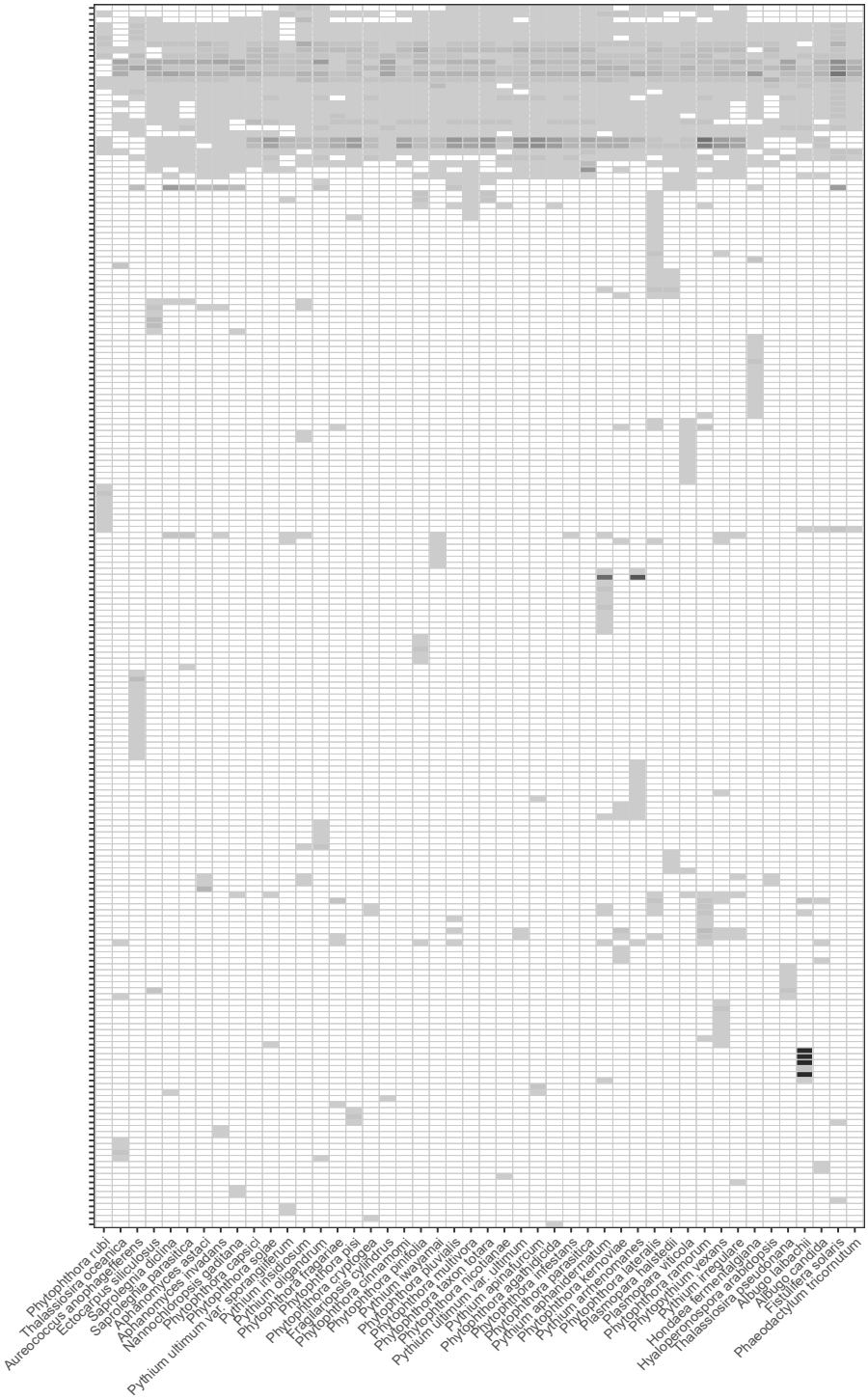


FIGURE S2 | The diversity of accessory domains in stramenopile MPs and their frequency of occurrence in MPs in the different species.

FIGURE S3 | The diversity of bigrams in stramenopile MPs and their frequency of occurrence in MPs in the different species..

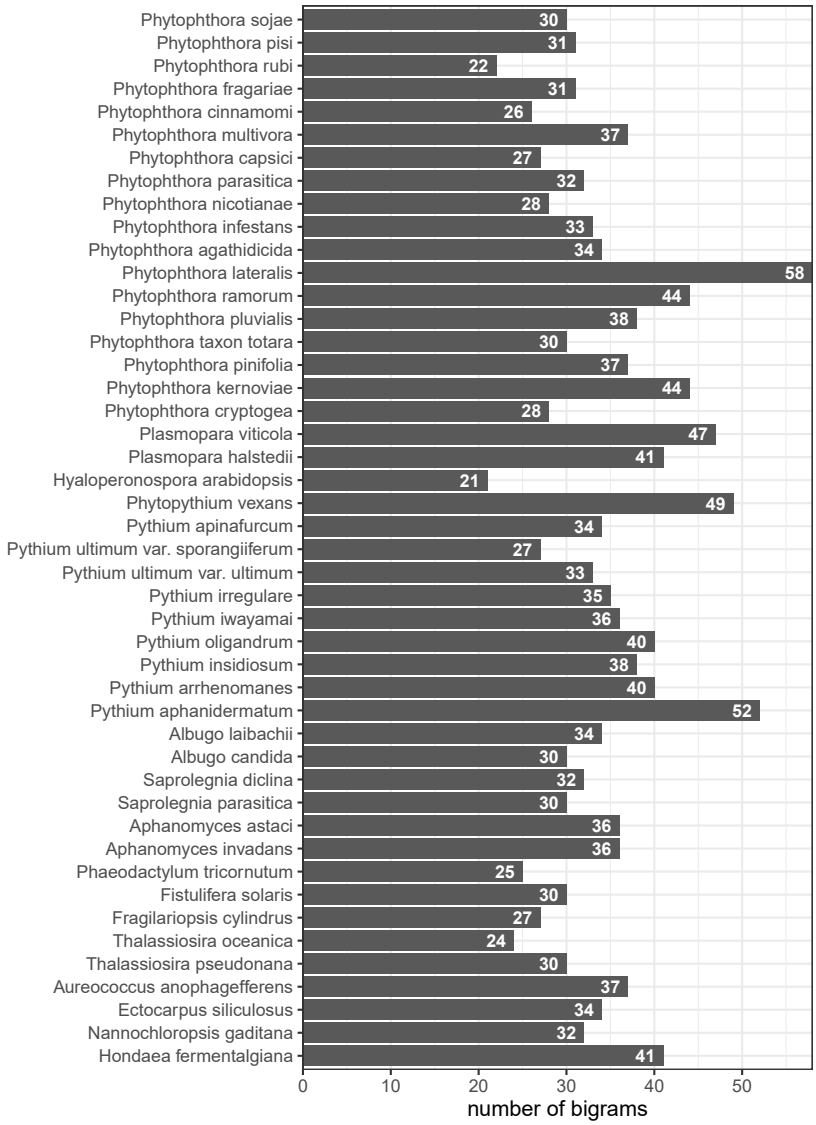


FIGURE S4 | The number of bigrams in MPs in each stramenopile species.

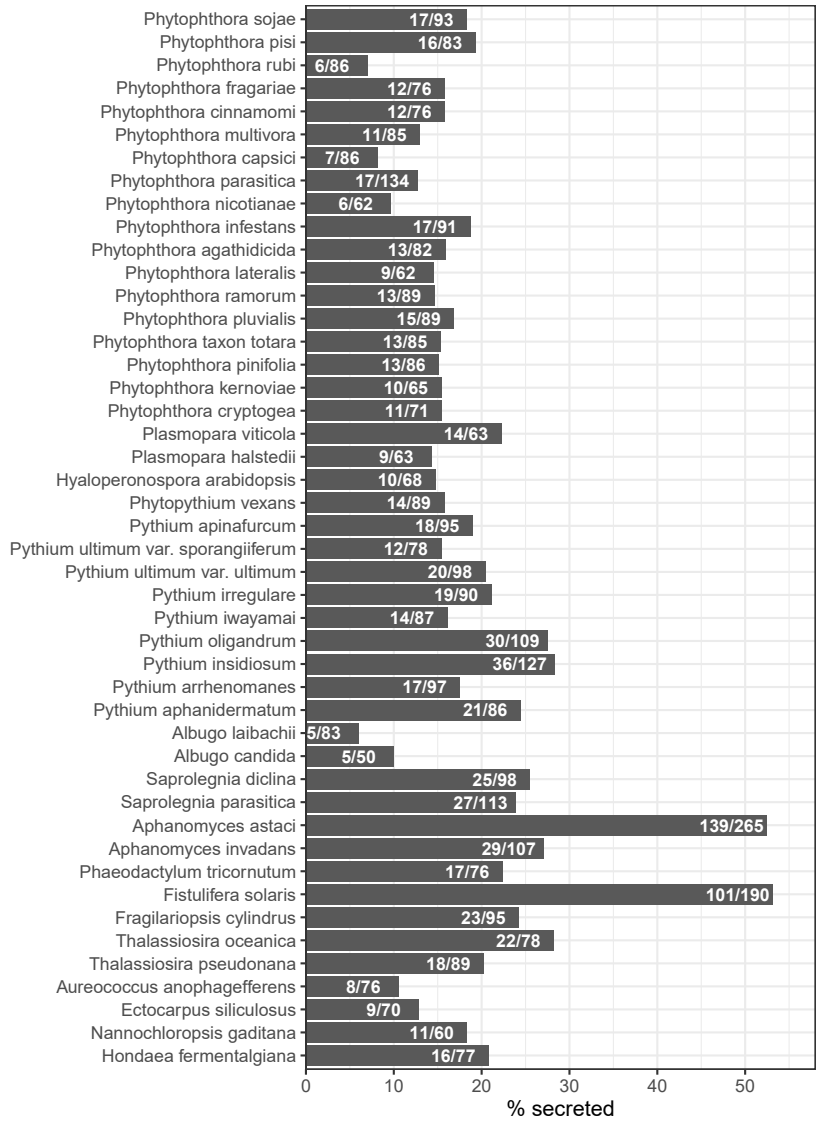


FIGURE S5 | The percentage of potentially secreted MPs per species based on SP prediction. For each species the ratio secreted MPs over the total number of MPs is shown.

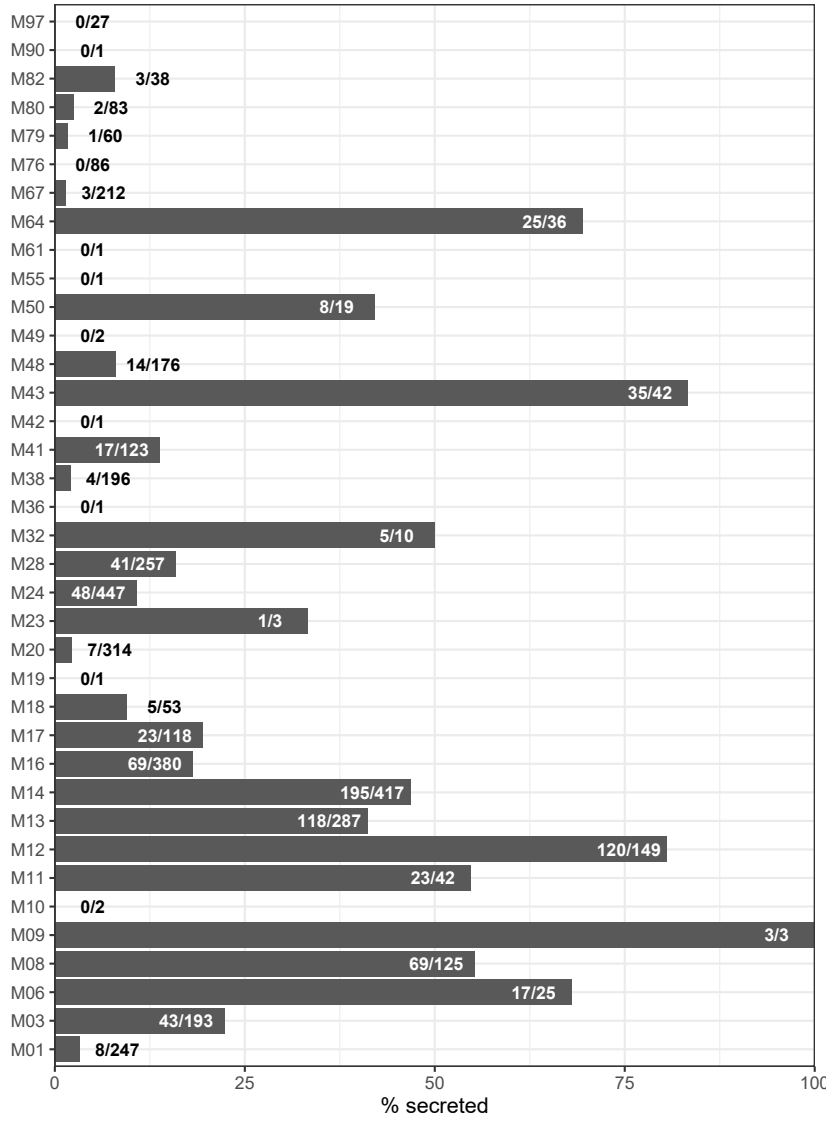


FIGURE S6 | The percentage of potentially secreted MPs per MP family based on SP prediction. For each MP family the ratio secreted MPs over the total number of stramenopile MPs in that family is shown.

TABLE S2A | Overview of the 91 MPs identified in *Phytophthora infestans*. Gene IDs, protein length, MEROPS classification, predicted localization, domain composition, number of accessory domains and orthogroup membership. Full version can be found online.

gene	length	family	clan	source	signalP TM	signalP no TM	Phobius TM	Phobius SP	TMHMM TM
PITG_01616	980	M01	MA	merops	no	no	no	no	no
PITG_01619	1008	M01	MA	merops	no	no	no	no	no
PITG_02917	884	M01	MA	merops	no	no	no	no	no
PITG_03837	639	M01	MA	merops	no	no	no	no	no
PITG_08804	899	M01	MA	merops	no	no	no	no	no
PITG_15392	670	M01	MA	merops	no	no	yes	no	yes
PITG_15413	677	M01	MA	merops	no	no	yes	no	yes
PITG_21860	505	M01	MA	merops	no	no	yes	no	yes
PITG_02711	728	M03	MA	merops	no	no	no	no	no
PITG_08197	704	M03	MA	merops	no	no	no	no	no
PITG_08599	712	M03	MA	merops	no	no	no	yes	yes
PITG_08774	1018	M08	MA	merops	no	no	yes	no	no
PITG_09118	738	M08	MA	merops	no	yes	no	yes	no
PITG_09119	681	M08	MA	merops	no	yes	yes	yes	no
PITG_13513	144	M08	MA	merops	no	no	no	no	no
PITG_14228	652	M08	MA	merops	no	yes	no	yes	no
PITG_09824	409	M12	MA	merops	no	yes	no	yes	no
PITG_09851	393	M12	MA	merops	no	yes	no	yes	no
PITG_13752	910	M12	MA	merops	no	yes	yes	yes	yes
PITG_07643	463	M13	MA	merops	no	yes	no	yes	no
PITG_08425	1223	M13	MA	merops	no	no	yes	no	no
PITG_09870	677	M13	MA	merops	no	yes	no	yes	no
PITG_10317	773	M13	MA	merops	no	no	yes	no	yes
PITG_18701	607	M13	MA	merops	no	yes	no	yes	no
PITG_01087	658	M41	MA	merops	no	no	yes	no	no
PITG_10147	874	M41	MA	merops	no	no	no	no	no
PITG_02177	415	M48	MA	merops	no	no	no	no	no
PITG_06937	300	M48	MA	merops	no	no	no	no	no
PITG_11607	217	M48	MA	merops	no	no	no	no	no
PITG_12517	485	M48	MA	merops	no	no	yes	no	yes
PITG_13164	290	M48	MA	merops	no	no	yes	no	yes
PITG_08420	188	M76	MA	merops	no	no	no	no	no
PITG_18055	196	M76	MA	merops	no	no	no	no	no
PITG_13744	283	M80	MA	merops	no	no	no	no	no
PITG_15925	478	M80	MA	merops	no	no	no	no	no
PITG_05650	905	M97	MA	merops	no	no	yes	no	yes
PITG_00756	482	M14	MC	merops	no	yes	no	yes	no
PITG_05548	1687	M14	MC	merops	no	no	no	no	no

	HECTAR localisation	domain composition	orthogroup
	other localisation	M1 (M01) [357-517] | Domain of unknown function (DUF3458) Ig-like fold [544-632] | Domain of unknown function (DUF3458_C) ARM repeats [644-960]	OG0000023
	other localisation	M1 (M01) [384-544] | Domain of unknown function (DUF3458) Ig-like fold [571-659] | Domain of unknown function (DUF3458_C) ARM repeats [671-987]	OG0000023
	other localisation	M1 (M01) [284-449] | ERAP1-like C-terminal domain [553-864]	OG0000011
	other localisation	M1 (M01) [280-435] | Leukotriene A4 hydrolase, C-terminal [519-631]	OG0000009
	other localisation	M1 (M01) [287-450] | ERAP1-like C-terminal domain [563-877]	OG0000011
	other localisation	M1 (M01) [267-426] | Leukotriene A4 hydrolase, C-terminal [498-614]	OG0000009
	other localisation	M1 (M01) [269-427] | Leukotriene A4 hydrolase, C-terminal [500-618]	OG0000009
	other localisation	M1 (M01) [97-255] | Leukotriene A4 hydrolase, C-terminal [328-446]	OG0000009
	mitochondrion	M3A (M03) [100-721]	OG0000016
	mitochondrion	M3A (M03) [105-700]	OG0000033
	signal peptide	M3A (M03) [102-695]	OG0000039
	other localisation	M8 (M08) [192-507] | EGF-like domain [687-713]	OG0000010
	signal peptide	M8 (M08) [113-490]	OG0000010
	signal peptide	M8 (M08) [117-496]	OG0000010
	other localisation	M8 (M08) [1-122]	OG0000010
	signal peptide	M8 (M08) [104-481]	OG0000010
	signal peptide	M12A (M12) [98-248]	OG0000007
	signal peptide	M12 (M12) [160-316]	OG0000007
	signal peptide	M12A (M12) [594-757]	OG0000054
	signal peptide	M13 (M13) [35-453]	OG0000000
	other localisation	M13 (M13) [87-1223]	OG0000000
	signal peptide	M13 (M13) [33-677]	OG0000000
	signal anchor	M13 (M13) [111-773]	OG0000000
	signal peptide	M13 (M13) [35-607]	OG0000000
	mitochondrion	ATPase family associated with various cellular activities (AAA) [257-386] | M41 (M41) [428-650]	OG0000006
	other localisation	FtsH Extracellular [160-310] | ATPase family associated with various cellular activities (AAA) [420-552] | M41 (M41) [596-821]	OG0000006
	mitochondrion	M48C (M48) [143-314]	OG0000050
	other localisation	M48A (M48) [81-275]	OG0000055
	other localisation	M48C (M48) [1-116]	OG0000050
	signal anchor	CAAX prenyl protease N-terminal, five membrane helices [48-268] | M48A (M48) [280-477]	OG0000026
	signal anchor	M48C (M48) [97-251]	OG0000022
	other localisation	M76 (M76) [22-186]	OG0000044
	other localisation	M76 (M76) [31-193]	OG0000034
	mitochondrion	M80 (M80) [14-206]	OG0000047
	other localisation	M80 (M80) [134-279] | PUB domain [391-457]	OG0000038
	other localisation	Domain of unknown function (DUF5117) [280-329] | M97 (M97) [351-569]	OG0000056
	signal peptide	M14C (M14) [197-444]	OG0000003
	other localisation	M14D (M14) [1079-1215]	OG0000002

gene	length	family	clan	source	signalP TM	signalP no TM	Phobius TM	Phobius SP	TMHMM TM
PITG_06195	654	M14	MC	merops	no	no	no	no	no
PITG_06850	531	M14	MC	merops	no	yes	no	yes	no
PITG_09251	863	M14	MC	merops	no	no	yes	no	yes
PITG_11126	828	M14	MC	merops	no	yes	yes	yes	yes
PITG_13564	277	M14	MC	merops	no	yes	no	yes	yes
PITG_23044	916	M14	MC	merops	no	no	no	no	no
PITG_00203	466	M16	ME	merops	no	no	no	yes	no
PITG_01569	1047	M16	ME	merops	no	no	no	no	no
PITG_04726	447	M16	ME	merops	no	no	no	no	no
PITG_08991	1008	M16	ME	merops	no	no	no	no	no
PITG_13437	1048	M16	ME	merops	no	no	no	no	no
PITG_16759	1078	M16	ME	merops	no	no	no	no	no
PITG_17056	1000	M16	ME	merops	no	no	no	no	no
PITG_20378	531	M16	ME	merops	no	no	no	no	no
PITG_04838	519	M17	MF	merops	no	no	no	no	no
PITG_11569	534	M17	MF	merops	no	no	no	no	no
PITG_01206	481	M24	MG	merops	no	no	no	no	no
PITG_04875	630	M24	MG	merops	no	no	no	no	no
PITG_07955	553	M24	MG	merops	no	no	no	yes	no
PITG_08392	1077	M24	MG	pfam	no	no	no	no	no
PITG_09504	453	M24	MG	merops	no	no	no	no	no
PITG_12211	394	M24	MG	merops	no	no	no	no	no
PITG_12220	378	M24	MG	merops	no	no	no	no	no
PITG_13014	393	M24	MG	merops	no	no	no	no	no
PITG_13658	291	M24	MG	merops	no	no	no	no	no
PITG_02114	462	M18	MH	merops	no	no	no	no	no
PITG_00028	422	M20	MH	merops	no	no	no	no	no
PITG_00029	416	M20	MH	merops	no	no	no	no	no
PITG_00577	865	M20	MH	merops	no	no	no	no	no
PITG_02852	407	M20	MH	merops	no	no	no	no	no
PITG_05858	419	M20	MH	merops	no	no	no	no	no
PITG_05861	437	M20	MH	merops	no	no	no	no	no
PITG_05866	192	M20	MH	merops	no	no	no	no	no
PITG_06978	682	M20	MH	merops	no	no	no	no	no
PITG_00273	869	M28	MH	merops	no	no	yes	no	yes
PITG_00289	776	M28	MH	merops	no	no	yes	no	yes
PITG_10813	875	M28	MH	merops	no	no	yes	no	yes

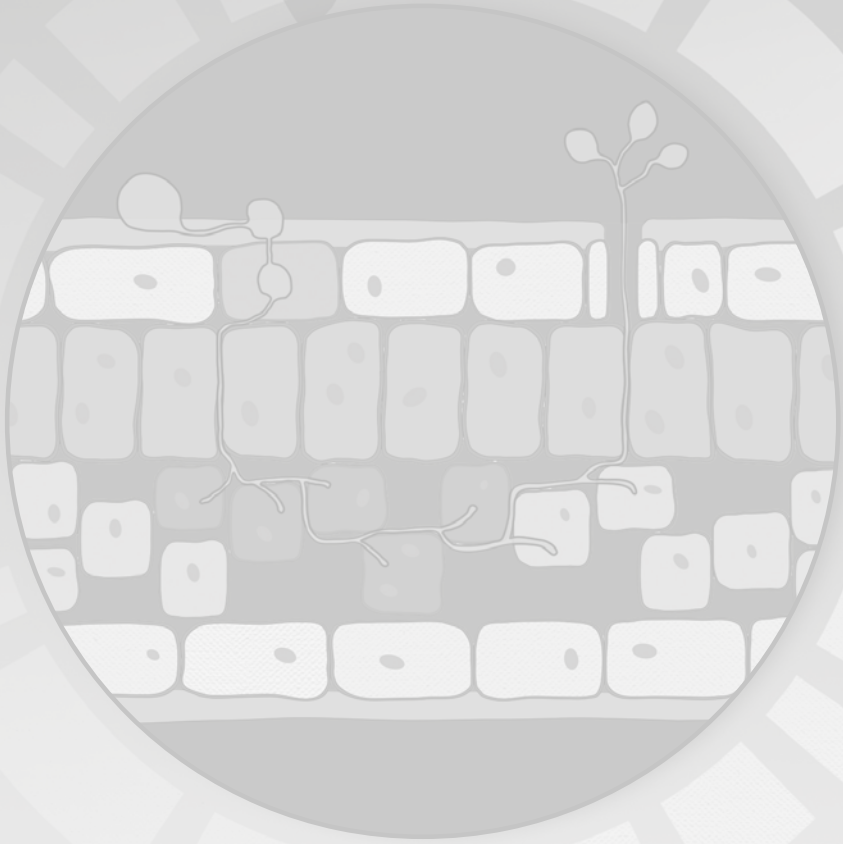
	HECTAR localisation	domain composition	orthogroup
	other localisation	M14D (M14) [266-428]	OG0000002
	signal peptide	M14B (M14) [35-285]	OG0000013
	other localisation	M14C (M14) [239-494]	OG0000035
	signal peptide	M14B (M14) [33-299]	OG0000013
	signal peptide	M14B (M14) [32-277]	OG0000013
	other localisation	M14D (M14) [294-430]	OG0000002
	mitochondrion	M16B (M16) [43-238]	OG0000028
	other localisation	M16A (M16) [23-257] | Middle or third domain of peptidase_M16 [437-721] | Peptidase M16 inactive domain [748-880]	OG0000001
	mitochondrion	M16B (M16) [57-250]	OG0000058
	other localisation	M16B (M16) [27-231] | Middle or third domain of peptidase_M16 [380-672] | Peptidase M16 inactive domain [689-865]	OG0000001
	other localisation	M16B (M16) [99-281] | Peptidase M16C associated [518-771]	OG0000024
	other localisation	M16A (M16) [29-252] | Middle or third domain of peptidase_M16 [439-737]	OG0000001
	other localisation	M16B (M16) [30-234] | Middle or third domain of peptidase_M16 [384-667] | Peptidase M16 inactive domain [679-842]	OG0000001
	mitochondrion	M16B (M16) [97-291]	OG0000031
	mitochondrion	Cytosol aminopeptidase family, N-terminal domain [60-161] | M17 (M17) [196-513]	OG0000020
	mitochondrion	M17 (M17) [204-526]	OG0000021
	mitochondrion	Aminopeptidase P, N-terminal domain [57-174] | M24A (M24) [213-465]	OG0000046
	mitochondrion	Creatinase/Prolidase N-terminal domain [44-169] | M24B (M24) [328-585]	OG0000027
	signal peptide	Aminopeptidase P, N-terminal domain [93-212] | M24B (M24) [254-546]	OG0000018
	other localisation	FACT complex subunit SPT16 N-terminal lobe domain [32-190] | PF00557.23 (M24) [206-438] | FACT complex subunit (SPT16/CDC68) [572-727] | Histone chaperone Rttp106-like [856-934]	OG0000040
	other localisation	M24A (M24) [125-449]	OG0000017
	other localisation	zf-MYND-like zinc finger, mRNA-binding [18-64] | M24A (M24) [136-375]	OG0000005
	other localisation	zf-MYND-like zinc finger, mRNA-binding [6-52] | M24A (M24) [121-360]	OG0000005
	other localisation	M24A (M24) [31-353]	OG0000030
	mitochondrion	M24A (M24) [54-291]	OG0000005
	other localisation	M18 (M18) [17-449]	OG0000025
	other localisation	M20F (M20) [7-412]	OG0000008
	other localisation	M20F (M20) [4-409]	OG0000008
	mitochondrion	Amidohydrolase family [96-481] | M20 (M20) [522-786]	OG0000052
	other localisation	M20F (M20) [6-402]	OG0000008
	other localisation	M20F (M20) [15-400]	OG0000012
	other localisation	M20F (M20) [13-428]	OG0000012
	other localisation	M20F (M20) [20-181]	OG0000012
	other localisation	CENP-B N-terminal DNA-binding domain [30-72] | M20D (M20) [190-652]	OG0000015
	other localisation	PA domain [222-306] | M28B (M28) [355-700] | Transferrin receptor-like dimerisation domain [753-862]	OG0000051
	other localisation	PA domain [207-270] | M28B (M28) [330-612] | Transferrin receptor-like dimerisation domain [658-775]	OG0000004
	signal anchor	M28B (M28) [113-339]	OG0000043

gene	length	family	clan	source	signalP TM	signalP no TM	Phobius TM	Phobius SP	TMHMM TM
PITG_12818	816	M28	MH	merops	no	no	yes	no	no
PITG_12824	757	M28	MH	merops	no	no	yes	no	no
PITG_14835	719	M28	MH	merops	no	yes	yes	yes	yes
PITG_00636	842	M38	MJ	pfam	no	no	no	no	no
PITG_03605	459	M38	MJ	pfam	no	no	no	no	no
PITG_05263	481	M38	MJ	pfam	no	no	no	no	no
PITG_07820	444	M38	MJ	pfam	no	no	no	no	no
PITG_08142	474	M38	MJ	pfam	no	no	no	no	no
PITG_03561	362	M67	MP	merops	no	no	no	no	no
PITG_05600	451	M67	MP	merops	no	no	no	no	no
PITG_08676	311	M67	MP	merops	no	no	no	no	no
PITG_11063	319	M67	MP	pfam	no	no	yes	no	no
PITG_16722	1368	M67	MP	merops	no	no	no	no	no
PITG_04001	315	M79	none	merops	no	no	yes	no	yes
PITG_17962	269	M79	none	merops	yes	no	yes	yes	yes
PITG_11862	446	M82	none	merops	no	no	yes	no	yes

	HECTAR localisation	domain composition	orthogroup
	signal anchor	PA domain [236-299] | M28B (M28) [359-644] | Transferrin receptor-like dimerisation domain [691-813]	OG0000004
	signal anchor	PA domain [183-263] | M28B (M28) [306-593] | Transferrin receptor-like dimerisation domain [641-754]	OG0000004
	signal peptide	M28A (M28) [235-459]	OG0000061
	other localisation	Urease, gamma subunit [1-99] | Urease beta subunit [134-230] | Urease alpha-subunit, N-terminal domain [272-391] | PF01979.19 (M38) [397-726]	OG0000029
	other localisation	PF01979.19 (M38) [62-399]	OG0000019
	other localisation	PF01979.19 (M38) [63-447]	OG0000014
	other localisation	PF01979.19 (M38) [66-438]	OG0000014
	other localisation	PF01979.19 (M38) [75-436]	OG0000019
	other localisation	M67A (M67) [44-201]	OG0000037
	other localisation	USP8 dimerisation domain [38-103] | M67C (M67) [276-394]	OG0000041
	other localisation	M67A (M67) [35-163] | Maintenance of mitochondrial structure and function [175-291]	OG0000032
	other localisation	PF01398.20 (M67) [13-122] | Maintenance of mitochondrial structure and function [170-281]	OG0000036
	other localisation	PWWP domain [36-167] | M67A (M67) [506-638]	OG0000053
	signal peptide	M79 (M79) [124-247]	OG0000045
	signal peptide	M79 (M79) [106-220]	OG0000063
	signal anchor	M82 (M82) [69-344]	OG0000059

Chapter 6

General discussion



Oomycetes are filamentous pathogens that pose serious threats to agriculture, animals, and ecosystems, and are challenging to control. Their dynamic genomes provide oomycetes remarkable adaptive capacity (Leesutthiphonchai *et al.*, 2018). Systems biology can provide insight into the mechanisms that allow oomycetes to proliferate and infect their host (Peyraud *et al.*, 2017). The chapters in this thesis describe the reconstruction, modelling, and comparison of oomycete metabolism to provide insight into the mechanisms related to life cycle and host interactions. In this chapter, I first discuss the process to reconstruct a computational model from the genome sequence of a species, and how we applied these steps to reconstruct the metabolic networks for oomycetes. I then elaborate on metabolic modelling of *Phytophthora infestans*-host interactions in particular, and how the results of our *P. infestans* metabolic model can be interpreted to provide insight into possible modes of host interaction. Finally, I present an outlook, and discuss a strategy to improve the *P. infestans* model in the coming years.

From genome sequence to genome-scale metabolic model

Genome-scale models (GEMs) are reconstructions of cellular systems that consider all genes encoded in the genome and allow to relate an emerging phenotype to the properties of the models (Yurkovich and Palsson, 2016). In other words, GEMs are used to investigate the genotype-phenotype relationship: how do the genes encoded by the genome result in the complex biological system that we observe? To answer this question, one needs to mine the available information within the genome, and integrate this information into genome-scale networks of interactions. Quality and insights derived from genome-scale models therefore critically depend on the quality of the genome sequence and gene annotation.

Obtaining a high-quality representation of the genome sequence (the genome assembly) is still challenging. Especially eukaryotic genomes are often hard to assemble because they are typically more complex than prokaryotic genomes. This complexity is mainly caused by a high abundance of repetitive elements and the fact that most eukaryotes are typically diploid and sometimes polyploid (or even aneuploid). The sequencing and comparative analyses of the first oomycete genomes in 2006 (*Phytophthora sojae* and *Phytophthora ramorum*) and 2009 (*Phytophthora infestans*) revealed that these *Phytophthora* species profoundly differ in genome size and content (Haas *et al.*, 2009; Tyler *et al.*, 2006). In particular *P. infestans* has a large genome compared to its close relatives, which is largely caused by the increased abundance of transposable elements, constituting roughly 74% of its genome. Transposable elements, in combination with a diploid, polyploid, or aneuploid genome, still hamper present-day genome assembly strategies (Nagarajan and Pop, 2013). Consequently, oomycete genome assemblies are still rather fragmented (**Figure 1A**) (McGowan *et al.*, 2019), in particular compared to fungal plant pathogens for which genome assemblies are nowadays often near-complete (Faino *et al.*, 2015). Only over the last few years, near-complete genome assemblies for a few oomycetes have been published (Fletcher *et al.*, 2019). In our own experience, 160-

fold genome coverage of PacBio long reads combined with 10x Genomics Chromium barcoded linked-reads were not sufficient to assemble the genome sequence of a *P. infestans* strain to an acceptable number of contigs (Rodenburg *et al.* unpublished). Later we learned that we had sequenced a (partially) triploid strain. It can be anticipated that the combined use of long and short reads, and novel techniques such as trio-binning (parental sequencing to enable phasing), Hi-C (spatial genome conformation) and optical mapping (comparing restriction sites of long DNA stretches) will yield better results (Burton *et al.*, 2013; Du and Liang, 2019; Koren *et al.*, 2018).

Despite the fragmented genome assemblies, the majority of oomycete genome sequences comprises most of the near-universal single-copy conserved orthologs, as determined by BUSCO (**Figure 1B**) (Seppey *et al.*, 2019), suggesting that a significant proportion of the coding genome is captured. Another major challenge is to correctly identify the open reading frames with associated exon boundaries (gene models) within the assembled genome sequences (Salzberg, 2019), and this is even more of a problem for fragmented genome assemblies (Denton *et al.*, 2014). Eukaryotic genomes are typically annotated using gene predictors trained on the parameters of high-quality gene models from closely related species and aligned transcriptome data (Yandell and Ence, 2012). Annotation of the first two sequenced *Phytophthora* genomes was performed using a gene predictor trained on expressed sequence tags (Tyler *et al.*, 2006). Many subsequent oomycete genome annotations were performed by gene predictors trained on the gene models in other oomycete genomes (McGowan and Fitzpatrick, 2017). However, the error rate in predicted gene models is still high (Salzberg, 2019), as was also demonstrated in **Chapter 5**, in which we manually corrected several putative metalloprotease genes predicted in the genome of *Phytophthora infestans*. Moreover, in **Chapters 2-4** we specifically searched for potentially missing gene models in the genome sequences of oomycetes, and identified several in each genome. Consequently, the field would benefit greatly from a manual curation effort of genome annotations. This would be valuable even when applied to only a single species, preferably one with a near-complete genome, which can then be used as a template to re-evaluate gene models of other oomycetes and train gene predictors. Especially community-based annotation would be a powerful resource (Rödelsperger *et al.*, 2019).

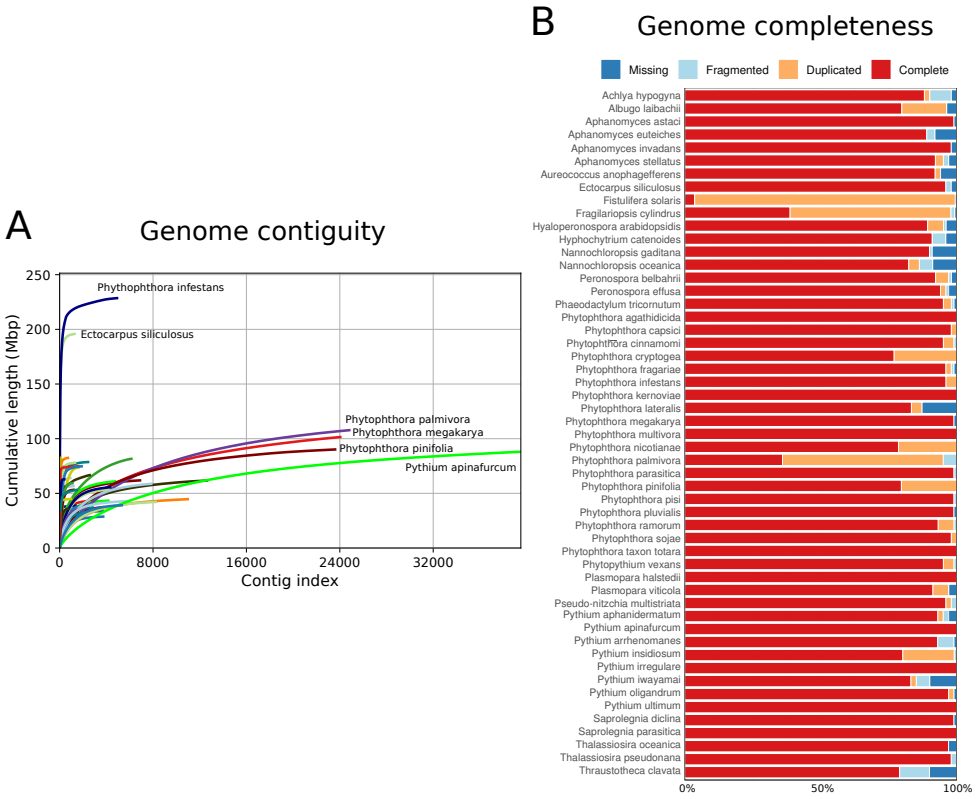


FIGURE 1 | Statistics for 56 published Stramenopile genome sequences including 42 oomycetes. **A:** Genome accumulation curves derived from Quast (Gurevich *et al.*, 2013), showing cumulative genome size when contigs are ordered from large to small. X-axis represents the number of contigs, Y-axis represents genome size. **B:** Presence of near-universal single copy orthologs in the genomes, as determined by BUSCO. Species are in alphabetical order. Colors indicate the completeness of the detected BUSCO genes.

Reliable functional annotation of the predicted proteome is an additional prerequisite for identifying the fundamental components of genome-scale models. In **Chapter 2**, we performed an automated proteome annotation for 42 oomycete species, which revealed that the large majority of predicted proteins could not be assigned to a KEGG orthologous group (KO), which are protein ortholog clusters with validated functions (Mao *et al.*, 2005). These findings illustrate that the large majority of Stramenopile and oomycete genes cannot be associated with any function. Many effector genes in oomycetes (e.g. those encoding RxLR effectors) lack functional associations (McGowan and Fitzpatrick, 2017). One should consider that distantly related protein orthologs are inherently harder to detect than closely related ones, which posed a problem in our homology-based annotations. However, even in the best studied eukaryotic model organism, the yeast *Saccharomyces cerevisiae,* about 20% of all genes lack any functional association (Wood *et al.*, 2019). Lineages that are evolutionary distant to model organisms — such as the oomycetes — have an even less functionally characterized

proteome, partly because of limitations of homology-based inference of protein functions, but most importantly, because of lack of experimental characterization. As an effective way to predict protein functions by homology, we primarily used hidden Markov models (HMMs) for protein annotation (**Chapter 2-5**). HMMs are trained on a multiple sequence alignment of a predefined cluster of homologous protein sequences, and weigh conserved sequence regions heavier than variable regions. Therefore, HMMs are particularly suitable for detecting protein domains, as these are often highly conserved to retain their biological function (Pearson, 2013). KOs are predefined ortholog clusters and are powerful resources to train HMMs (Aramaki *et al.*, 2019). In **Chapter 2, 3,** and **4** we used KO-based HMMs to identify orthologs of metabolic enzymes in oomycetes and their close relatives. In **Chapter 2** we made use of KofamScan (Aramaki *et al.*, 2019), which employs an adaptive threshold on the probability assigned by a KO HMM to decide whether a sequence is orthologous or not. This threshold was determined by a heuristic method that aimed to minimize the F1 measure, balancing precision and recall, when positive and negative training samples are analyzed. However, we noted that this method missed putative orthologs, as could later be determined by manual BLAST confirmation of missing KOs. We chose to apply a more suitable thresholding method in **Chapter 3** and **4**, in which we combined a static threshold with a relative threshold, depending on the E-values of all proteins matching the respective HMM. This method was adopted from the RAVEN toolbox (Agren *et al.*, 2013; Thines *et al.*, 2020), but we empirically optimized the thresholds for the predicted proteomes used in our studies. Previously, oomycetes genomes have been shown to be enriched for novel domain combinations (Seidl *et al.*, 2011). In some cases this results in two enzymatic domains in one protein, possibly to facilitate substrate channeling (Judelson *et al.*, 2009b). In other cases an enzymatic domain is combined with a characteristic signal transduction module (**Chapter 5**) (van den Hoogen *et al.*, 2018; Meijer and Govers, 2006). The thresholding method we applied enabled the assignment of multiple KOs to a single protein, though with a strict threshold, and thus is more suitable to capture proteomic complexity in oomycetes. In summary, genome/proteome annotation has been and likely will remain a challenging task for oomycetes, as incomplete genomes and a lack of lineage-specific knowledge trouble homology-based inference. Future genome assemblies and revisited annotations will likely give a more accurate view of the oomycete proteome. Nevertheless, the use of adapted algorithms and approaches enabled us to chart a genome-scale repertoire of enzymes for each of the oomycetes (**Chapters 3-5**)**.**

In omics-based bioinformatics studies, it is conventional to look for overrepresentation (enrichment) of functional annotations in differentially abundant molecules such as mRNA or proteins (Bordbar *et al.*, 2014; Reed *et al.*, 2006). These data provide insight into the biological systems that are active during the specific sampling conditions (e.g. time, developmental stage or tissue) of an organism (Rodenburg *et al.*, 2018b). In the last decade, there has been a continuous flow of published genome sequences for oomycetes, accompanied by numerous studies publishing comparative genomics, transcriptomics, proteomics, and metabolomics research (McGowan and Fitzpatrick, 2020). These large-scale omics datasets are analyzed to understand how oomycetes evolve, reproduce and interact with their hosts. Despite omics

studies being indispensable to investigate the transcriptional/translational responses of both pathogen and host during infection, these studies are often biased, as usually only a subset of functions is investigated and these are not necessarily representative or causative for the complex phenotype. Moreover, proteomic and metabolomic samples typically only capture the most ubiquitous molecules. In fact, the differential abundance of any molecule may be influenced by subtle changes in the environmental or experimental conditions, and therefore the biological implications remain rather speculative.

In systems biology, high-throughput data can be integrated into computational models that describe the state of the whole system. Cellular metabolism is arguably the best described cellular system and the availability of reaction information (i.e. the conversions catalyzed by metabolic enzymes) makes GEMs particularly powerful tools to investigate this system (DeBerardinis and Thompson, 2012). Metabolism is profoundly connected to all other systems in the cell. A computational model of cell metabolism can therefore be used as a proxy to describe the phenotypic state of an organism (McKnight, 2010). A GEM is based on a metabolic network, assuming that enzymatic activity and substrate specificity of orthologs is conserved. The known metabolism of model organisms serve as Rosetta Stones for other organisms (Ideker *et al.*, 2001). Investigation of the static metabolic network can already provide insights into the biology of an organism that could be easily overlooked when only smaller subsets of the data would be considered. For instance, in **Chapter 2** we showed that oomycetes with an obligate parasitic lifestyle tend to lose genes coding for enzymes that predominantly function at the periphery of their metabolic network. More detailed insights into the metabolism of a species can be derived by reconstructing a full GEM (Orth *et al.*, 2010). Because of their integrative nature (relating genes with reactions and metabolites), GEMs serve two main purposes:

The first purpose of GEMs is to serve as a knowledge base for species-specific information about an organism's biochemical capacity, deduced from its genome and prior knowledge, and to provide a scaffold for the integration of additional omics data (O'Brien *et al.*, 2015). Prior knowledge typically integrated in a GEM includes, for instance, the nutrients taken up and the metabolites produced in the form of biomass or secondary metabolites. Importantly, prior knowledge should be used to correct the model where automated methods are limited, such as the inference of species-specific enzymatic substrates, gene-protein-reaction associations (i.e. which genes catalyze what reactions, including subunits and isozymes), enzyme subcellular localization, and reaction directionality and transporter substrates (Thiele and Palsson, 2010). Unfortunately, we found that prior knowledge and literature about the biochemical capacity of *Phytophthora infestans* are rather limited (**Chapter 3** and **4**). Note that GEMs can and should be updated when new data and information become available, which is an essential aspect of systems biology (see below).

The second main purpose of GEMs is enabling flux simulations that can be used to investigate system complexity and dynamics (Orth *et al.*, 2010). For instance, this can help unravel which fluxes are thermodynamically optimal for growth (biomass production) – sometimes in

various transcriptomic contexts. Moreover, these simulations can be used to investigate the system's robustness to induced perturbations (e.g. gene deletions). The testable hypotheses forthcoming from simulations are called model-driven hypotheses. This process is key for the cyclic process in which model predictions are experimentally validated, driving technological advance, allowing for the integration of new data, and again enabling generation of new hypotheses (Kitano, 2002).

In summary, GEMs are knowledge bases that can be used to calculate the metabolic fluxes of an organism. A continuous cycle of improvements with additional data and knowledge can eventually lead to a highly predictive model to provide a deeper understanding of the molecular systems of an organism.

Modelling P. infestans metabolism to predict pathogen-host interactions

The primary motivation for building a GEM for a pathogen – in this thesis *Phytophthora infestans* – is to understand its parasitic lifestyle. The infection of a plant by a parasite, such as a fungus or an oomycete, often involves a prolonged symbiosis in which the parasite feeds off the plant for growth and reproduction. Oomycetes have a similar morphology and ecological niche as fungi (Judelson and Blanco, 2005). However, compared to fungi there is a knowledge gap in oomycetes and many fundamental properties of their biochemical capacity and physiology are still unknown. For instance, the cytosol of *P. infestans* seems to contract to the tips of germ tubes (Kots *et al.*, 2017), suggesting that metabolism occurs predominantly at specific subcellular locations. This knowledge gap has multiple causes, one of which concerns the amenability to experimentation; many oomycetes are hard to culture and require complex media for *in vitro* growth, such as rye or pea broth. This limitation obviously complicates biochemical assays to investigate their metabolism, for which knowledge of the precise growth substrates is essential. Several oomycete pathogens are obligate biotrophs that exclusively grow inside their living host and are thus unculturable *in vitro* (McDowell, 2011). Another challenge is targeted mutagenesis for functional gene analyses. DNA transformation is feasible in several *Phytophthora* and *Pythium* species, but in most species transformation efficiencies are relatively low. Until recently, gene silencing or overexpression were the only methods available to manipulate the expression of a target gene in oomycetes. Because of the variability in silencing or overexpression levels the phenotypic characterisation of the transformants is not always straightforward. The first successful application of CRISPR-Cas9 mediated gene editing in oomycetes was published in 2016 (Fang and Tyler, 2016). This was a leap forward and nowadays it is applied in a few *Phytophthora* species (Pettongkhao *et al.*, 2020; W. Wang *et al.*, 2019). However, for still unknown reasons CRISPR-Cas mediated gene editing in *P. infestans* seems problematic (van den Hoogen and Govers, 2018) and so far successful employment in this species has not been reported.

The availability of knock-out mutants and information about their growth phenotypes on defined substrates greatly contributes to the acquisition of novel biochemical knowledge (Nakahigashi *et al.*, 2009), and is therefore pivotal for the reconstruction of GEMs and the validation of their predictions (Monk *et al.*, 2014). Unfortunately, due to the challenges described above with respect to targeted mutagenesis, experimental data on metabolism in *P. infestans* are very limited. The information that we found in publications included data of minimal *in vitro* growth substrates (**Chapter 3**) (Hohl, 1991), verified subcellular localizations of enzymes (Abrahamian *et al.*, 2017; López-Calcagno *et al.*, 2009), and capacity to produce a mixture of long-chain polyunsaturated fatty acids (Griffiths *et al.*, 2003c; Sun *et al.*, 2013). Perhaps the most important limitation to our work was the lack of knowledge on biomass composition, i.e. the stoichiometry of *P. infestans* biomass precursors (Feist and Palsson, 2010). The biomass composition relates the fluxes in the model to a hypothetical growth rate, and as such it can be used as a proxy for metabolic fitness. Because a precise description of *P. infestans* biomass composition was not available, we estimated it from literature, but ignored relative abundance (stoichiometry) (**Chapter 3**). This disabled quantitative flux predictions, but still allowed us to investigate the model for connectivity and importance of different nutrients (**Chapter 4**). Similar challenges were faced by others modelling pathogens. Tymoshenko *et al.* (2015) who published a GEM for human parasite *Toxoplasma gondii* also reconstructed a biomass composition from literature, ignoring stoichiometry. For a GEM of *Leishmania donovani*, Sharma *et al.* (2017) chose to infer the biomass composition from a *Plasmodium* GEM. In retrospect, we could have adopted the biomass composition from curated GEMs of closely related organisms such as the brown algae *Phaeodactylum tricornutum* (Levering *et al.*, 2016) or *Ectocarpus siliculosus* (Prigent *et al.*, 2014), although this would have introduced new biases and uncertainties. After all the similarity of biomass composition between closely related organisms and *P. infestans* is unknown and, to make it even more complicated,, the biomass composition of *P. infestans* in different life stages is radically different (Grenville-Briggs *et al.*, 2008).

In oomycetes the sporangia, which are asexual spores, likely rely on stored nutrient reserves, such as glucans and fatty acids, that are catabolized for energy production (Judelson, 2017). When sporangia disperse and reach a suitable plant surface, zoospores are released and encyst. The cysts then germinate and form an appressorium at the tip of the germ tube to penetrate the epidermal cells of the plant, likely mediated by secreted cell wall degrading enzymes (Ospina-Giraldo *et al.*, 2010). Haustoria (vesicle-like structures) emerge from hyphae that colonize the apoplast and enter the mesophyll cells (Judelson, 2017). It is often assumed that these haustoria are the main site of nutrient uptake, as is the case for various plant pathogenic fungi (Wang *et al.*, 2018b). However, many oomycetes do not form haustoria (Fawke *et al.*, 2015), and haustoria make up only a very small proportion of the total hyphal biomass (~2%), raising the question whether haustoria are truly the main site of nutrient uptake (Judelson and Ah-Fong, 2018). The plant apoplast is a nutrient-rich environment, and might be the main site of nutrient uptake (Chen, 2014). Haustoria are nonetheless very important for the host-pathogen interaction. It is the site from where the pathogen deposits so-called cytoplasmic

effectors into the plant cell for suppression of immune responses (Boevink *et al.*, 2020). The host recognizes the intracellular host-pathogen interface created by the haustorium as the site where defence responses have to be activated and for example, relocates the nucleus to the interface (Wang *et al.*, 2017). Next to cytoplasmic effectors the pathogen secretes apoplastic effectors, so also in the apoplast host and pathogen interact and likely this involves exchange of signals and compounds.

Because of the specialized tasks of the haustorium and hyphae in the apoplast, hyphal cells likely have a "division of labour", which implies that the biological processes are tailored for the specific region of infection. This phenomenon was recently modelled for the fungal plant pathogen *Sclerotium sclerotiorum*, by mapping the transcriptome of the apex and the center of infection to a multi-cell GEM (Peyraud *et al.*, 2019). During reconstruction of the *P. infestans* GEMs (**Chapter 3** and **4**), we did not explicitly discriminate between the different sites of infection. Novel approaches, such as the integration of single-cell RNA-Seq into our GEM could provide insight into the metabolic processes delegated throughout the hyphae (Penaranda and Hung, 2019; Rohlenova *et al.*, 2020).

Our models demonstrated how they can be used to integrate the current knowledge of an organism (or pathosystem), and how to analyze transcriptome data in a system-wide context. In concordance with related transcriptome studies (Abrahamian *et al.*, 2016; Ah-Fong, Kim, *et al.*, 2017), the transcriptome-based submodels reflected reduced metabolic activity in the sporangial stages of *P. infestans,* and nutritional changes in the transition from a biotrophic to a necrotrophic stage of infection on tomato leaves. Importantly, because we analyzed these transcriptomic changes in the context of a GEM, results were subject to the imposed model constraints (steady-state, reaction thermodynamics) and thereby to the topology of the metabolic network (Hyduke *et al.*, 2013). Clearly, any model-based hypotheses remain to be tested, but we do demonstrate a transcriptome analysis from an alternative angle. The transcriptomic changes are interpreted in terms of ensuing differences of metabolic fluxes, and this system-wide approach may give more insight than the differential expression analysis of individual genes.

The models for *P. infestans* we reconstructed were subdivided to represent the spatial distribution of metabolic pathways in different subcellular compartments. The transporters and channels that transport metabolic substrates across membranes are modelled by introduction of transport reactions (Thiele and Palsson, 2010). Unfortunately, transporters typically have a wide substrate range, which cannot be reliably predicted from their protein sequence, and are therefore often manually added to a GEM based on prior knowledge. This was particularly challenging when creating the pathogen-host interaction model (**Chapter 4**), considering that membrane transport is pivotal to pathogen nutrition. Little is known about *P. infestans* nutrition *in planta*, and manual addition of host-pathogen transport reactions would bias fluxes towards a predefined set of nutrient transporters. Since one of our goals was to predict the nutrient pool of *P. infestans* during tomato infection, we chose not to manually add transport reactions, and based our conclusions on the optimal fluxes in the

model. In other words, the transport reactions in our models were largely based on network topology. Depending on the objective function, the most optimal set of transporters had a nonzero flux. The downside of this approach was that we could not consider bidirectional transport, because this would also imply unrestricted metabolite exchange between host and pathogen. This could lead to scenarios where the host would utilize *P. infestans* metabolism for profit, which, from a biological point of view, is not plausible. In reality, however, pathogen-host metabolite exchange is likely to occur both ways, e.g. to secrete metabolites as waste products or virulence factors. There is a clear knowledge gap on the metabolic exchanges that *P. infestans* maintains with its environment, and this should be addressed by a broad substrate screening. For instance, mass spectrometry could be used to perform comparative metabolomics on a growth medium, at different time points of mycelial growth.

Outlook

Systems biology has been recognized years ago as a suitable method to study plant pathogens in general and *P. infestans* specifically (Pinzón *et al.*, 2009; Pritchard and Birch, 2011). In the last few years, the potato and tomato late blight pathosystem with *P. infestans* as pathogen was subject of several systems biology studies (D. Botero *et al.*, 2018; K. Botero *et al.*, 2018; Castro *et al.*, 2019; Rodenburg *et al.*, 2018a; Rodenburg *et al.*, 2019; Seidl *et al.*, 2013; Thines *et al.*, 2020). However, systems biology of this pathosystem – as goes for many – is still in its infancy, and many fundamental factors are still unknown. The success of a systems biology approach is subject to the level of knowledge on the organism to be modelled. Thus, in order to arrive at highly predictive models for *P. infestans*, *in vitro* experiments need to be performed to gain basal knowledge. Valuable information would be, for instance, the substrates *P. infestans* can assimilate from its environment, as well as its biomass composition and how this changes throughout its lifecycle. There is a lot we can learn from the more advanced metabolic research in other pathosystems. For instance, there are now several GEMs for *Plasmodium* spp. – some of which are also integrated with GEMs of the host, the red blood cell (Abdel-Haleem *et al.*, 2018; Huthmacher *et al.*, 2010; Plata *et al.*, 2010). Initially, these GEMs were reconstructed based on metabolomics, identifying growth substrates (Olszewski *et al.*, 2009). Later, *Plasmodium* GEMs were further curated, refined and integrated with new omics data and novel biochemical knowledge (Bazzani *et al.*, 2012; Carey *et al.*, 2017; Stanway *et al.*, 2019). Interestingly, these models have pinpointed several essential reactions, some of which turned out to be leads for promising drug targets (O'Hara *et al.*, 2014). For protozoan parasites, isotope-labelled growth experiments have been successful to dissect their metabolism during parasitic growth (Kloehn *et al.*, 2016). For *P. infestans*, similar analyses would provide intriguing novel avenues. Considering the current advances in oomycete research, we can speculate on the achievements that are attainable in the coming ten years.

In the future, several steps are needed to create a higher-quality GEM for *P. infestans* (Thiele and Palsson, 2010). Obviously, a complete functional characterization of the substrates and characteristics of each individual *P. infestans* metabolic enzyme would be ideal, but this seems infeasible in the near future. Nevertheless, significant achievements could be gained from *in silico* and *in vitro* procedures, designed specifically for the purpose of building a high-quality GEM. Below we summarize the points discussed above in a stepwise protocol (**Figure 2**):

The first prerequisite is a (near-)complete, gapless genome sequence of *P. infestans*. The genome needs to be resequenced and assembled using novel technologies and more advanced assembly methods to attain the complete coding information (Thomma *et al.*, 2016). Gene prediction should be guided by RNA sequencing and homology-based evidence, and predicted protein sequences need to be functionally annotated using sequence information from the available biochemical databases. Manual gene model curation of the predicted metabolic enzymes should be performed to yield accurate insight into its biochemical capacity (Fernandes *et al.*, 2019). Such manual curation is a time- and labour- intensive process, but will ultimately lead to more accurate hypotheses.

The identified putative enzymes lead to a reconstruction of a draft metabolic network, which should be separated over several cellular compartments. We consider at least the cytosol, extracellular space and the mitochondria essential, since these are the main metabolic hotspots. Enzyme annotation and subcellular localization should be curated according to literature. The associated reactions and metabolites should be curated according to the established protocol (Thiele and Palsson, 2010) and protein subunits of metabolic enzymes should be annotated to correctly consider the gene-protein-reaction associations. The constructed metabolic network can be inspected to identify potentially missing enzymes by comparing the reconstructed pathway to reference pathways in model organisms, and the annotations should be revisited for missing enzymes. In addition, intracellular transport reactions should be included according to common biochemical knowledge (e.g. textbooks), GEMs of other species, and network topology. It is unlikely that many more transporters are to be characterized, given the labour-intensive process (Savory *et al.*, 2018).

In vitro growth assays should be performed to characterize *P. infestans* growth, using the same strain as the one sequenced. The ideal growth medium is similar in composition to a potato or tomato leaf, to mimic natural growth. The growth medium should be analyzed by untargeted metabolomics (mass spectrometry and/or nuclear magnetic resonance) over multiple time points to provide insight into the presence and abundance of specific substances. Metabolites that strongly change in abundance during *P. infestans* growth and between subsequent sampling stages are likely assimilated or secreted. This can be indicative for a nutrient transporter on the plasma membrane that is capable of transporting the respective metabolite. In addition, radioactively labelled metabolites can be added to the medium to test the assimilation of specific nutrients (Ah-Fong *et al.*, 2019), such as carbohydrates and lipids. For metabolites for which changes in abundance are measured, uptake and demand (transport) reactions should be added to the model.

The relative *P. infestans* biomass composition and growth rates should be measured. A promising approach for this is Fourier-Transform Infrared Spectroscopy (FTIR) (Mayers *et al.*, 2013). This method was optimized for analyses of brown algae, and was successfully used in the reconstruction of a GEM for the diatom *Phaeodactylum tricornutum* to quantify the percentages of carbohydrate, protein, DNA/RNA, and fatty acids per gram of cellular dry weight (Levering *et al.*, 2016). Once the main classes of biomass components are quantified and are related to growth rate, more specific metabolites can be calculated based on traditional metabolomics methods.

Predicted phenotypes, e.g. by specific nutrient starvation, should be validated by growth experiments and gene/reaction essentiality should be validated in knock-out or knock-down mutants, for instance as was done for the nitrate assimilation cluster (Abrahamian *et al.*, 2016). We anticipate that CRISPR-Cas9 gene editing will be successfully employed in the coming years, but as an alternative, gene silencing mutants can be generated to investigate metabolic perturbations. The model should be updated with novel findings, and discrepancies should lead to corrections of the model. For instance, when the knock-out of a predicted essential gene is not lethal *in vitro*, there are likely alternative enzymes or metabolic routes that compensate for this mutation. The model should be inspected on incorrect annotations or missing reactions accordingly.

The refined GEM for *P. infestans* could also be integrated with a similarly refined GEM for tomato or potato. More sophisticated constraints and objective functions could be deployed to simulate a more realistic symbiosis for this pathosystem, such as multi-objective simulations to address the competition for nutrients (Jamshidi and Raghunathan, 2015).

In addition to these steps, there is a large and rapidly increasing number of methods and algorithms that can be applied to GEMs to gain insights into the complex system of pathogen-host interactions (Lewis *et al.*, 2012). For instance, regulatory networks could be inferred from (anti-)correlated expression patterns in dual RNA-Seq data and other experimental data, and integrated into GEMs to further constrain the fluxes, in order to learn how metabolism is regulated during infection (Peyraud *et al.*, 2018). Collectively, we anticipate that these steps will lead to a high-quality GEM of *P. infestans*, that can be used to identify novel targets for late blight control, and help understand how oomycetes in general interact with plants.

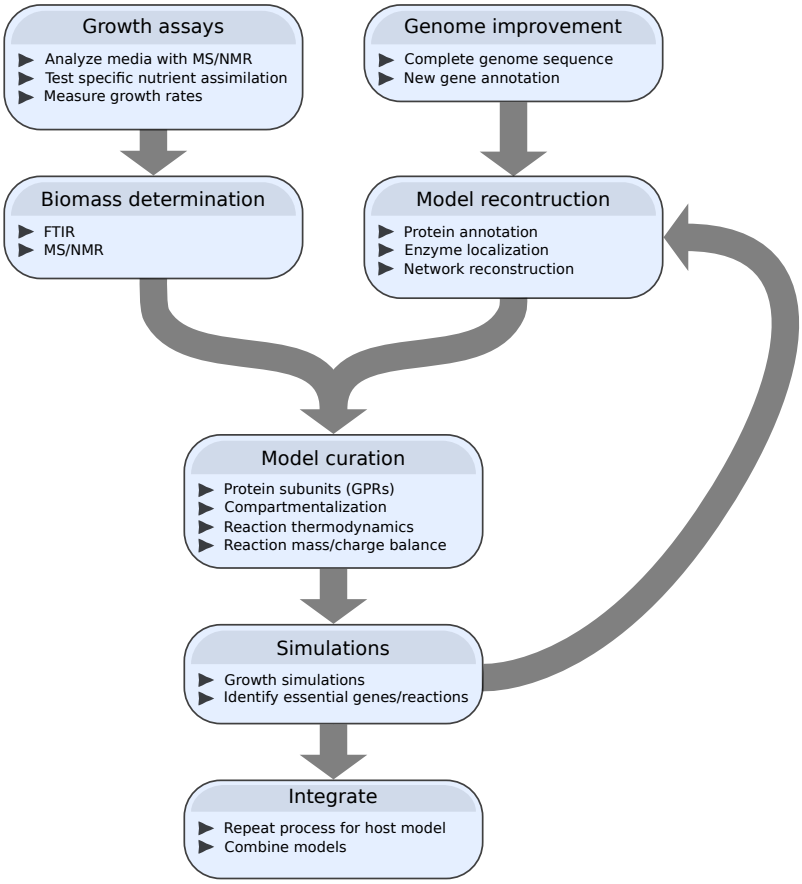
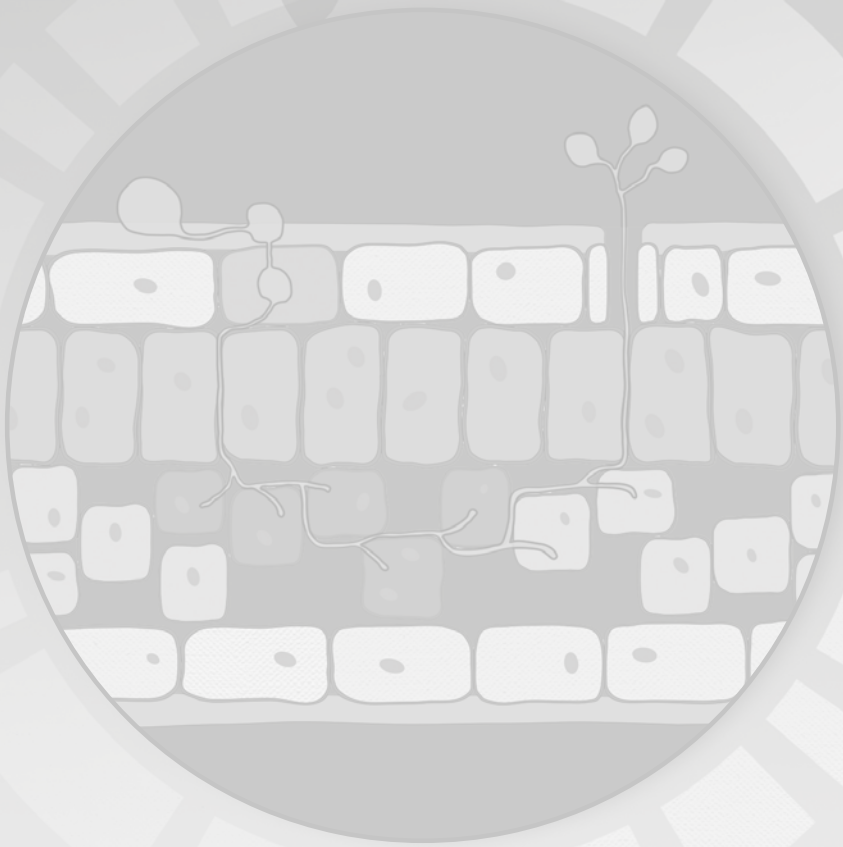


FIGURE 2 | A proposed workflow (including possible methods and analyses) for reconstructing a high-quality genome scale metabolic model (GEM) for *P. infestans* and and, potentially, an integrated GEM for a *P. infestans*/host interaction. Model reconstruction is an iterative process of simulation and model improvement.

Conclusions

Systems biology approaches are relatively new in plant pathology, yet models are promising tools to study plant-pathogen interactions. Here we discussed the interpretation of a GEM for *P. infestans*, reconstructed from its genome sequence and available information in literature. There is still much to gain from basal metabolic/biochemical knowledge, which should be exploited and integrated to build higher-quality models in the future. This will contribute to an increased understanding of oomycetes in general

References



References

Abdel-Haleem, A.M., Hefzi, H., Mineta, K., Gao, X., Gojobori, T., Palsson, B.O., Lewis, N.E. and Jamshidi, N. (2018) Functional interrogation of *Plasmodium* genus metabolism identifies species- and stage-specific differences in nutrient essentiality and drug targeting. *PLoS Comput. Biol.* **14**.

Abrahamian, M., Ah-Fong, A.M., Davis, C., Andreeva, K. and Judelson, H.S. (2016) Gene expression and silencing studies in *Phytophthora infestans* reveal infection-specific nutrient transporters and a role for the nitrate reductase pathway in plant pathogenesis. *PLoS Pathog.* **12**, e1006097.

Abrahamian, M., Kagda, M., Ah-Fong, A.M.V. and Judelson, H.S. (2017) Rethinking the evolution of eukaryotic metabolism: Novel cellular partitioning of enzymes in stramenopiles links serine biosynthesis to glycolysis in mitochondria. *BMC Evol. Biol.* **17**, 241.

Aderem, A. (2005) Systems Biology: Its Practice and Challenges. *Cell* **121**, 511–513.

Adhikari, B.N., Hamilton, J.P., Zerillo, M.M., Tisserat, N., Lévesque, C.A. and Buell, C.R. (2013) Comparative genomics reveals insight into virulence strategies of plant pathogenic oomycetes. *PLoS ONE* **8**, e75072.

Agren, R., Bordel, S., Mardinoglu, A., Pornputtapong, N., Nookaew, I. and Nielsen, J. (2012) Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT. *PLoS Comput. Biol.* **8**, e1002518.

Agren, R., Liu, L., Shoaie, S., Vongsangnak, W., Nookaew, I. and Nielsen, J. (2013) The RAVEN toolbox and its use for generating a genome-scale metabolic model for *Penicillium chrysogenum*. *PLoS Comput. Biol.* **9**.

Ah-Fong, A.M., Kagda, M.S., Abrahamian, M. and Judelson, H.S. (2019) Niche-specific metabolic adaptation in biotrophic and necrotrophic oomycetes is manifested in differential use of nutrients, variation in gene content, and enzyme evolution. *PLoS Pathog.* **15**, e1007729.

Ah-Fong, A.M., Kim, K.S. and 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**, 198–198.

Ah-Fong, A.M., Shrivastava, J. and 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.

Albalat, R. and Cañestro, C. (2016) Evolution by gene loss. *Nat. Rev. Genet.* **17**, 379–391.

Albert, R. (2007) Network Inference, Analysis, and Modeling in Systems Biology. *Plant Cell* **19**, 3327–3338.

Ali, S.S., Shao, J., Lary, D.J., et al. (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.

Aller, S., Scott, A., Sarkar-Tyson, M. and Soyer, O.S. (2018) Integrated human-virus metabolic stoichiometric modelling predicts host-based antiviral targets against Chikungunya, Dengue and Zika viruses. *J. R. Soc. Interface* **15**, 20180125.

Andrews, S. (2010) Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data,.

Ankrah, N.Y.D., Chouaia, B. and Douglas, A.E. (2018) The cost of metabolic interactions in symbioses between insects and bacteria with reduced genomes. *mBio* **9**, e01433–18.

Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S. and Ogata, H. (2019) KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* **btz859**.

Armbrust, E.V., Berges, J.A., Bowler, C., et al. (2004) The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* **306**, 79–86.

Asai, S., Rallapalli, G., Piquerez, S.J.M., et al. (2014) Expression profiling during arabidopsis/downy mildew interaction reveals a highly-expressed effector that attenuates responses to salicylic acid. *PLoS Pathog.* **10**, e1004443.

Azevedo, R.A., Lancien, M. and Lea, P.J. (2006) The aspartic acid metabolic pathway, an exciting and essential pathway in plants. *Amino Acids* **30**, 143–162.

Bartocci, E. and Lió, P. (2016) Computational Modeling, Formal Analysis, and Tools for Systems Biology Kwiatkowska, M.Z., ed, ACM.

Basenko, E., Pulman, J., Shanmugasundram, A., et al. (2018) FungiDB: An Integrated Bioinformatic Resource for Fungi and Oomycetes. *J. Fungi* **4**, 39.

Bastian, M., Heymann, S. and Jacomy, M. (2009) Gephi: an open source software for exploring and manipulating networks. *Third Int. AAAI Conf. Weblogs Soc. Media*, 361–362.

Basu, S., Patil, S., Mapleson, D., et al. (2017) Finding a partner in the ocean: molecular and evolutionary bases of the response to sexual cues in a planktonic diatom. *New Phytol.* **215**, 140–156.

Baxter, L., Tripathy, S., Ishaque, N., et al. (2010) Signatures of adaptation to obligate biotrophy in the *Hyaloperonospora arabidopsidis* genome. *Science* **330**, 1549–1551.

Bazzani, S., Hoppe, A. and Holzhütter, H.-G. (2012) Network-based assessment of the selectivity of metabolic drug targets in *Plasmodium falciparum* with respect to human liver metabolism. *BMC Syst. Biol.* **6**, 118.

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

Belda, E., Silva, F.J., Peretó, J. and Moya, A. (2012) Metabolic networks of *Sodalis glossinidius*: A systems biology approach to reductive evolution. *PLoS ONE* **7**, e30652.

Bellincampi, D., Cervone, F. and Lionetti, V. (2014) Plant cell wall dynamics and wall-related susceptibility in plant-pathogen interactions. *Front. Plant Sci.* **5**, 228–228.

Benhamou, N., Floch, G. le, Vallance, J., Gerbore, J., Grizard, D. and Rey, P. (2012) *Pythium oligandrum*: an example of opportunistic success. *Microbiology,* **158**, 2679–2694.

Benitez, J.A. and Silva, A.J. (2016) *Vibrio cholerae* hemagglutinin(HA)/protease: An extracellular metalloprotease with multiple pathogenic activities. *Toxicon Off. J. Int. Soc. Toxinology* **115**, 55–62.

Berger, H., Yacoub, A., Gerbore, J., Grizard, D., Rey, P., Sessitsch, A. and Compant, S. (2016) Draft genome sequence of biocontrol agent *Pythium oligandrum* strain P037, an oomycota. *Genome Announc.* **4**.

Beste, D.J.V., Nöh, K., Niedenführ, S., Mendum, T.A., Hawkins, N.D., Ward, J.L., Beale, M.H., Wiechert, W. and McFadden, J. (2013) 13C-flux spectral analysis of host-pathogen metabolism reveals a mixed diet for intracellular *Mycobacterium tuberculosis*. *Chem. Biol.* **20**, 1012–21.

Bhardwaj, R., Das, M., Singh, S., Chiranjivi, A.K., Prabhu, S.V., Singh, S.K. and Dubey, V.K. (2017) Evaluation of CAAX prenyl protease II of *Leishmania donovani* as potential drug target: Infectivity and growth of the parasite is significantly lowered after the gene knockout. *Eur. J. Pharm. Sci.* **102**, 156–160.

Bland, N.D., Pinney, J.W., Thomas, J.E., Turner, A.J. and Isaac, R.E. (2008) Bioinformatic analysis of the neprilysin (M13) family of peptidases reveals complex evolutionary and functional relationships. *BMC Evol. Biol.* **8**, 16–16.

Blume, M. and Seeber, F. (2018) Metabolic interactions between *Toxoplasma gondii* and its host. *F1000Research* **7**, 1719.

Boevink, P.C., Birch, P.R., Turnbull, D. and Whisson, S.C. (2020) Devastating intimacy: the cell biology of plant-*Phytophthora* interactions. *New Phytol.*

Bordbar, A., Lewis, N.E., Schellenberger, J., Palsson, B.O. and Jamshidi, N. (2010) Insight into human alveolar macrophage and *M. tuberculosis* interactions via metabolic reconstructions. *Mol. Syst. Biol.* **6**.

Bordbar, A., Monk, J.M., King, Z.A. and Palsson, B.O. (2014) Constraint-based models predict metabolic and associated cellular functions. *Nat. Rev. Genet.* **15**, 107–120.

Borenstein, E. and Feldman, M.W. (2009) Topological signatures of species interactions in metabolic networks. *J. Comput. Biol. J. Comput. Mol. Cell Biol.* **16**, 191–200.

Botero, D., Valdés, I., Rodríguez, M.-J., Henao, D., Danies, G., González, A.F. and Restrepo, S. (2018) A genome-scale metabolic reconstruction of *Phytophthora infestans* with the integration of transcriptional data reveals the key metabolic patterns involved in the interaction of its host. *Front. Genet.* **9**, 244.

Botero, K., Restrepo, S. and Pinzón, A. (2018) A genome-scale metabolic model of potato late blight suggests a photosynthesis suppression mechanism. *BMC Genomics* **19**, 863.

Bowler, C., Allen, A.E., Badger, J.H., et al. (2008) The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* **456**, 239–244.

Bray, N.L., Pimentel, H., Melsted, P. and Pachter, L. (2016) Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527.

Breitling, R. (2010) What is Systems Biology? *Front. Physiol.* **1**.

Brouwer, H., Coutinho, P.M., Henrissat, B. and Vries, R.P. de (2014) Carbohydrate-related enzymes of important *Phytophthora* plant pathogens. *Fungal Genet. Biol.* **72**, 192–200.

Brouwer, S., Cork, A.J., Ong, C.-L.Y., Barnett, T.C., West, N.P., McIver, K.S. and Walker, M.J. (2018) Endopeptidase PepO Regulates the SpeB Cysteine Protease and Is Essential for the Virulence of Invasive M1T1 *Streptococcus pyogenes*. *J. Bacteriol.* **200**, e00654-17.

Buchfink, B., Xie, C. and Huson, D.H. (2015) Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60.

Burton, J.N., Adey, A., Patwardhan, R.P., Qiu, R., Kitzman, J.O. and Shendure, J. (2013) Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat. Biotechnol.* **31**, 1119–1125.

Carbonell, P., Lecointres, G. and Faulon, J.L. (2011) Origins of specificity and promiscuity in metabolic networks. *J. Biol. Chem.* **286**, 43994–44004.

Carella, P., Gogleva, A., Hoey, D.J., Bridgen, A.J., Stolze, S.C., Nakagami, H. and Schornack, S. (2019) Conserved biochemical defenses underpin host responses to oomycete infection in an early-divergent land plant lineage. *Curr. Biol.* **29**, 2282–2294.e5.

Carere, J., Colgrave, M.L., Stiller, J., Liu, C., Manners, J.M., Kazan, K. and Gardiner, D.M. (2016) Enzyme-driven metabolomic screening: a proof-of-principle method for discovery of plant defence compounds targeted by pathogens. *New Phytol.* **212**, 770–779.

Carey, M.A., Papin, J.A. and Guler, J.L. (2017) Novel *Plasmodium falciparum* metabolic network reconstruction identifies shifts associated with clinical antimalarial resistance. *BMC Genomics* **18**, 543.

Casadevall, A. (2008) Evolution of intracellular pathogens. *Annu. Rev. Microbiol.* **62**, 19–33.

Caspi, R., Altman, T., Billington, R., et al. (2014) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res.* **42**, D459–71.

Castro, J.C., Valdés, I., Gonzalez-García, L.N., Danies, G., Cañas, S., Winck, F.V., Ñústez**, C.E., Restrepo, S. and Riaño-Pachón, D.M.** (2019) Gene regulatory networks on transfer entropy (GRNTE): a novel approach to reconstruct gene regulatory interactions applied to a case study for the plant pathogen *Phytophthora infestans*. *Theor. Biol. Med. Model.* **16**, 7.

Cesur, M.F., Abdik, E., Güven-Gülhan, Ü., Durmus, S. and Çakir**, T.** (2018) Computational systems biology of metabolism in infection. In pp. 235–282. Springer, Cham.

Chan, X.W.A., Wrenger, C., Stahl, K., Bergmann, B., Winterberg, M., Müller, I.B. and Saliba, K.J. (2013) Chemical and genetic validation of thiamine utilization as an antimalarial drug target. *Nat. Commun.* **4**, 2060.

Charrad, M., Ghazzali, N., Boiteau, V. and Niknafs, A. (2014) NbClust: An R package for determining the relevant number of clusters in a data set. *J. Stat. Softw.* **61**.

Chauhan, N. and Singh, S. (2019) Integrative computational framework for understanding metabolic modulation in *Leishmania*. *Front. Bioeng. Biotechnol.* **7**.

Chavali, A.K., D 'auria, K.M., Hewlett, E.L., Pearson, R.D. and Papin, J.A. (2012) A metabolic network approach for the identification and prioritization of antimicrobial drug targets. *Trends Microbiol.* **20**, 113–123.

Chen, L.Q. (2014) SWEET sugar transporters for phloem transport and pathogen nutrition. *New Phytol.* **201**, 1150–1155.

Chen, X., Fang, X., Zhang, Y., et al. (2019) Overexpression of a soybean 4-coumaric acid: coenzyme A ligase (GmPl4L) enhances resistance to *Phytophthora sojae* in soybean. *Funct. Plant Biol.* **46**, 304–313.

Chen, Y., Chi, H. yun, Meesapyodsuk, D. and Qiu, X. (2013) *Phytophthora infestans* cholinephosphotransferase with substrate specificity for very-long-chain polyunsaturated fatty acids. *Appl. Environ. Microbiol.* **79**, 1573–1579.

Cock, J.M., Sterck, L., Rouzé, P., et al. (2010) The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. *Nature* **465**, 617–621.

Cook, D.E., Mesarich, C.H. and Thomma, B.P.H.J. (2015) Understanding plant immunity as a surveillance system to detect invasion. *Annu. Rev. Phytopathol.* **53**, 541–563.

Corradi, N. (2015) Microsporidia: eukaryotic intracellular parasites shaped by gene loss and horizontal gene transfers. *Annu. Rev. Microbiol.* **69**, 167–183.

Csardi, G. and Nepusz, T. (2006) The igraph software package for complex network research. *InterJournal Complex Syst.* **1695**, 1–9.

Dahlin, P., Srivastava, V., Ekengren, S., McKee, L.S. and Bulone, V. (2017) Comparative analysis of sterol acquisition in the oomycetes *Saprolegnia parasitica* and *Phytophthora infestans*. *PLoS ONE* **12**, e0170873.

Danies, G., Small, I.M., Myers, K., Childers, R. and Fry, W.E. (2013) Phenotypic characterization of recent clonal lineages of *Phytophthora infestans* in the United States. *Plant Dis.* **97**, 873–881.

Dean, P., Major, P., Nakjang, S., Hirt, R.P. and Embley, T.M. (2014) Transport proteins of parasitic protists and their role in nutrient salvage. *Front. Plant Sci.* **5**, 153.

Dean, P., Sendra, K.M., Williams, T.A., et al. (2018) Transporter gene acquisition and innovation in the evolution of microsporidia intracellular parasites. *Nat. Commun.* **9**, 1–12.

Dean, R., Van Kan, J.A.L., Pretorius, Z.A., et al. (2012) The Top 10 fungal pathogens in molecular plant pathology. *Mol. Plant Pathol.* **13**, 414–430.

DeBerardinis, R.J. and Thompson, C.B. (2012) Cellular metabolism and disease: what do metabolic outliers teach us? *Cell* **148**, 1132–1144.

Denton, J.F., Lugo-Martinez, J., Tucker, A.E., Schrider, D.R., Warren, W.C. and Hahn, M.W. (2014) Extensive error in the number of genes inferred from draft genome assemblies. *PLoS Comput. Biol.* **10**, e1003998–e1003998.

Derevnina, L., Dagdas, Y.F., Concepcion, J.C.D. la, et al. (2016) Nine things to know about elicitins. *New Phytol.* **212**, 888–895.

Derevnina, L., Petre, B., Kellner, R., et al. (2016) Emerging oomycete threats to plants and animals. *Philos. Trans. R. Soc. B Biol. Sci.* **371**.

Deu, E. (2017) Proteases as antimalarial targets: strategies for genetic, chemical, and therapeutic validation. *FEBS J.* **284**, 2604–2628.

Dharmawardhana, P., Ren, L., Amarasinghe, V., Monaco, M., Thomason, J., Ravenscroft, D., McCouch, S., Ware, D. and Jaiswal, P. (2013) A genome scale metabolic network for rice and accompanying analysis of tryptophan, auxin and serotonin biosynthesis regulation under biotic stress. *Rice* **6**, 1–15.

Diéguez-Uribeondo, J., García, M.A., Cerenius, L., et al. (2009) Phylogenetic relationships among plant and animal parasites, and saprotrophs in *Aphanomyces* (oomycetes). *Fungal Genet. Biol.* **46**, 365–376.

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

Dix, A., Vlaic, S., Guthke, R. and Linde, J. (2016) Use of systems biology to decipher host-pathogen interaction networks and predict biomarkers. *Clin. Microbiol. Infect. Off. Publ. Eur. Soc. Clin. Microbiol. Infect. Dis.* **22**, 600–606.

Du, H. and Liang, C. (2019) Assembly of chromosome-scale contigs by efficiently resolving repetitive sequences with long reads. *Nat. Commun.* **10**, 1–10.

Duan, G., Christian, N., Schwachtje, J., Walther, D. and Ebenhöh, O. (2013) The metabolic interplay between plants and phytopathogens. *Metabolites* **3**, 1–23.

Dunphy, L.J. and Papin, J.A. (2018) Biomedical applications of genome-scale metabolic network reconstructions of human pathogens. *Curr. Opin. Biotechnol.* **51**, 70–79.

Durmus, S., Çakır**, T.,** Özgür**, A. and Guthke, R.** (2015) A review on computational systems biology of pathogen-host interactions. *Front. Microbiol.* **6**, 235.

Dussert, Y., Gouzy, J., Richart-Cervera, S., et al. (2016) Draft genome sequence of *Plasmopara viticola*, the grapevine downy mildew pathogen. *Genome Announc.* **4**.

Eberl, G. (2018) Robustness in living organisms is homeostasis. *Semin. Immunol.* **36**, 56–57.

Ebrahim, A., Lerman, J.A., Palsson, B.O. and Hyduke, D.R. (2013) COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Syst. Biol.* **7**, 74.

Eddy, S.R. (1998) Profile hidden Markov models. *Bioinforma. Oxf. Engl.* **14**, 755–763.

Edwards, D.R., Handsley, M.M. and Pennington, C.J. (2008) The ADAM metalloproteinases. *Mol. Aspects Med.* **29**, 258–289.

Edwards, J.S. and Palsson, B.O. (1999) Systems properties of the *Haemophilus influenzae* Rd metabolic genotype. *J. Biol. Chem.* **274**, 17410–17416.

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

Emms, D.M. and Kelly, S. (2015) OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157–157.

Emms, D.M. and Kelly, S. (2017) STRIDE: Species tree root inference from gene duplication events. *Mol. Biol. Evol.* **34**, 3267–3278.

Escotte-Binet, S., Huguenin, A., Aubert, D., Martin, A.P., Kaltenbach, M., Florent, I. and Villena, I. (2018) Metallopeptidases of *Toxoplasma gondii*: in silico identification and gene expression. *Parasite Paris Fr.* **25**, 26–26.

Faino, L., Seidl, M.F., Datema, E., Berg, G.C.M. van den, Janssen, A., Wittenberg, A.H.J. and Thomma, B.P.H.J. (2015) Single-molecule real-time sequencing combined with optical mapping yields completely finished fungal genome. *mBio* **6**.

Fang, Y. and Tyler, B.M. (2016) Efficient disruption and replacement of an effector gene in the oomycete *Phytophthora sojae* using CRISPR/Cas9. *Mol. Plant Pathol.* **17**, 127–139.

Fatima, U. and Senthil-Kumar, M. (2015) Plant and pathogen nutrient acquisition strategies. *Front. Plant Sci.* **6**, 750.

Fawke, S., Doumane, M. and Schornack, S. (2015) Oomycete interactions with plants: infection strategies and resistance principles. *Microbiol. Mol. Biol. Rev.* **79**, 263–280.

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

Feist, A.M. and Palsson, B.O. (2010) The Biomass Objective Function. *Curr. Opin. Microbiol.* **13**, 344–349.

Feng, C., Lamour, K.H., Bluhm, B.H., Sharma, S., Shrestha, S., Dhillon, B.D.S. and Correll, J.C. (2018) Genome sequences of three races of *Peronospora effusa*: a resource for studying the evolution of the spinach downy mildew pathogen. *Mol. Plant-Microbe Interact. MPMI* **31**, 1230–1231.

Fernandes, B.S., Dias, O., Costa, G., et al. (2019) Genome-wide sequencing and metabolic annotation of *Pythium irregulare* CBS 494.86: Understanding eicosapentaenoic acid production. *BMC Biotechnol.* **19**, 41.

Finn, R.D., Coggill, P., Eberhardt, R.Y., et al. (2016) The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Res.* **44**, D279–D285.

Fletcher, K., Gil, J., Bertier, L.D., et al. (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. and Michelmore, R. (2018) Comparative genomics of downy mildews reveals potential adaptations to biotrophy. *BMC Genomics* **19**, 851.

Fothergill, P.G. and Child, J.H. (1964) Comparative studies of mineral nutrition of three species of *Phytophthora*. *J. Gen. Microbiol.* **36**, 67–.

Franceschetti, M., Maqbool, A., Jiménez-Dalmaroni, M.J., Pennington, H.G., Kamoun, S. and Banfield, M.J. (2017) Effectors of filamentous plant pathogens: Commonalities amid diversity. *Microbiol. Mol. Biol. Rev. MMBR* **81**, e00066-16.

Frantzeskakis, L., Pietro, A.D., Rep, M., Schirawski, J., Wu, C.-H. and Panstruga, R. (2020) Rapid evolution in plant–microbe interactions – a molecular genomics perspective. *New Phytol.* **225**, 1134–1142.

Frischkorn, K.R., Harke, M.J., Gobler, C.J. and Dyhrman, S.T. (2014) De novo assembly of *Aureococcus anophagefferens* transcriptomes reveals diverse responses to the low nutrient and low light conditions present during blooms. *Front. Microbiol.* **5**, 375.

Fry, W.E. (2016) *Phytophthora infestans*: New tools (and old ones) lead to new understanding and precision management. *Annu. Rev. Phytopathol.* **54**, 529–547.

Galeano Garcia, P., Neves dos Santos, F., Zanotta, S., Eberlin, M. and Carazzone, C. (2018) Metabolomics of *Solanum lycopersicum* infected with *Phytophthora infestans* leads to early detection of late blight in asymptomatic plants. *Molecules* **23**, 3330.

Gao, R., Cheng, Y., Wang, Y., Wang, Y., Guo, L. and Zhang, G. (2015) Genome sequence of *Phytophthora fragariae* var. fragariae, a quarantine plant-pathogenic fungus. *Genome Announc.* **3**.

Garavito, M.F., Narvaez-Ortiz, H.Y., Pulido, D.C., Löffler, M., Judelson, H.S., Restrepo, S. and Zimmermann, B.H. (2019) *Phytophthora infestans* dihydroorotate dehydrogenase is a potential target for chemical control – a comparison with the enzyme from *Solanum tuberosum*. *Front. Microbiol.* **10**, 1479.

García Sánchez, C.E. and Torres Sáez, R.G. (2014) Comparison and analysis of objective functions in flux balance analysis. *Biotechnol. Prog.* **30**, 985–991.

Gardner, M.J., Hall, N., Fung, E., et al. (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* **419**, 498–511.

Gaulin, E., Pel, M.J.C., Camborde, L., et al. (2018) Genomics analysis of *Aphanomyces* spp. identifies a new class of oomycete effector associated with host adaptation. *BMC Biol.* **16**, 43–43.

Goldberg, T., Hecht, M., Hamp, T., et al. (2014) LocTree3 prediction of localization. *Nucleic Acids Res.* **42**, W350–W355.

Gómez Luciano, L.B., Tsai, I.J., Chuma, I., et al. (2019) Blast fungal genomes show frequent chromosomal changes, gene gains and losses, and effector gene turnover. *Mol. Biol. Evol.* **36**, 1148–1161.

Gomez, M.A., Contreras, I., Hallé, M., Tremblay, M.L., McMaster, R.W. and Olivier, M. (2009) *Leishmania* GP63 alters host signaling through cleavage-activated protein tyrosine phosphatases. *Sci. Signal.* **2**, ra58–ra58.

Goyer, A. (2010) Thiamine in plants: Aspects of its metabolism and functions. *Phytochemistry* **71**, 1615–1624.

Grant, B.R., Greenaway, W. and Whatley, F.R. (1988) Metabolic changes during development of *Phytophthora palmivora* examined by gas chromatography/mass spectrometry. *Microbiology* **134**, 1901–1911.

Grenville-Briggs, L.J., Anderson, V.L., Fugelstad, J., et al. (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., Bruce, C.R., Williams, A., Whisson, S.C., Birch, P.R.J. and Van West, P. (2005) Elevated amino acid biosynthesis in *Phytophthora infestans* during appressorium formation and potato infection. *Fungal Genet. Biol.* **42**, 244–256.

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

Griffiths, R.G., Dancer, J., O'Neill, E. and Harwood, J.L. (2003a) A mandelamide pesticide alters lipid metabolism in *Phytophthora infestans*. *New Phytol.* **158**, 345–353.

Griffiths, R.G., Dancer, J., O'Neill, E. and Harwood, J.L. (2003b) Effect of culture conditions on the lipid composition of *Phytophthora infestans*. *New Phytol.* **158**, 337–344.

Gschloessl, B., Guermeur, Y. and Cock, J.M. (2008) HECTAR: a method to predict subcellular targeting in heterokonts. *BMC Bioinformatics* **9**, 393–393.

Gu, C., Kim, G.B., Kim, W.J., Kim, H.U. and Lee, S.Y. (2019) Current status and applications of genome-scale metabolic models. *Genome Biol.* **20**, 121.

Guan, J.C., Hasnain, G., Garrett, T.J., Chase, C.D., Gregory, J., Hanson, A.D. and McCarty, D.R. (2014) Divisions of labor in the thiamin biosynthetic pathway among organs of maize. *Front. Plant Sci.* **5**, 370.

Gubbels, M.-J. (2006) A MORN-repeat protein is a dynamic component of the *Toxoplasma gondii* cell division apparatus. *J. Cell Sci.* **119**, 2236–2245.

Gurevich, A., Saveliev, V., Vyahhi, N. and Tesler, G. (2013) QUAST: quality assessment tool for genome assemblies. *Bioinforma. Oxf. Engl.* **29**, 1072–1075.

Haas, B.J., Kamoun, S., Zody, M.C., et al. (2009) Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature* **461**, 393–398.

Halim, V.A., Eschen-Lippold, L., Altmann, S., Birschwilks, M., Scheel, D. and Rosahl, S. (2007) Salicylic acid is important for basal defense of *Solanum tuberosum* against *Phytophthora infestans*. *Mol. Plant. Microbe Interact.* **20**, 1346–1352.

Hallé, M., Gomez, M.A., Stuible, M., Shimizu, H., McMaster, W.R., Olivier, M. and Tremblay, M.L. (2009) The *Leishmania* surface protease GP63 cleaves multiple intracellular proteins and actively participates in p38 mitogen-activated protein kinase inactivation. *J. Biol. Chem.* **284**, 6893–6908.

Hartman, H.B., Fell, D.A., Rossell, S., et al. (2014) Identification of potential drug targets in *Salmonella enterica* sv. *Typhimurium* using metabolic modelling and experimental validation. *Microbiol. U. K.* **160**, 1252–1266.

Hartmann, M., Herrlich, A. and Herrlich, P. (2013) Who decides when to cleave an ectodomain? *Trends Biochem. Sci.* **38**, 111–120.

Hartmann, M., Zeier, T., Bernsdorff, F., et al. (2018) Flavin monooxygenase-generated n-hydroxypipecolic acid is a critical element of plant systemic immunity. *Cell* **173**, 456–469.e16.

Haverkort, A.J., Boonekamp, P.M., Hutten, R., Jacobsen, E., Lotz, L.A.P., Kessel, G.J.T., Visser, R.G.F. and Van Der Vossen, E.A.G. (2008) Societal costs of late blight in potato and prospects of durable resistance through cisgenic modification,.

Hayashida, K., Bartlett, A.H., Chen, Y. and Park, P.W. (2010) Molecular and cellular mechanisms of ectodomain shedding. *Anat. Rec. Hoboken NJ 2007* **293**, 925–937.

Heirendt, L., Arreckx, S., Pfau, T., et al. (2017) Creation and analysis of biochemical constraint-based models: the COBRA Toolbox v3.0. *ArXiv Prepr.* **1710.04038**.

Higashiyama, S., Nanba, D., Nakayama, H., Inoue, H. and Fukuda, S. (2011) Ectodomain shedding and remnant peptide signalling of EGFRs and their ligands. *J. Biochem. (Tokyo)* **150**, 15–22.

Hiltunen, J.K., Schonauer, M.S., Autio, K.J., Mittelmeier, T.M., Kastaniotis, A.J. and Dieckmann, C.L. (2009) Mitochondrial fatty acid synthesis type II: More than just fatty acids. *J. Biol. Chem.* **284**, 9011–9015.

Hochberg, B. (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc.*

Hodgson, W.A. (1958) Growth of four races of *Phytophthora infestans* (mont.) de bary in synthetic media. *Can. J. Plant Sci.* **38**, 145–154.

Hohl, H.R. (1991) Nutrition. In *Phytophthora infestans*, the cause of late blight of potato. Advances in Plant Pathology. Vol. 7 (Ingram, D.S. and Williams, P.H., eds), pp. 53–83.

Horn, F., Heinekamp, T., Kniemeyer, O., Pollmächer, J., Valiante, V. and Brakhage, A.A. (2012) Systems Biology of Fungal infection. *Front. Microbiol.* **3**.

Huthmacher, C., Hoppe, A., Bulik, S. and Holzhütter, H.G. (2010) Antimalarial drug targets in *Plasmodium falciparum* predicted by stage-specific metabolic network analysis. *BMC Syst. Biol.* **4**, 120.

Hyduke, D.R., Lewis, N.E. and Palsson, B.O. (2013) Analysis of omics data with genome-scale models of metabolism. *Mol. Biosyst.* **9**, 167–174.

Ideker, T., Galitski, T. and Hood, L. (2001) A new approach to decoding life: systems biology. *Annu. Rev. Genomics Hum. Genet.* **2**, 343–372.

Jamshidi, N. and Raghunathan, A. (2015) Cell scale host-pathogen modeling: another branch in the evolution of constraint-based methods. *Front. Microbiol.* **6**.

Jander, G. and Joshi, V. (2009) Aspartate-derived amino acid biosynthesis in *Arabidopsis thaliana*. *Arab. Book* **7**, e0121.

Jang, C., Chen, L. and Rabinowitz, J.D. (2018) Metabolomics and Isotope Tracing. *Cell* **173**, 822–837.

Jashni, M.K., Dols, I.H.M., Iida, Y., Boeren, S., Beenen, H.G., Mehrabi, R., Collemare, J. and Wit, P.J.G.M. de (2015) Synergistic action of a metalloprotease and a serine protease from *Fusarium oxysporum* f. sp. *lycopersici* cleaves chitin-binding tomato chitinases, reduces their antifungal activity, and enhances fungal virulence. *Mol. Plant. Microbe Interact.* **28**, 996–1008.

Jia, Y., McAdams, S.A., Bryan, G.T., Hershey, H.P. and Valent, B. (2000) Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. *EMBO J.* **19**, 4004–4014.

Jia, Y., Zhou, E., Lee, S. and Bianco, T. (2016) Coevolutionary dynamics of rice blast resistance gene Pi-ta and *Magnaporthe oryzae* avirulence gene AVR-Pita 1. *Phytopathology* **106**, 676–683.

Jiang, R.H.Y., Bruijn, I. de, Haas, B.J., et al. (2013) Distinctive expansion of potential virulence genes in the genome of the oomycete fish pathogen *Saprolegnia parasitica*. *PLoS Genet.* **9**, e1003272–e1003272.

Jones, P., Binns, D., Chang, H.-Y., et al. (2014) InterProScan 5: genome-scale protein function classification. *Bioinforma. Oxf. Engl.* **30**, 1236–1240.

Joshi, P.B., Kelly, B.L., Kamhawi, S., Sacks, D.L. and McMaster, W.R. (2002) Targeted gene deletion in *Leishmania major* identifies leishmanolysin (GP63) as a virulence factor. *Mol. Biochem. Parasitol.* **120**, 33–40.

Judelson, H.S. (2017) Metabolic diversity and novelties in the oomycetes. *Annu. Rev. Microbiol.* **71**, 21–39.

Judelson, H.S. and Ah-Fong, A.M. (2018) Exchanges at the plant-oomycete interface that influence disease. *Plant Physiol.*, pp.00979.2018.

Judelson, H.S. and Blanco, F.A. (2005) The spores of *Phytophthora*: Weapons of the plant destroyer. *Nat. Rev. Microbiol.* **3**, 47–58.

Judelson, H.S., Narayan, R.D., Ah-Fong, A.M. and Kim, K.S. (2009a) Gene expression changes during asexual sporulation by the late blight agent *Phytophthora infestans* occur in discrete temporal stages. *Mol. Genet. Genomics* **281**, 193–206.

Judelson, H.S., Tani, S. and Narayan, R.D. (2009b) Metabolic adaptation of *Phytophthora infestans* during growth on leaves, tubers and artificial media. *Mol. Plant Pathol.* **10**, 843–855.

Käll, L., Krogh, A. and Sonnhammer, E.L.L. (2004) A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.* **338**, 1027–1036.

Kaltdorf, M., Srivastava, M., Gupta, S.K., et al. (2016) Systematic identification of anti-fungal drug targets by a metabolic network approach. *Front. Mol. Biosci.* **3**, 22.

Kamoun, S., Furzer, O., Jones, J.D.G., et al. (2015) The Top 10 oomycete pathogens in molecular plant pathology. *Mol. Plant Pathol.* **16**, 413–434.

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. and Tanabe, M. (2015) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **44**, gkv1070.

Karp, P.D., Paley, S., Krummenacker, M., et al. (2009) Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology. *Brief Bioinform* **11**, 40.

Kastaniotis, A.J., Autio, K.J., Kerätär, J.M., et al. (2017) Mitochondrial fatty acid synthesis, fatty acids and mitochondrial physiology. *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids* **1862**, 39–48.

Katoh, K. and Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780.

Kavvas, E.S., Seif, Y., Yurkovich, J.T., Norsigian, C., Poudel, S., Greenwald, W.W., Ghatak, S., Palsson, B.O. and Monk, J.M. (2018) Updated and standardized genome-scale reconstruction of *Mycobacterium tuberculosis* H37Rv, iEK1011, simulates flux states indicative of physiological conditions. *BMC Syst. Biol.* **12**.

Kay, J., Meijer, H.J.G., Have, A. ten and Kan, J.A.L. van (2011) The aspartic proteinase family of three *Phytophthora* species. *BMC Genomics* **12**, 254–254.

Kemen, E., Gardiner, A., Schultz-Larsen, T., et al. (2011) Gene gain and loss during evolution of obligate parasitism in the white rust pathogen of *Arabidopsis thaliana*. *PLoS Biol.* **9**, e1001094

Kemen, E. and Jones, J.D.G. (2012) Obligate biotroph parasitism: can we link genomes to lifestyles? *Trends Plant Sci.* **17**, 448–457.

Kim, D., Langmead, B. and Salzberg, S.L. (2015) HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360.

Kim, J. and Sanderson, M.J. (2008) Penalized likelihood phylogenetic inference: Bridging the parsimony-likelihood gap. *Syst. Biol.* **57**, 665–674.

King, Z.A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J.A., Ebrahim, A., Palsson, B.O. and Lewis, N.E. (2016) BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res.* **44**, D515–D522.

Kitano, H. (2002) Systems biology: a brief overview. *Science* **295**, 1662–1664.

Kleessen, S., Irgang, S., Klie, S., Giavalisco, P. and Nikoloski, Z. (2015) Integration of transcriptomics and metabolomics data specifies the metabolic response of *Chlamydomonas* to rapamycin treatment. *Plant J.* **81**, 822–835.

Klein, C.C., Cottret, L., Kielbassa, J., Charles, H., Gautier, C., Ribeiro de Vasconcelos, A.T., Lacroix, V. and Sagot, M.-F. (2012) Exploration of the core metabolism of symbiotic bacteria. *BMC Genomics* **13**, 438.

Kloehn, J., Blume, M., Cobbold, S., Saunders, E., Dagley, M. and McConville, M. (2016) Using metabolomics to dissect host–parasite interactions. *Curr. Opin. Microbiol.* **32**, 59–65.

Koduru, L., Kim, H.Y., Lakshmanan, M., Mohanty, B., Lee, Y.Q., Lee, C.H. and Lee, D.-Y. (2020) Genome-scale metabolic reconstruction and in silico analysis of the rice leaf blight pathogen, *Xanthomonas oryzae*. *Mol. Plant Pathol.*

Koren, S., Rhie, A., Walenz, B.P., et al. (2018) De novo assembly of haplotype-resolved genomes with trio binning. *Nat. Biotechnol.* **36**, 1174–1182.

Kots, K., Meijer, H.J.G., Bouwmeester, K., Govers, F. and Ketelaar, T. (2017) Filamentous actin accumulates during plant cell penetration and cell wall plug formation in *Phytophthora infestans*. *Cell. Mol. Life Sci.* **74**, 909–920.

Lamour, K.H., Mudge, J., Gobena, D., et al. (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.

Larhlimi, A., David, L., Selbig, J. and Bockmayr, A. (2012) F2C2: a fast tool for the computation of flux coupling in genome-scale metabolic networks. *BMC Bioinformatics* **13**, 57.

Lazar, M.A. and Birnbaum, M.J. (2012) De-Meaning of Metabolism. *Science* **336**, 1651–1652.

Lee, E., Helt, G.A., Reese, J.T., et al. (2013) Web Apollo: a web-based genomic annotation editing platform. *Genome Biol.* **14**, R93.

Leesutthiphonchai, W., Vu, A.L., Ah-Fong, A.M.V. and Judelson, H.S. (2018) How does *Phytophthora infestans* evade control efforts? Modern insight into the late blight disease. *Phytopathology* **108**, 916–924

Lelis, T., Peng, J., Barphagha, I., Chen, R. and Ham, J.H. (2019) The virulence function and regulation of the metalloprotease gene prtA in the plant-pathogenic bacterium *Burkholderia glumae*. *Mol. Plant. Microbe Interact.* **32**, 841–852.

Leonard, G., Labarre, A., Milner, D.S., et al. (2018) Comparative genomic analysis of the 'pseudofungus' *Hyphochytrium catenoides*. *Open Biol.* **8**.

Levering, J., Broddrick, J., Dupont, C.L., et al. (2016) Genome-scale model reveals metabolic basis of biomass partitioning in a model diatom. *PLoS ONE* **11**, e0155038.

Lévesque, C.A., Brouwer, H., Cano, L., et al. (2010) Genome sequence of the necrotrophic plant pathogen *Pythium ultimum* reveals original pathogenicity mechanisms and effector repertoire. *Genome Biol.* **11**, R73.

Levy, R., Carr, R., Kreimer, A., Freilich, S. and Borenstein, E. (2015) NetCooperate: a network-based tool for inferring host-microbe and microbe-microbe cooperation. *BMC Bioinformatics* **16**, 164.

Lewis, N.E., Nagarajan, H. and Palsson, B.O. (2012) Constraining the metabolic genotype–phenotype relationship using a phylogeny of in silico methods. *Nat. Rev. Microbiol.* **10**, 291–305.

Li, H., Handsaker, B., Wysoker, A., et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079.

Li, R., Tee, C.S., Jiang, Y.L., Jiang, X.Y., Venkatesh, P.N., Sarojam, R. and Ye, J. (2015) A terpenoid phytoalexin plays a role in basal defense of *Nicotiana benthamiana* against Potato Virus X. *Sci. Rep.* **5**.

Li, Y., Orlando, B.J. and Liao, M. (2019) Structural basis of lipopolysaccharide extraction by the LptB(2)FGC complex. *Nature* **567**, 486–490.

Liu, H., Ma, X., Yu, H., Fang, D., Li, Y., Wang, X., Wang, W., Dong, Y. and Xiao, B. (2016) Genomes and virulence difference between two physiological races of *Phytophthora nicotianae*. *GigaScience* **5**, 3.

Liu, T., Song, T., Zhang, X., et al. (2014) Unconventionally secreted effectors of two filamentous pathogens target plant salicylate biosynthesis. *Nat. Commun.* **5**, 4686.

Lommer, M., Specht, M., Roy, A.-S., et al. (2012) Genome and low-iron response of an oceanic diatom adapted to chronic iron limitation. *Genome Biol.* **13**, R66.

López-Calcagno, P.E., Moreno, J., Cedeño, L., Labrador, L., Concepción, J.L. and Avilán, L. (2009) Cloning, expression and biochemical characterization of mitochondrial and cytosolic malate dehydrogenase from *Phytophthora infestans*. *Mycol. Res.* **113**, 771–781.

Lowe-Power, T.M., Jacobs, J.M., Ailloud, F., Fochs, B., Prior, P. and Allen, C. (2016) Degradation of the plant defense signal salicylic acid protects *Ralstonia solanacearum* from toxicity and enhances virulence on tobacco. *mBio* **7**, e00656-16.

Machado, D., Herrgard, M.J. and Rocha, I. (2016) Stoichiometric representation of gene-protein-reaction associations leverages constraint-based analysis from reaction to gene-level phenotype prediction. *PLoS Comput. Biol.* **12**, e1005140.

Magnúsdóttir, S., Heinken, A., Kutt, L., et al. (2016) Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nat. Biotechnol.* **35**, 81–89.

Mao, X., Cai, T., Olyarchuk, J.G. and Wei, L. (2005) Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics* **21**, 3787–3793.

Marano, A.V., Jesus, A.L., Souza, J.I. de, Jerônimo, G.H., Gonçalves, D.R., Boro, M.C., Rocha, S.C.O. and Pires-Zottarelli, C.L.A. (2016) Ecological roles of saprotrophic *Peronosporales* (Oomycetes, Straminipila) in natural environments. *Fungal Ecol.* **19**, 77–88.

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

Matson, M.E.H., Small, I.M., Fry, W.E. and Judelson, H.S. (2015) Metalaxyl resistance in *Phytophthora infestans*: Assessing role of RPA190 gene and diversity within clonal lineages. *Phytopathology* **105**, 1594–1600.

Mayers, J.J., Flynn, K.J. and Shields, R.J. (2013) Rapid determination of bulk microalgal biochemical composition by Fourier-Transform Infrared spectroscopy. *Bioresour. Technol.* **148**, 215–220.

McCarthy, C.G.P. and Fitzpatrick, D.A. (2017) Phylogenomic reconstruction of the oomycete phylogeny derived from 37 genomes. *mSphere* **2**, e00095–17.

McDowell, J.M. (2011) Genomes of obligate plant pathogens reveal adaptations for obligate parasitism. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 8921–8922.

McGowan, J., Byrne, K.P. and 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. and Fitzpatrick, D.A. (2017) Genomic, network, and phylogenetic analysis of the oomycete effector arsenal. *mSphere* **2**, e00408-17.

McGowan, J. and Fitzpatrick, D.A. (2020) Recent advances in oomycete genomics. In Advances in Genetics., Academic Press.

McGwire, B.S., Chang, K.-P. and Engman, D.M. (2003) Migration through the extracellular matrix by the parasitic protozoan *Leishmania* is enhanced by surface metalloprotease gp63. *Infect. Immun.* **71**, 1008–1010.

McKnight, S.L. (2010) On Getting There from Here. *Science* **330**, 1338–1339.

McRose, D., Guo, J., Monier, A., et al. (2014) Alternatives to vitamin B1 uptake revealed with discovery of riboswitches in multiple marine eukaryotic lineages. *ISME J.* **8**, 2517–2529.

Meier, A., Erler, H. and Beitz, E. (2018) Targeting channels and transporters in protozoan parasite infections. *Front. Chem.* **6**, 88.

Meijer, H.J.G. and Govers, F. (2006) Genomewide analysis of phospholipid signaling genes in *Phytophthora* spp.: novelties and a missing link. *Mol. Plant-Microbe Interact. MPMI* **19**, 1337–1347.

Meijer, H.J.G., Hassen, H.H. and Govers, F. (2011) *Phytophthora infestans* has a plethora of phospholipase D enzymes including a subclass that has extracellular activity. *PloS One* **6**, e17767–e17767.

Meijer, H.J.G., Mancuso, F.M., Espadas, G., Seidl, M.F., Chiva, C., Govers, F. and Sabidó, E. (2014) Profiling the secretome and extracellular proteome of the potato late blight pathogen *Phytophthora infestans*. *Mol. Cell. Proteomics MCP* **13**, 2101–2113.

Meijer, H.J.G., Schoina, C., Wang, S., Bouwmeester, K., Hua, C. and Govers, F. (2019) *Phytophthora infestans* small phospholipase D-like proteins elicit plant cell death and promote virulence. *Mol. Plant Pathol.* **20**, 180–193.

Miedes, E., Vanholme, R., Boerjan, W. and Molina, A. (2014) The role of the secondary cell wall in plant resistance to pathogens. *Front. Plant Sci.* **5**, 358.

Misner, I., Blouin, N., Leonard, G., Richards, T.A. and Lane, C.E. (2014) 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.

Miyoshi, S. and Shinoda, S. (2000) Microbial metalloproteases and pathogenesis. *Microbes Infect.* **2**, 91–98.

Mock, T., Otillar, R.P., Strauss, J., et al. (2017) Evolutionary genomics of the cold-adapted diatom *Fragilariopsis cylindrus*. *Nature* **541**, 536–540.

Monk, J., Nogales, J. and Palsson, B.O. (2014) Optimizing genome-scale network reconstructions. *Nat. Biotechnol.* **32**, 447–452.

Monod, M., Capoccia, S., Léchenne, B., Zaugg, C., Holdom, M. and Jousson, O. (2002) Secreted proteases from pathogenic fungi. *Int. J. Med. Microbiol.* **292**, 405–419.

Morales-Cruz, A., Amrine, K.C.H., Blanco-Ulate, B., Lawrence, D.P., Travadon, R., Rolshausen, P.E., Baumgartner, K. and Cantu, D. (2015) Distinctive expansion of gene families associated with plant cell wall degradation, secondary metabolism, and nutrient uptake in the genomes of grapevine trunk pathogens. *BMC Genomics* **16**, 469.

Moretti, S., Martin, O., Van Du Tran, T., Bridge, A., Morgat, A. and Pagni, M. (2016) MetaNetX/MNXref – reconciliation of metabolites and biochemical reactions to bring together genome-scale metabolic networks. *Nucleic Acids Res.* **44**, D523–D526.

Mukherjee, S., Retwitzer, M.D., Barash, D. and Sengupta, S. (2018) Phylogenomic and comparative analysis of the distribution and regulatory patterns of TPP riboswitches in fungi. *Sci. Rep.* **8**, 5563.

Nagarajan, N. and Pop, M. (2013) Sequence assembly demystified. *Nat. Rev. Genet.* **14**, 157–167.

Nakahigashi, K., Toya, Y., Ishii, N., et al. (2009) Systematic phenome analysis of *Escherichia coli* multiple-knockout mutants reveals hidden reactions in central carbon metabolism. *Mol. Syst. Biol.* **5**, 306.

Negrel, L., Halter, D., Wiedemann-Merdinoglu, S., Rustenholz, C., Merdinoglu, D., Hugueney, P. and Baltenweck, R. (2018) Identification of lipid markers of *Plasmopara viticola* infection in grapevine using a non-targeted metabolomic approach. *Front. Plant Sci.* **9**, 360.

Nelson, D.L. and Cox, M.M. (2017) Lehninger Principles of Biochemistry, W. H. Freeman.

Nielsen, H. (2017) Predicting secretory proteins with signalP. In Methods in Molecular Biology., pp. 59–73. Humana Press, New York, NY.

Nielsen, J. (2017) Systems biology of metabolism. *Annu. Rev. Biochem.* **86**, 245–275.

Nijhout, H.F., Best, J.A. and Reed, M.C. (2019) Systems biology of robustness and homeostatic mechanisms. *WIREs Syst. Biol. Med.* **11**, e1440.

Nisar, N., Li, L., Lu, S., Khin, N.C. and Pogson, B.J. (2015) Carotenoid metabolism in plants. *Mol. Plant* **8**, 68–82.

Niu, X., Ah-Fong, A.M.V., Lopez, L.A. and Judelson, H.S. (2018) Transcriptomic and proteomic analysis reveals wall-associated and glucan-degrading proteins with potential roles in *Phytophthora infestans* sexual spore development. *PLoS ONE* **13**, e0198186.

O'Brien, E.J., Monk, J.M. and Palsson, B.O. (2015) Using genome-scale models to predict biological capabilities. *Cell* **161**, 971–987.

O'Hara, J.K., Kerwin, L.J., Cobbold, S.A., Tai, J., Bedell, T.A., Reider, P.J. and Llinás, M. (2014) Targeting NAD+ metabolism in the human malaria parasite *Plasmodium falciparum*. *PLoS ONE* **9**.

Olive, A.J. and Sassetti, C.M. (2016) Metabolic crosstalk between host and pathogen: Sensing, adapting and competing. *Nat. Rev. Microbiol.* **14**, 221–234.

Olszewski, K.L., Morrisey, J.M., Wilinski, D., Burns, J.M., Vaidya, A.B., Rabinowitz, J.D. and Llinás, M. (2009) Host-parasite interactions revealed by *Plasmodium falciparum* metabolomics. *Cell Host Microbe* **5**, 191–199.

Orbach, M.J., Farrall, L., Sweigard, J.A., Chumley, F.G. and Valent, B. (2000) a telomeric avirulence gene determines efficacy for the rice blast resistance gene Pi-ta. *Plant Cell* **12**, 2019.

Orth, J.D., Thiele, I. and Palsson, B.O. (2010) What is flux balance analysis? *Nat. Biotechnol.* **28**, 245–8.

Ospina-Giraldo, M.D., Griffith, J.G., Laird, E.W. and 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.

Paradis, E. and Schliep, K. (2019) ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R Schwartz, R., ed. *Bioinformatics* **35**, 526–528.

Pearson, W.R. (2013) An introduction to sequence similarity ("homology") searching. *Curr. Protoc. Bioinforma.* **Chapter 3**, Unit3.1.

Penaranda, C. and Hung, D.T. (2019) Single-cell rna sequencing to understand host–pathogen interactions. *ACS Infect. Dis.* **5**, 336–344.

Peng, D., Lin, J., Huang, Q., Zheng, W., Liu, G., Zheng, J., Zhu, L. and Sun, M. (2015) A novel metalloproteinase virulence factor is involved in *Bacillus thuringiensis* pathogenesis in nematodes and insects. *Environ. Microbiol.* **18**, 846–862.

Percudani, R. and Peracchi, A. (2003) A genomic overview of pyridoxal-phosphate-dependent enzymes. *EMBO Rep.* **4**, 850–4.

Pettongkhao, S., Navet, N., Schornack, S., Tian, M. and Churngchow, N. (2020) A secreted protein of 15 kDa plays an important role in *Phytophthora palmivora* development and pathogenicity. *Sci. Rep.* **10**, 2319.

Petutschnig, E.K., Stolze, M., Lipka, U., et al. (2014) A novel *Arabidopsis* CHITIN ELICITOR RECEPTOR KINASE 1 (CERK1) mutant with enhanced pathogen-induced cell death and altered receptor processing. *New Phytol.* **204**, 955–967.

Peyraud, R., Cottret, L., Marmiesse, L. and Genin, S. (2018) Control of primary metabolism by a virulence regulatory network promotes robustness in a plant pathogen. *Nat. Commun.* **9**, 418.

Peyraud, R., Cottret, L., Marmiesse, L., Gouzy, J. and Genin, S. (2016) A resource allocation trade-off between virulence and proliferation drives metabolic versatility in the plant pathogen *Ralstonia solanacearum*. *PLoS Pathog.* **12**, e1005939.

Peyraud, R., Dubiella, U., Barbacci, A., Genin, S., Raffaele, S. and Roby, D. (2017) Advances on plant-pathogen interactions from molecular toward systems biology perspectives. *Plant J.* **90**, 720–737.

Peyraud, R., Mbengue, M., Barbacci, A. and Raffaele, S. (2019) Intercellular cooperation in a fungal plant pathogen facilitates host colonization. *Proc. Natl. Acad. Sci. U. S. A.*, 201811267.

Pinzón, A., Barreto, E., Bernal, A., Achenie, L., González Barrios, A.F., Isea, R. and Restrepo, S. (2009) Computational models in plant-pathogen interactions: the case of *Phytophthora infestans*. *Theor. Biol. Med. Model.* **6**, 24.

Pinzón, A., Barreto, E., Bernal, A., Achenie, L., Gonzlez Barrios, A.F., Isea, R. and Restrepo, S. (2009) Computational models in plant-pathogen interactions: The case of *Phytophthora infestans*. *Theor. Biol. Med. Model.* **6**, 24.

Pinzón, A., Rodriguez-R, L.M., González, A., Bernal, A. and Restrepo, S. (2011) Targeted metabolic reconstruction: a novel approach for the characterization of plant–pathogen interactions. *Brief. Bioinform.* **12**, 151–162.

Placzek, S., Schomburg, I., Chang, A., Jeske, L., Ulbrich, M., Tillack, J. and Schomburg, D. (2017) BRENDA in 2017: New perspectives and new tools in BRENDA. *Nucleic Acids Res.* **45**, D380–D388.

Plata, G., Hsiao, T.L., Olszewski, K.L., Llinás, M. and Vitkup, D. (2010) Reconstruction and flux-balance analysis of the *Plasmodium falciparum* metabolic network. *Mol. Syst. Biol.* **6**, 408.

Poulin, R. (2007) Evolutionary ecology of parasites 2nd ed., Princeton University Press.

Poulin, R. and Randhawa, H.S. (2015) Evolution of parasitism along convergent lines: From ecology to genomics. *Parasitology* **142**, S6–S15.

Pratapa, A., Balachandran, S. and Raman, K. (2015) Fast-SL: An efficient algorithm to identify synthetic lethal sets in metabolic networks. *Bioinformatics* **31**, 3299–3305.

Price, M.N., Dehal, P.S. and Arkin, A.P. (2010) FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490.

Prigent, S., Collet, G., Dittami, S.M., et al. (2014) The genome-scale metabolic network of *Ectocarpus siliculosus* (EctoGEM): a resource to study brown algal physiology and beyond. *Plant J.* **80**, 367–381.

Pritchard, L. and Birch, P.R.J. (2011) A systems biology perspective on plant-microbe interactions: biochemical and structural targets of pathogen effectors. *Plant Sci.* **180**, 584–603.

Pryor, E.E., Horanyi, P.S., Clark, K.M., et al. (2013) Structure of the integral membrane protein CAAX protease Ste24p. *Science* **340**, 1600–1604.

Quinn, L., O'Neill, P.A., Harrison, J., Paskiewicz, K.H., McCracken, A.R., Cooke, L.R., Grant, M.R. and Studholme, D.J. (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.

Radakovits, R., Jinkerson, R.E., Fuerstenberg, S.I., Tae, H., Settlage, R.E., Boore, J.L. and Posewitz, M.C. (2012) Draft genome sequence and genetic transformation of the oleaginous alga *Nannochloropsis gaditana*. *Nat. Commun.* **3**, 1–11.

Raffaele, S. and Kamoun, S. (2012) Genome evolution in filamentous plant pathogens: why bigger can be better. *Nat. Rev. Microbiol.* **10**, 417–430.

Raffaele, S., Win, J., Cano, L.M. and 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.

Raghunathan, A., Reed, J., Shin, S., Palsson, B.O. and Daefler, S. (2009) Constraint-based analysis of metabolic capacity of *Salmonella typhimurium* during host-pathogen interaction. *BMC Syst. Biol.* **3**, 38.

Raghunathan, A., Shin, S. and Daefler, S. (2010) Systems approach to investigating host-pathogen interactions in infections with the biothreat agent *Francisella*. Constraints-based model of *Francisella tularensis*. *BMC Syst. Biol.* **4**.

Rawlings, N.D. and Barrett, A.J. (1995) Evolutionary families of metallopeptidases. *Proteolytic Enzym. Aspartic Met. Pept.*, 183–228.

Rawlings, N.D., Barrett, A.J. and Finn, R. (2016) Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* **44**, D343–D350.

Rawlings, N.D., Barrett, A.J., Thomas, P.D., Huang, X., Bateman, A. and Finn, R.D. (2018) The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res.* **46**, D624–D632.

Rawlings, N.D. and Morton, F.R. (2008) The MEROPS batch BLAST: A tool to detect peptidases and their non-peptidase homologues in a genome. *Biochimie* **90**, 243–259.

Rawlings, N.D. and Salvesen, G. (2013) Handbook of Proteolytic Enzymes,.

Reed, J.L., Famili, I., Thiele, I. and Palsson, B.O. (2006) Towards multidimensional genome annotation. *Nat. Rev. Genet.* **7**, 130–141.

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

Richards, T.A. and Talbot, N.J. (2013) Horizontal gene transfer in osmotrophs: Playing with public goods. *Nat. Rev. Microbiol.* **11**, 720–727.

Rienksma, R.A., Schaap, P.J., Martins dos Santos, V.A.P. and Suarez-Diez, M. (2018) Modeling the metabolic state of *Mycobacterium tuberculosis* upon infection. *Front. Cell. Infect. Microbiol.* **8**.

Robbins, W.J. (1938) Thiamin and growth of species of *Phytophthora*. *Bull. Torrey Bot. Club* **65**, 267.

Robinson, M.D. and Oshlack, A. (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25.

Robinson, S.M. and Bostock, R.M. (2015) β-glucans and eicosapolyenoic acids as MAMPs in plant-oomycete interactions: past and present. *Front. Plant Sci.* **5**, 1–6.

Rödelsperger, C., Athanasouli, M., Lenuzzi, M., et al. (2019) Crowdsourcing and the feasibility of manual gene annotation: A pilot study in the nematode *Pristionchus pacificus*. *Sci. Rep.* **9**, 1–9.

Rodenburg, S.Y.A., Seidl, M.F., Judelson, H.S., Vu, A.L., Govers, F. and Ridder, D. de (2019) Metabolic model of the *Phytophthora infestans*-tomato interaction reveals metabolic switches during host colonization. *mBio* **10**, e00454–19.

Rodenburg, S.Y.A., Seidl, M.F., Ridder, D. de and Govers, F. (2018a) Genome-wide characterization of *Phytophthora infestans* metabolism: a systems biology approach. *Mol. Plant Pathol.* **19**, 1403–1413.

Rodenburg, S.Y.A., Terhem, R.B., Veloso, J., Stassen, J.H.M. and Kan, J.A.L. van (2018b) Functional analysis of mating type genes and transcriptome analysis during fruiting body development of *Botrytis cinerea*. *mBio* **9**.

Rohlenova, K., Goveia, J., García-Caballero, M., et al. (2020) Single-cell rna sequencing maps endothelial metabolic plasticity in pathological angiogenesis. *Cell Metab.* **31**, 862-877.e14.

Rolfsson, O., Palsson, B.O. and Thiele, I. (2011) The human metabolic reconstruction Recon 1 directs hypotheses of novel human metabolic functions. *BMC Syst. Biol.* **5**, 155.

Rost, B. (1999) Twilight zone of protein sequence alignments. *Protein Eng. Des. Sel.* **12**, 85–94.

Rovenich, H., Boshoven, J.C. and Thomma, B.P.H.J. (2014) Filamentous pathogen effector functions: Of pathogens, hosts and microbiomes. *Curr. Opin. Plant Biol.* **20**, 96–103.

Rujirawat, T., Patumcharoenpol, P., Lohnoo, T., et al. (2015) Draft genome sequence of the pathogenic oomycete *Pythium insidiosum* strain Pi-S, isolated from a patient with pythiosis. *Genome Announc* **3**.

Sahoo, S., Aurich, M.K., Jonsson, J.J. and Thiele, I. (2014) Membrane transporters in a human genome-scale metabolic knowledgebase and their implications for disease. *Front. Physiol.* **5**.

Saier, M.H., Reddy, V.S., Tsu, B.V., Ahmed, M.S., Li, C. and Moreno-Hagelsieb, G. (2016) The Transporter Classification Database (TCDB): Recent advances. *Nucleic Acids Res.* **44**, D372–D379.

Salzberg, S.L. (2019) Next-generation genome annotation: we still struggle to get it right. *Genome Biol.* **20**, 92.

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

Sanderson, M.P., Dempsey, P.J. and Dunbar, A.J. (2006) Control of ErbB signaling through metalloprotease mediated ectodomain shedding of EGF-like factors. *Growth Factors* **24**, 121–136.

Sanz-Martín, J.M., Pacheco-Arjona, J.R., Bello-Rico, V., Vargas, W.A., Monod, M., Díaz-Mínguez, J.M., Thon, M.R. and Sukno, S.A. (2016) A highly conserved metalloprotease effector enhances virulence in the maize anthracnose fungus *Colletotrichum graminicola*. *Mol. Plant Pathol.* **17**, 1048–1062.

Sato, S., Tabata, S., Hirakawa, H., et al. (2012) The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* **485**, 635–641.

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

Savchenko, T., Walley, J.W., Chehab, E.W., et al. (2010) Arachidonic acid: An evolutionarily conserved signaling molecule modulates plant stress signaling networks. *Plant Cell* **22**, 3193–3205.

Savory, F.R., Leonard, G. and 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. and Richards, T.A. (2018) Ancestral function and diversification of a horizontally acquired oomycete carboxylic acid transporter. *Mol. Biol. Evol.* **35**, 1887–1900.

Schellenberger, J., Que, R., Fleming, R.M.T., et al. (2011) Quantitative prediction of cellular metabolism with constraint-based models: The COBRA Toolbox v2.0. *Nat. Protoc.* **6**, 1290–1307.

Schoina, C., Verbeek-de Kruif, N., Govers, F. and Bouwmeester, K. (2019) Clade 5 aspartic proteases of *Phytophthora infestans* are virulence factors implied in RXLR effector cleavage. *Eur. J. Plant Pathol.* **154**, 17–29.

Seidl, M.F., Schneider, A., Govers, F. and Snel, B. (2013) A predicted functional gene network for the plant pathogen *Phytophthora infestans* as a framework for genomic biology. *BMC Genomics* **14**, 483.

Seidl, M.F., Van den Ackerveken, G., Govers, F. and Snel, B. (2011) A domain-centric analysis of oomycete plant pathogen genomes reveals unique protein organization. *Plant Physiol.* **155**, 628–644.

Seidl, M.F., Van den Ackerveken, G., Govers, F. and 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.

Seppey, M., Manni, M. and Zdobnov, E.M. (2019) BUSCO: Assessing genome assembly and annotation completeness. *Methods Mol Biol.* **1962**, 227–245.

Shanab, S.M.M., Hafez, R.M. and Fouad, A.S. (2018) A review on algae and plants as potential source of arachidonic acid. *J. Adv. Res.* **11**, 3–13.

Sharma, M., Shaikh, N., Yadav, S., Singh, S. and Garg, P. (2017) A systematic reconstruction and constraint-based analysis of *Leishmania donovani* metabolic network: identification of potential antileishmanial drug targets. *Mol. Biosyst.* **13**, 955–969.

Sharma, R., Xia, X., Cano, L.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.

Shende, R., Wong, S.S.W., Rapole, S., et al. (2018) *Aspergillus fumigatus* conidial metalloprotease Mep1p cleaves host complement proteins. *J. Biol. Chem.* **293**, 15538–15555.

Shlomi, T., Cabili, M.N., Herrgard, M.J., Palsson, B.O. and Ruppin, E. (2008) Network-based prediction of human tissue-specific metabolism. *Nat. Biotechnol.* **26**, 1003–1010.

Slater, G.S.C. and Birney, E. (2005) Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**, 31.

Spanu, P.D. (2012) The genomics of obligate (and nonobligate) biotrophs. *Annu. Rev. Phytopathol.* **50**, 91–109.

Stanway, R.R., Bushell, E., Chiappino-Pepe, A., et al. (2019) Genome-scale identification of essential metabolic processes for targeting the *Plasmodium* liver stage. *Cell* **179**, 1112-1128.e26.

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

Subramanian, A., Jhawar, J. and Sarkar, R.R. (2015) Dissecting *Leishmania infantum* energy metabolism - a systems perspective. *PLOS ONE* **10**, e0137976.

Sun, Q., Liu, J., Zhang, Q., Qing, X., Dobson, G., Li, X. and Qi, B. (2013) Characterization of three novel desaturases involved in the delta-6 desaturation pathways for polyunsaturated fatty acid biosynthesis from *Phytophthora infestans*. *Appl. Microbiol. Biotechnol.* **97**, 7689–7697.

Tanaka, T., Maeda, Y., Veluchamy, A., et al. (2015) Oil accumulation by the oleaginous diatom *Fistulifera solaris* as revealed by the genome and transcriptome. *Plant Cell* **27**, 162–176.

Tekir, S.D.D. and Ülgen**, K.Ö.** (2013) Systems biology of pathogen-host interaction: Networks of protein-protein interaction within pathogens and pathogen-human interactions in the post-genomic era. *Biotechnol. J.* **8**, 85–96.

Thiele, I. and Palsson, B.O. (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.* **5**, 93–121.

Thiele, I., Swainston, N., Fleming, R.M.T., et al. (2013) A community-driven global reconstruction of human metabolism. *Nat. Biotechnol.* **31**, 419–425.

Thiele, I., Vlassis, N. and Fleming, R.M.T. (2014) fastGapFill: efficient gap filling in metabolic networks. *Bioinformatics* **30**, 2529–2531.

Thines, M. and Choi, Y.J. (2015) Evolution, diversity, and taxonomy of the *Peronosporaceae*, with focus on the genus *Peronospora*. *Phytopathology* **106**, 6–18.

Thines, M., Sharma, R., Rodenburg, S.Y.A., et al. (2020) The genome of *Peronospora belbahrii* reveals high heterozygosity, a low number of canonical effectors, and tc-rich promoters. *Mol. Plant-Microbe Interactions* **33**, 742–753.

Thomma, B.P.H.J., Seidl, M.F., Shi-Kunne, X., Cook, D.E., Bolton, M.D., Kan, J.A.L. van and Faino, L. (2016) Mind the gap; seven reasons to close fragmented genome assemblies. *Fungal Genet. Biol.* **90**, 24–30.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M.J., Salzberg, S.L., Wold, B.J. and Pachter, L. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515.

Tyler, B.M., Tripathy, S., Zhang, X., et al. (2006) *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science* **313**, 1261–1266.

Tymoshenko, S., Oppenheim, R.D., Agren, R., Nielsen, J., Soldati-Favre, D. and Hatzimanikatis, V. (2015) Metabolic needs and capabilities of *Toxoplasma gondii* through combined computational and experimental analysis. *PLOS Comput. Biol.* **11**, e1004261.

Uzuhashi, S., Endoh, R., Manabe, R. and Ohkuma, M. (2017) Draft genome sequences of the oomycete *Pilasporangium apinafurcum* strains JCM 30513 and JCM 30514, formerly classified as *Pythium apinafurcum*. *Genome Announc.* **5**.

Vaattovaara, A., Leppälä, J., Salojärvi, J. and Wrzaczek, M. (2019) High-throughput sequencing data and the impact of plant gene annotation quality. *J. Exp. Bot.* **70**, 1069–1076.

van den Hoogen, D.J. Meijer, H.J.G., Seidl, M.F. and Govers, F. (2018) The ancient link between G-protein-coupled receptors and C-terminal phospholipid kinase domains. *mBio* **9**, e02119-17.

van den Hoogen, D.J. and Govers, F. (2018) GPCR-bigrams: Enigmatic signaling components in oomycetes. *PLoS Pathog.* **14**, e1007064–e1007064.

van West, P., De Jong, A.J., Judelson, H.S., Emons, A.M.C. and Govers, F. (1998) The ipiO gene of *Phytophthora infestans* is highly expressed in invading hyphae during infection. *Fungal Genet. Biol.* **23**, 126–138.

Wanders, R.J.A., Waterham, H.R. and Ferdinandusse, S. (2016) Metabolic interplay between peroxisomes and other subcellular organelles including mitochondria and the endoplasmic reticulum. *Front. Cell Dev. Biol.* **3**, 83.

Wang, D., Ning, K., Li, J., et al. (2014) *Nannochloropsis* genomes reveal evolution of microalgal oleaginous Traits. *PLOS Genet.* **10**, e1004094.

Wang, H., Marcišauskas, S., Sánchez, B.J., Domenzain, I., Hermansson, D., Agren, R., Nielsen, J. and Kerkhoven, E.J. (2018a) RAVEN 2.0: A versatile toolbox for metabolic network reconstruction and a case study on *Streptomyces coelicolor PLoS Comput. Biol.* **14**, e1006541.

Wang, L., Chen, H., Li, J., Shu, H., Zhang, X., Wang, Y., Tyler, B.M. and Dong, S. (2020) Effector gene silencing mediated by histone methylation underpins host adaptation in an oomycete plant pathogen. *Nucleic Acids Res.* **48**, 1790–1799.

Wang, R., Zhang, M., Liu, H., Xu, J., Yu, J., He, F., Zhang, X., Dong, S. and Dou, D. (2018) PsAAT3, an oomycete-specific aspartate aminotransferase, is required for full pathogenicity of the oomycete pathogen *Phytophthora sojae*. *Fungal Biol.* **122**, 1228–1229.

Wang, S., Boevink, P.C., Welsh, L., Zhang, R., Whisson, S.C. and 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., Welsh, L., Thorpe, P., Whisson, S.C., Boevink, P.C. and Birch, P.R.J. (2018b) The *Phytophthora infestans* haustorium is a site for secretion of diverse classes of infection-associated proteins. *mBio* **9**, e01216–18.

Wang, W., Xue, Z., Miao, J., Cai, M., Zhang, C., Li, T., Zhang, B., Tyler, B.M. and Liu, X. (2019) PcMuORP1, an oxathiapiprolin-resistance gene, functions as a novel selection marker for *Phytophthora* transformation and CRISPR/Cas9 Mediated genome editing. *Front. Microbiol.* **10**.

Wang, Y., Tyler, B.M. and Wang, Y. (2019) Defense and counterdefense during plant-pathogenic oomycete infection. *Annu. Rev. Microbiol.* **73**, 667–696.

Warrilow, A.G.S., Hull, C.M., Rolley, N.J., Parker, J.E., David Nes, W., Smith, S.N., Kelly, D.E. and Kelly, S.L. (2014) Clotrimazole as a potent agent for treating the oomycete fish pathogen *Saprolegnia parasitica* through inhibition of sterol 14α-demethylase (CYP51). *Appl. Environ. Microbiol.* **80**, 6154–6166.

Watson, E., Yilmaz, L.S. and Walhout, A.J.M. (2015) Understanding metabolic regulation at a systems level: metabolite sensing, mathematical predictions, and model organisms. *Annu. Rev. Genet.* **49**, 553–575.

Wegner, A., Meiser, J., Weindl, D. and Hiller, K. (2015) How metabolites modulate metabolic flux. *Curr. Opin. Biotechnol.* **34**, 16–22.

Weßling, R., Epple, P., Altmann, S., et al. (2014) Convergent targeting of a common host protein-network by pathogen effectors from three kingdoms of life. *Cell Host Microbe* **16**, 364–375.

West, P. van and Beakes, G.W. (2014) Animal pathogenic oomycetes. *Fungal Biol.* **118**, 525–526.

Whisson, S.C., Boevink, P.C., Wang, S. and Birch, P.R. (2016) The cell biology of late blight disease. *Curr. Opin. Microbiol.* **34**, 127–135.

Winterbach, W., Mieghem, P.V., Reinders, M., Wang, H. and Ridder, D. de (2013) Topology of molecular interaction networks. *BMC Syst. Biol.* **7**, 90.

Wood, V., Lock, A., Harris, M.A., Rutherford, K., Bähler, J. and Oliver, S.G. (2019) Hidden in plain sight: what remains to be discovered in the eukaryotic proteome? *Open Biol.* **9**, 180241.

Yandell, M. and Ence, D. (2012) A beginner's guide to eukaryotic genome annotation. *Nat. Rev. Genet.* **13**, 329–342.

Yilmaz, J.L., Lim, Z.L., Beganovic, M., Breazeale, S., Andre, C., Stymne, S., Vrinten, P. and Senger, T. (2017) Determination of substrate preferences for desaturases and elongases for production of docosahexaenoic acid from oleic acid in engineered canola. *Lipids* **52**, 207–222.

Yilmaz, L.S. and Walhout, A.J.M. (2016) A *Caenorhabditis elegans* genome-scale metabolic network model. *Cell Syst.* **2**, 297–311.

Yilmaz, L.S. and Walhout, A.J.M. (2017) Metabolic network modeling with model organisms. *Curr. Opin. Chem. Biol.* **36**, 32–39.

Yizhak, K., Gabay, O., Cohen, H. and Ruppin, E. (2013) Model-based identification of drug targets that revert disrupted metabolism and its application to ageing. *Nat. Commun.* **4**, 513–519.

Yousef, L.F., Wojno, M., Dick, W.A. and Dick, R.P. (2012) Lipid profiling of the soybean pathogen *Phytophthora sojae* using Fatty Acid Methyl Esters (FAMEs). *Fungal Biol.* **116**, 613–619.

Yuan, H., Cheung, C.Y.M., Poolman, M.G., Hilbers, P.A.J. and Van Riel, N.A.W. (2016) A genome-scale metabolic network reconstruction of tomato (*Solanum lycopersicum* L.) and its application to photorespiratory metabolism. *Plant J.* **85**, 289–304.

Yurkovich, J.T. and Palsson, B.O. (2016) Solving puzzles with missing pieces: the power of systems biology. *Proc. IEEE* **104**, 2–7.

Zhang, C., Bidkhori, G., Benfeitas, R., Lee, S., Arif, M., Uhlén, M. and Mardinoglu, A. (2018) ESS: A tool for genome-scale quantification of essentiality score for reaction/genes in constraint-based modeling. *Front. Physiol.* **9**, 1355.

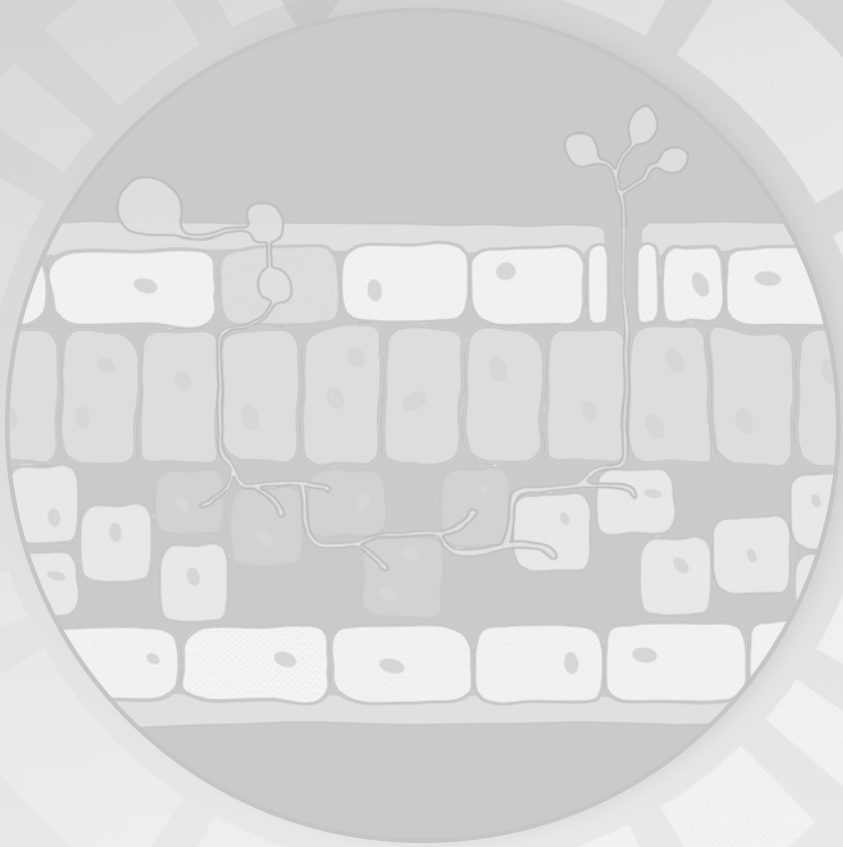
Zhang, C. and Hua, Q. (2015) Applications of genome-scale metabolic models in biotechnology and systems medicine. *Front. Physiol.* **6**, 413.

Zhang, Y., Bak, D.D., Heid, H. and Geider, K. (1999) Molecular characterization of a protease secreted by *Erwinia amylovora*. *Mol. Biol.* **289**, 1239–1251.

Zimorski, V., Rauch, C., Hellemond, J.J. van, Tielens, A.G.M. and Martin, W.F. (2017) The mitochondrion of *Euglena gracilis*. In Advances in Experimental Medicine and Biology. (Schwartzbach, S.D. and Shigeoka, S., eds), pp. 19–37. Springer International Publishing.

Zoledowska, S., Presta, L., Fondi, M., Decorosi, F., Giovannetti, L., Mengoni, A. and Lojkowska, E. (2019) Metabolic modeling of *Pectobacterium parmentieri* SCC3193 provides insights into metabolic pathways of plant pathogenic bacteria. *Microorganisms* **7**.

Zuluaga, A.P., Vega-Arreguin, J.C., Fei, Z., Matas, A.J., Patev, S., Fry, W.E. and Rose, J.K.C. (2016) Analysis of the tomato leaf transcriptome during successive hemibiotrophic stages of a compatible interaction with the oomycete pathogen *Phytophthora infestans*. *Mol. Plant Pathol.* **17**, 42–54.



Summary

Metabolism is the set of biochemical reactions of an organism that enables it to assimilate nutrients from its environment, and to generate building blocks for growth and proliferation. It is a complex network that is intertwined with the many molecular and cellular processes in cells, the fundamental units of life. Systems biology is a rapidly developing discipline that aims at capturing the complexity of cells, organisms or communities by reconstructing models based on information gathered by high-throughput analyses (omics data) and from literature (**Chapter 1**). One type of model is a genome-scale metabolic model (GEM) that simulates metabolism as a framework of metabolic fluxes, i.e. the "mass-flow" through a network of biochemical reactions. GEMs are nowadays widely applied and have been reconstructed for various microbial pathogens, with the aim to get insight into mechanisms of pathogenicity.

Oomycetes are filamentous organisms that belong to the stramenopile lineage in the SAR supergroup. They include many devastating plant and animal pathogens that are a continuous threat to agriculture, aquaculture, and natural ecosystems. Comparative genomics has revealed several intruiging characteristics, such as dynamic genomes with large differences in gene repertoires and proteins with novel and unique domain combinations. We identified and compared the metabolic enzymes of a broad range of oomycetes that differ in lifestyle and host preference, and investigated their metabolic networks from an evolutionary perspective (**Chapter 2**). We observed lineage-specific pathway loss, and convergent loss of metabolic enzymes in obligate biotrophs, pathogens that are entirely dependent on their host for survival. This biotrophic lifestyle has led to extreme gene loss also in metabolic genes and consequently, obligate biotrophs have a reduced metabolic capacity and greater host dependency. The network-based analysis that we performed showed that those gene losses predominantly affected the periphery of the metabolic network, something that could not be shown when solely comparing genomes.

Metabolism is dynamic with a continuous flow of nutrient uptake, processing and biosynthesis. To gain insight into this dynamics, we constructed a GEM of the oomycete *Phytophthora infestans,* the causal agent of potato and tomato late-blight (**Chapter 3**). We extracted information on its metabolism from literature, and used homology-based enzyme annotation to arrive at a draft metabolic model, divided reactions over several subcellular compartments. We inspected the model topology to gain insight into the biochemical processes in each compartment. Integration of transcriptome data of different *P. infestans* life stages revealed a sharp contrast in metabolic activity between sporangia and hyphae. This GEM allows us to simulate growth (biomass production) on a minimal defined growth medium and to predict the essential metabolic genes and reactions. As such, the model has the potential to contribute to the discovery of novel targets for disease control.

Conceivably, the metabolism of *P. infestans* growing in a minimal growth medium differs from that during *in planta* growth. To investigate how the metabolism of the pathogen and the host

are intertwined we integrated our *P. infestans* GEM with a GEM of tomato, that we adopted from literature (**Chapter 4**). Using modelling techniques such as flux coupling analysis we identified several metabolic reactions in tomato that were of importance to the fluxes in *P. infestans* metabolism, including thiamine biosynthesis to supply the thiamine auxotrophic pathogen with this essential vitamin. Integration of dual-transcriptome data of a time course of a full infection cycle revealed various switches in metabolism and differential nutrient usage over time. The integrated model predicts that during late infection, *P. infestans* switches to the import of amino acids rather than inorganic nutrients. However, this goes at the cost of system robustness, i.e. the ability to function under perturbations.

Next to metabolic enzymes, an organism exploits many other classes of enzymes for maintenance of the basic cellular and molecular processes. Pathogens often utilize enzymes to break physical barriers for gaining entry to their hosts or for modifying the host cell machinery to suppress defense. One class of enzymes that has been implicated in pathogenicity comprises metalloproteases (MPs). To gain insight into the MP repertoire of oomycete pathogens we screened the proteomes of 47 Stramenopile species (**Chapter 5**) and found a miscellaneous set of MPs, with some MP families being expanded in oomycetes with a particular lifestyle. We then investigated the expression profiles of *P. infestans* MPs, and selected those with a predicted secretion signal for a role in virulence in *in planta* assays. Three MPs had a negative effect on virulence, while eight had a positive effect, suggesting that MPs play a role in host-pathogen interactions.

The interactions that oomycetes have with their host are part of complex molecular systems. Models can be used to study these systems. which are based on the available genomic information and integrated with experimental data and prior knowledge (**Chapter 6**). The current state of oomycete genomes and annotations begs for improvement. However, despite these limitations we were able to build a model for *P. infestans*-tomato metabolism, which demonstrated that models of plant pathogens are feasible and can be used to study their mechanisms of host interaction. Our model should be further curated when more knowledge is obtained from experimental work. Therefore, we propose a framework of *in silico* and *in vitro* steps that would yield useful data for future model improvements.

Samenvatting

Metabolisme is de collectie van biochemische reacties in een organisme dat het gebruikt om voedingsstoffen op te nemen uit zijn omgeving, en deze om te zetten naar bouwstenen voor groei en voortplanting. Het is een complex netwerk dat samenhangt met vele moleculaire en cellulaire systemen in de cel, de fundamentele elementen van het leven. Systeembiologie is een wetenschappelijk veld waarin geprobeerd wordt om door middel van modellen de complexiteit van de cel, een organisme of een ecosysteem na te bootsen. Hiervoor worden grote biologische datasets gebruikt in combinatie met kennis uit de literatuur (**Hoofdstuk 1**). Een specifiek type model is het metabolische model (GEM), dat het metabolisme van de cel nabootst en dit beschouwt als een systeem van fluxen, de massa-doorvoer in het netwerk van biochemische reacties. GEMs worden tegenwoordig veel gebruikt, en zijn gebouwd voor verschillende microben, waaronder pathogenen, om te begrijpen hoe deze ziekte veroorzaken in plant of dier.

Oömyceten zijn draadvormige organismes die toebehoren aan het rijk van de Stramenopila. Hieronder vallen vele pathogenen voor plant en dier, en vormen een belangrijk risico voor de agricultuur, aquacultuur en ecosystemen. Door het vergelijken van de genomen zijn er verschillende interessante factoren ontdekt, zoals dat deze genomen erg dynamisch zijn en verschillen in grootte, en voor eiwitten coderen die uit unieke domein combinaties bestaan. We hebben we de metabole enzymen geïdentificeerd en vergeleken in een brede selectie van oömyceten, die verschillen in levensstijl en gastheer, om zo inzicht te krijgen in hun metabole systemen in evolutionaire context (**Hoofdstuk 2**). We hebben zo geobserveerd dat er in bepaalde evolutionaire takken specifieke metabole routes verloren zijn gegaan. Obligate biotrofe oömyceten, volledig afhankelijk van hun gastheer, hebben ondanks hun evolutionaire afstand dezelfde enzymen verloren. Dit patroon was ook zichtbaar in metabole enzymen, waardoor obligate biotrofen een verminderde metabole capaciteit hebben en een grotere afhankelijkheid van hun gastheer. We hebben toen deze enzymen beschouwd in context van een metabool netwerk, en observeerden dat deze verloren enzymen voornamelijk aan de rand van het netwerk lagen, iets wat niet had kunnen worden gezien op basis van de genomen alleen.

Metabolisme is dynamisch, en er wordt een constante stroom van nutriënten omgezet in bouwstenen. Om inzicht te krijgen in deze dynamiek, hebben we een GEM gereconstrueerd voor de oömyceet *Phytophthora infestans,* de veroorzaker van tomaat- en aardappelrot (**Hoofdstuk 3**). Op basis van de literatuur en door middel van homologie tussen enzymen hebben we een model gebouwd, onderverdeeld in verschillende cellulaire compartimenten. Door de topologie van het model te bestuderen hebben we inzicht gekregen in de biochemische processen in elk van deze compartimenten. De integratie van transcriptoom data van de verschillende levensstadia van *P. infestans* legde een scherp contrast bloot in metabole activiteit tussen sporangia en hyfen. Dit GEM hielp ons verder om groei te simuleren

(de productie van biomassa) op een minimaal groeimedium, en om de essentiële genen en reacties te voorspellen. Daarmee heeft dit model veel potentie in de zoektocht naar nieuwe strategieën voor bestrijdingsmiddelen.

De groei van *P. infestans* op een minimaal groeimedium is natuurlijk anders dan in de plant. Om te onderzoeken hoe het metabolisme van de pathogeen en zijn gastheer met elkaar zijn verweven, hebben we ons *P. infestans* GEM geïntegreerd met een GEM van tomaat, dat we uit de literatuur hebben overgenomen (**Hoofdstuk 4**). Door middel van methodes zoals flux-koppeling analyse hebben we verschillende reacties geïdentificeerd in tomaat, die van belang zijn voor het metabolisme van *P. infestans*. Bijvoorbeeld de thiamine biosynthese, om de auxotrofe pathogeen te voorzien van dit vitamine. Door het integreren van duaal-transcriptoom data van een infectie cyclus ontdekten we dat er verschillende schakelpunten zijn in het metabolisme waarin andere nutriënten worden gebruikt. Het model voorspelde dat laat in de infectiecyclus aminozuren worden geïmporteerd in plaats van anorganische stoffen. Dit ging ten koste van de robuustheid van het systeem, de capaciteit om normaal te functioneren tijdens perturbaties.

Naast metabole enzymen, heeft een organisme ook vele andere enzymen voor onderhoud of regulatie van cellulaire en moleculaire processen. Pathogenen gebruiken vaak enzymen om fysieke barrières af te breken, om toegang te krijgen tot de gastheer of om de zijn afweersystemen te onderdrukken. Een specifieke groep van enzymen die waarschijnlijk een rol spelen in pathogeniciteit, zijn de metalloproteases (MPs). Om inzicht te krijgen in het repertoire van MPs in oömyceten, hebben we de proteomen gescreend van 47 Stramenopila soorten (**Hoofdstuk 5**), waarbij we een veelzijdige set van MPs vonden. Sommige MP families kwamen vaker voor in oömyceten van een specifieke levensstijl. We onderzochten ook de expressiepatronen van *P. infestans* MPs, en keken specifiek naar MPs met een voorspeld secretie signaal voor een mogelijke rol in virulentie, in een studie in planten. Drie MPs hadden een negatief effect op virulentie, en acht een positief effect, wat suggereert dat MPs inderdaad een rol spelen in de interactie tussen pathogeen en gastheer.

De interacties die oömyceten hebben met hun gastheer zijn onderdeel van een complex moleculair systeem. Modellen kunnen gebruikt worden om deze systemen te bestuderen, gebaseerd op de beschikbare informatie uit genomen en experimentele data uit literatuur (**Hoofdstuk 6**). De huidige staat van de genoomsequenties van oömyceten genomen vraagt om verbetering. Ondanks deze belemmering waren we in staat een model te maken van *P. infestans*-tomaat metabolisme, wat demonstreerde dat modellen van plantpathogenen haalbaar zijn en gebruikt kunnen worden om de interactie tussen pathogenen en hun gastheer te bestuderen. Ons model moet verder worden gecureerd wanneer er meer experimentele data beschikbaar is. Daarom stellen we een plan voor met computationele en experimentele stappen die nuttige data zullen produceren, en kunnen worden gebruikt om het model te verbeteren.

Acknowledgements

Finally the moment has arrived that I can write this part of my thesis. It has been quite a journey that has changed me from a naive student to a grown up. Though this thesis has my name on it, I would not have come nearly as far without my family, friends and colleagues. Here I will try to express my gratitude to those who have supported me over the last few years.

First and foremost, I want to thank my supervisors, because without them this thesis would not have come together. **Francine** and **Dick** and **Michael**, thank you for believing in me through tough times. More than once I felt distraught and on the verge of a mental breakdown, but you always managed to motivate me and keep it together. **Francine**, with your ever critical eye for detail you brought my research to a higher level. You were always open to give advice whenever needed, either personal or professional. You are an endless source of information about plant pathogens, and you have taught me so much. **Dick**, you kept your head cool and always made me see the light at the end of the tunnel. You are a great manager, and I have deep respect for how you manage both a chair group as well as a family. In our meetings you always seemed to ask just the questions that I did not expect, and this has made me stronger and more critical at my own work. Whenever I had trouble understanding an algorithm or math problem, your explanations made it look so simple. It was a great pleasure to work under your supervision. **Michael**, as a fellow bioinformation in the world of plant pathology, you were a great source of inspiration, and this was indispensable for this thesis. You always came up with new and original research ideas whenever I presented some new results, and always triggered me to exploit existing data in completely novel ways. It was always fun to discuss scientific and other topics.

Jan, doing my MSc thesis under your supervision has sparked my interest in phytopathology, and I consider you the reason I ended up at the phyto lab, thank you for that. **Xiaoqian**, since my MSc thesis you were my partner in crime in the bioinformatics office. Only after you left I realized how much I liked writing scripts for you. Sorry that I have killed all your plants. I am happy to have you as my friend, and though our career paths have diverged, we will continue to have fun in private life. **Tim**, I envy your ever sober look on things. You also taught me that Germans do make jokes. **Jasper**, sometimes I think the Dutch and Belgians are not so different after all. We had great times in the office and also outside work. **Laura**, it is always great fun when we hang out. These few hours of driving will not stop us in the future. **David,** we only worked together for a short period, but it was great. Thanks amigo! **Hui**, the hotpot dinners were great! I learned many new types of edible noodles and mushrooms. **Nick** and **Martin** I always enjoyed the chats, especially over a few beers. To the "Pinf-group" members of my generation: **Elysa, Chara, Johan, Kiki, Shuqing, Klaas, Rob, Natalie** and others, it was a great pleasure working with you, and those who already graduated set a great example to me. I am also grateful to all other (ex-)members of phytopathology for the last years.

Miguel, **Mehmet, Janani, Siavash** thanks for the good times at the congresses. **Rens, Vittorio,** and all others from the Bioinformatics Group, thanks for the fun at the retreats, the chats and the nerd jokes!

Thanks to all my new colleagues at The Hyve for their mental support and understanding over the last year.

Besides all these people I met through my career, there are also people in my private life that are priceless to me. My siblings and their partners: **Luuk** and **Nathalie**, **Melissa** and **Jan-Luuk**, **Iris**, I know I can always count on you, you are among the most precious people in my life. **Richard, Rik** and **Manon**, **Mark** and **Janine**, **Roel** and **Serise, Arjan**, most of you I have known for so long now. Thanks for the support and all side-activities that have kept me sane.

Mama and **papa**, thanks for your unconditional love and support. Thanks for raising me the way I am, to pursue my interests, ask questions, explore, and try to understand. Without you I would not have come as far as I am today. **Opa Jan**, thanks for raising me with computers. I am also grateful to my two **grandmas** that did not live to see me graduate but have supported me over the years.

Finally, I cannot even begin to express my gratitude to **Kim**, my beautiful fiance. Without you, my life would look so different. We have been together long enough to see each other grow from students into two scientists. You have taught me perseverance and dedication. With this ending a new period in our life starts, and I cannot wait to find out what the future holds for us.

About the author

Sander Rodenburg was born on the 13th of Februari, 1991 in Leiden, The Netherlands. As a young kid, he said he wanted to become an "inventor". He was always interested in the working of things, and built large structures from K'Nex and Lego to recreate mechanisms he encountered in daily life. During his highschool years, he took a science route, and did his graduation project on the photosynthesis of plants. At home, Sander was often found at his computer, scripting self-made maps for games he played. In 2008, he combined both worlds, science and programming, and started his bachelor programme in bioinformatics at the Hogeschool of Arnhem and Nijmegen (HAN). During this study, he did an internship at the VUmc Cancer Center in Amsterdam, and in his final year he did a minor in information and communication technology (ICT) at the Fontys hogeschool in Eindhoven. He graduated with his BSc thesis at the Eindhoven University of Technology, working with synthetic biology. In 2012, he started his MSc programme in bioinformatics at Wageningen University. He performed his minor MSc thesis at the Phytopathology Lab, under the supervision of Dr Jan van Kan, where he investigated the transcriptome of the fungus *Botrytis cinerea* during its sexual reproductive cycle, familiarizing him with the field of plant pathology. He later performed his major MSc thesis under the supervision of Prof. Dick de Ridder (Bioinformatics Group) at Genetwister, a plant breeding R&D company. Here he developed a pipeline for image analyses on flower photographs to yield quantitative data for genome-wide association studies. After achieving his MSc degree in 2015, he applied for an open PhD position in a collaboration between the Phytopathology Lab and the Bioinformatics Group. Under the supervision of Professors Francine Govers and Dick de Ridder, he started his PhD on developing a physical interaction model for the tomato late blight pathogen *Phytophthora infestans*. Five years later, this resulted in this thesis. Today, Sander is working as a data scientist at The Hyve, a bio-IT company that facilitates scientific communities by open-source software.

Publications

Charikleia Schoina*, **Sander Y.A. Rodenburg***, Harold J.G. Meijer, Michael F. Seidl, Lysette T. Lacambra, Klaas Bouwmeester, Francine Govers (2020) Mining oomycete proteomes for metalloproteases leads to identification of candidate virulence factors in *Phytophthora infestans*. *Submitted*.
*Equal contribution

Sujaya Srinivasan, Natallia Kalinava, Rafael Aldana, Zhipan Li, Sjoerd van Hagen, **Sander Y.A. Rodenburg**, Megan Wind-Rotolo, Ariella S. Sasson, Hao Tang, Xiaozhong Qian, Stefan Kirov (2020) Mis-annotated multi nucleotide variants in public cancer genomics datasets can lead to inaccurate mutation calls with significant implications. *bioRxiv* 2020.06.05.136549.

Sander Y.A. Rodenburg, Dick de Ridder, Francine Govers, Michael F. Seidl (2020) Oomycete metabolism is highly dynamic and reflects lifestyle adaptations. *bioRxiv* 2020.02.12.941195.

Marco Thines, Rahul Sharma, **Sander Y.A. Rodenburg**, Anna Gogleva, Howard S. Judelson, Xiaojuan Xia, Johan van den Hoogen, Miloslav Kitner, Joël Klein, Manon Neilen, Dick de Ridder, Michael F. Seidl, Guido van den Ackerveken, Francine Govers, Sebastian Schornack, David J. Studholme (2020) The genome of *Peronospora belbahrii* reveals high heterozygosity, a low number of canonical effectors, and TC-rich promoters. *Molecular Plant-Microbe Interactions* 33.5: 742-753.

Sander Y.A. Rodenburg, Michael F. Seidl, Howard S. Judelson, Andrea L. Vu, Francine Govers, Dick de Ridder (2019) Metabolic model of the *Phytophthora infestans*-tomato interaction reveals metabolic switches during host colonization. *MBio* 10(4), e00454-19.

Sander Y.A. Rodenburg, Michael F. Seidl, Dick de Ridder, Francine Govers (2018) Genome-wide characterization of *Phytophthora infestans* metabolism: a systems biology approach. *Molecular plant pathology* 19.6: 1403-1413.

Sander Y.A. Rodenburg, Razak B. Terhem, Javier Veloso, Joost H.M. Stassen, Jan A.L. van Kan (2018) Functional analysis of mating type genes and transcriptome analysis during fruiting body development of *Botrytis cinerea*. *MBio*, 9(1).

Education Statement of the Graduate School Experimental Plant Sciences

Issued to: **Sander Y.A. Rodenburg**
Date: **15 September 2020**
Group: **Phytopathology & Bioinformatics**
University: **Wageningen University & Research**

1) Start-Up Phase	date	cp
▶ **First presentation of your project**		
Title: Blight vs. tomato: deducing a physical interaction network between *Phytophthora infestans* and *Solanum lycopersicum,* Wageningen, NL	13 Nov 2015	1,5
▶ **Writing or rewriting a project proposal**		
▶ **MSc courses**		
Subtotal Start-Up Phase		1,5

2) Scientific Exposure	date	cp
▶ **EPS PhD student days**		
EPS Get2Gether, Soest, NL	28–29 Jan 2016	0,6
EPS Get2Gether, Soest, NL	9–10 Feb 2017	0,6
EPS Get2Gether, Soest, NL	15–16 Feb 2018	0,6
▶ **EPS theme symposia**		
EPS Theme 2 Symposium & Willie Commelin Scholten Day 'Interactions between plants and biotic agents', Leiden, NL	22 Jan 2016	0,3
EPS Theme 4 Symposium 2016 'Genome biology', Wageningen, NL	16 Dec 2016	0,3
▶ **Lunteren Days and other national platforms**		
Annual Meeting Experimental Plant Sciences, Lunteren, NL	11–12 Apr 2016	0,6
Dutch Bioinformatics & Systems Biology conference (BioSB) + YoungCB retreat, Lunteren, NL	18–20 Apr 2016	0,8
Annual Meeting Experimental Plant Sciences, Lunteren, NL	10–11 Apr 2017	0,6
Dutch Bioinformatics & Systems Biology conference (BioSB) conference, Lunteren, NL	4–5 Apr 2017	0,6
Annual Meeting Experimental Plant Sciences, Lunteren, NL	9–10 Apr 2018	0,6
Annual Meeting Experimental Plant Sciences, Lunteren, NL	8–9 Apr 2019	0,6
▶ **Seminars (series), workshops and symposia**		
Seminar Gero Steinberg: 'Long-distance endosome trafficking drives fungal effector production during plant infection'	5 May 2015	0,1
Seminar Jane Parker: 'Plant intracellular immunity: evolutionary and molecular underpinnings'	21 Jan 2016	0,1
Seminar Laura Grenville-Briggs: 'Molecular oomycete-host interactions: The Good, The Bad and The Ugly'	19 Feb 2016	0,1

	Seminar Wenbo Ma: 'Effectors as molecular probes to understand pathogenesis'	20 Jun 2016	0,1
	Seminar Gabino Sanchez-Perez	14 Nov 2016	0,1
	Seminar Marjon de Vos	20 Oct 2017	0,1
	Seminar Remco Stam	26 Jan 2018	0,1
	Seminar Guido van der Ackerveken	8 Mar 2019	0,1
	Seminar Daniel Croll	21 Jun 2019	0,1
	Seminar - Celebrating five years of bioinformatics collaboration @EPS	10 Jul 2019	0,2
	WURomics symposium 'Technology-driven innovation for plant breeding', Wageningen, NL	15 Dec 2016	0,3
	Host-Microbes meeting, Wageningen, NL	27 Oct 2017	0,3
▶	**Seminar plus**		
▶	**International symposia and congresses**		
	Oomycete Molecular Genetics Network (OMGN) conference, Malmö, SE	15–17 Jun 2016	0,8
	15th European Conference on Computational Biology (ECCB), The Hague, NL	4–7 Sep 2016	0,9
	17th European Conference on Computational Biology (ECCB), Athens, GR	8–12 Sep 2018	1,2
▶	**Presentations**		
	Poster: Annual Meeting Experimental Plant Sciences, Interaction network reconstruction of the *Phytophthora infestans* – tomato pathosystem, Lunteren, NL	11–12 Apr 2016	1,0
	Poster: 15th European Conference on Computational Biology (ECCB), Reconstruction of the metabolic network of the *Phytophthora infestans* – tomato pathosystem, The Hague, NL	6 Sep 2016	1,0
	Poster: Annual Meeting Experimental Plant Sciences, Reconstruction of a metabolic model of the plant pathogen *Phytophthora infestans,* Lunteren, NL	10–11 Apr 2017	1,0
	Poster: Annual Meeting Experimental Plant Sciences, Modelling the metabolism of the tomato - *Phytophthora infestans* interaction, Lunteren, NL	9–10 Apr 2018	1,0
	Talk: Oomycete Molecular Genetics Network (OMGN) conference, Metabolic network construction of the *Phytophthora infestans* – Tomato pathosystem, Malmö, SE	15 Jun 2016	1,0
	Talk: B-wise, Deducing a physical interaction network between *Phytophthora infestans* and *Solanum lycopersicum,* Wageningen, NL	6 Dec 2016	1,0
	Talk: Annual Meeting Experimental Plant Sciences, Reconstruction of a metabolic model for the plant pathogen *Phytophthora infestans,* Lunteren, NL	11 Apr 2017	1,0
	Talk: Host Microbe Genetics Meeting, Genome-wide characterization of *Phytophthora infestans* metabolism: a systems approach, Wageningen, NL	27 Oct 2017	1,0
	Talk: Seminar - Celebrating five years of bioinformatics collaboration @EPS, Metabolic model of the *Phytophthora infestans*-tomato interaction reveals metabolic switches during host colonization, Wageningen, NL	10 Jul 2019	1,0
▶	**3rd year interview**		
▶	**Excursions**		
	YoungCB: The Hyve company visit	3 Sep 2015	0,2
	KeyGene company visit	12 Oct 2017	0,2
		Subtotal Scientific Exposure	20,2

3) In-Depth Studies	date	cp
► **Advanced scientific courses & workshops**		
BioSB: In silico life: constraint-based modelling at genome scale	9–11 May 2016	0,9
SPS Summer School: From gene expression to genomic network	17–22 Jul 2016	1,5
Data analyses and visualizations in R (for biologist)	12–13 Dec 2016	0,6
► **Journal club**		
Phytophthora research group literature discussion	2015–2018	1,0
Bioinformatics group literature discussion	2015–2018	1,0
► **Individual research training**		
Lab visit Judelson Lab - University of California, Riverside, USA	2 May – 4 Jul 2018	3,0
Subtotal In-Depth Studies		8,0

4) Personal Development	date	cp
► **General skill training courses**		
EPS Introduction course	11 Feb 2016	0,3
Wageningen Graduate Schools course Brain training	23 Mar 2016	0,3
Wageningen Graduate Schools course Career perspectives	8 Nov – 6 Dec 2018	1,6
Wageningen Graduate Schools course Scientific artwork: Photoshop and Illustrator	1–2 Mar 2016	0,6
Wageningen University & Research Library course Adobe InDesign Essentials Training	9–10 Oct 2017	0,4
► **Organisation of meetings, PhD courses or outreach activities**		
► **Membership of EPS PhD Council**		
Subtotal Personal Development		3,2

5) Teaching & Supervision Duties	date	cp
► **Courses**		
Genomics	26 Oct – 15 Dec 2015	1,0
Introduction to Bioinformatics	4 Sep – 27 Oct 2017	1,0
Genomics	19 Mar – 8 May 2018	1,0
► **Supervision of BSc/MSc students**		
Daniel Del Hoyo	8 Jan – 30 Jun 2018	3,0
Subtotal Teaching & Supervision Duties		6,0

TOTAL NUMBER OF CREDIT POINTS*	**38,9**

Herewith the Graduate School declares that the PhD candidate has complied with the educational requirements set by the Educational Committee of EPS with a minimum total of 30 ECTS credits.

** A credit represents a normative study load of 28 hours of study.*

This research was performed at the Laboratory of Phytopathology and the Bioinformatics Group of Wageningen University & Research. This work was financially supported by a private donor (via the University Fund Wageningen) and by Wageningen University & Research.

Cover illustration: Sander Rodenburg
Cover and Layout design: Iliana Boshoven-Gkini | AgileColor.com
Printed: GVO drukkers & vormgever B.V.