Predicting lifestyle and host from positive selection data and genome properties in oomycetes

Daniel Gómez-Pérez[1*,§] **and Eric Kemen**[1*,§]

[1] University of Tübingen

***For correspondence:** daniel.gomez-perez@uni-tuebingen.de (DGP); eric-kemen@uni-tuebingen.de (EK)

Present address: [§] Microbial interactions in plant ecosystems, ZMBP/IMIT, Tübingen

Abstract Host and niche shifts are a source of genomic and phenotypic diversification as evidenced in parasitism. Most characteristic is metabolism reduction as parasites adapt to a particular host. However, selection pressures acting on such organisms are not fully understood. Here, we developed a comparative genomic approach to study underlying adaptive trends in oomycetes, eukaryotes with a broad range of economically important parasitic lifestyles. Our analysis reveals converging evolution on biological processes for oomycetes with similar niche and host. We find that certain functions, in particular carbohydrate metabolism, transport, and signaling, are important for host and environmental adaption in oomycetes. Given an increasing availability of fungal and oomycete genomes from environmental samples, we have developed a model for the prediction of filamentous pathogen lifestyles based on genome properties. Understanding how genomes and selective pressures correlate with lifestyle may be crucial to identify new emerging diseases and pandemic threats.

Introduction

The adaptation of organisms as they evolve to occupy different niches or adopt different lifestyles is reflected on their genome. Expansion or contraction of gene families has been cited as a general mechanism for such adaptations (***Zhang et al., 2018***; ***Props et al., 2019***). Expansions arise mainly from gene duplication and, in some cases, from acquisition via horizontal gene transfer, whereas gene loss can happen by accumulation of loss-of-function mutations through genetic drift (***Scannell et al., 2006***; ***Treangen and Rocha, 2011***; ***Sheridan et al., 2020***). Fundamentally, both of these processes are driven by adaptive evolution, whereby beneficial mutations are selected for and deleterious removed from the gene pool, ultimately leading to phenotypic diversification (***Behe, 2010***). More concretely, trends in the evolution of coding genes can be studied by measuring the ratio of non-synonymous (dN) to synonymous (dS) amino acid rates in the comparison to closely related sequences, usually represented as ω (***Yang, 1998***). A ratio higher than one ($dN/dS = \omega > 1$) implies positive selection and thus functional diversification, while a ratio lower than one ($dN/dS = \omega < 1$) indicates the presence of purifying selection and thus a tighter constraint for the diversification of the gene sequence. Most genes in an organism are under strong purifying selection, as a change in a key amino acid of a protein would have a detrimental effect (***Hughes et al., 2003***). However, a small portion of them, those that have been subject to recent diversification, shows signs of an increased nonsynonymous mutation rate. Codon models that take into account statistical rate variations and are commonly used in comparative genomic studies (***Kosiol et al., 2008***). The study of positively selected genes together with their functional annotation evidences which of their functions were important for adaptation when performed on related organisms that have different lifestyles and hosts.

Oomycetes are eukaryotic organisms belonging, together with diatoms and brown algae, to the group of Stramenopiles (***Cavalier-Smith and Chao, 2006***). Since their origin from a marine autotrophic common ancestor around 400 million years ago, oomycetes have adapted to multiple environments and lifestyles, and most of them are economically impactful plant and animal parasites (***Matari and Blair, 2014***; ***Bebber and Gurr, 2015***; ***Derevnina et al., 2016***). Therefore, they represent a relevant and appropriate system to study the genetic impact of lifestyle and host adaptation on genetically close genomes. Four main phylogenetic orders compose the oomycetes: Albuginales, Peronosporales, Saprolegniales and Phythiales. The Albuginales and some Peronosporales independently evolved the ability to survive exclusively on living host material, also known as obligate biotrophy (***Kemen and Jones, 2012***). Most Peronosporales are, however, hemibiotrophs, i.e., they display an initial biotrophic phase followed by a necrotrophic one, during which they feed on the decaying living matter of their host (***Lee and Rose, 2010***). All Albuginales, Peronosporales, and most Pythiales are plant parasitic organisms (***Adhikari et al., 2013***). On the other hand, most Saprolegniales are capable of infecting animals (***Steciow et al., 2014***). Additionally, a free-living saprophyte is known, *Thraustotheca clavata*, which does not need a host at any point in its life cycle (***Misner et al., 2015***).

Obligate biotrophs have a significantly reduced primary metabolism. Comparative genome studies have reported a significant and convergent loss of the enzymatic arsenal for independent lineages of the oomycetes following this lifestyle (***McGowan et al., 2020***). The picture is not so clear for the heterotrophs and their adaptations to different hosts. *Pythium insidiosum*, a mammal parasite responsible for pythiosis, shows a relatively recent divergence from *Pythium aphanidermatum* and *Pythium arrhenomanes*, both of which are plant pathogens (***Weiblen et al., 2020***). How such drastic host shifts occur in a small evolutionary timescale is not fully understood. Some put forward the explanation that the large reservoir of noncoding DNA material readily evolves into small secreted proteins, known as effectors, facilitating oomycete-host interactions (***Haas et al., 2009***). Additionally, readiness to take up genetic material through horizontal gene transfer from fungi and bacteria has been reported at multiple time points in the oomycete lineages (***Savory et al., 2015***).

There is a high degree of convergent evolution between oomycetes and fungi (***Money et al., 2004***). Both share many of the niches mentioned, including pathogens of animals and plants, as well as lifestyles lifestyles, including saprotrophy, hemibiotrophy and obligate biotrophy. Oomycetes and fungi have developed similar strategies to overcome the same challenges, including a similar filamentous and reproductive morphology, as well as similar infection strategies (***Latijnhouwers et al., 2003***). As mentioned above, convergence is probably promoted by genetic exchange, as the source of many genes in oomycetes that play a role in host adaptation can be traced back to pathogenic fungi (***Richards et al., 2006***). Based on the parallels between the adaptive strategies of these two eukaryotic phyla, we can infer similar mechanisms in oomycetes as those further characterized in fungi.

How genome information relates to lifestyle and host adaptation is one of the big questions in ecology, and may be relevant to predict the appearance of new emerging diseases. Understanding the genome characteristics and selective pressures in organisms that have undergone host and niche shifts may offer insights into this question. In this study, we report the first whole-genome positive selection screening of the proteome of the oomycetes phylum, including 28 representative members and an outgroup of eight non-oomycete Stramenopiles (***Table 1***). We compared the genes inferred as being under positive selection to the background annotated genes to find enriched biological functions that may correlate to their adaption to different hosts and lifestyles. Additionally, we developed a method to predict with high accuracy plant pathogenic lifestyle from the genome of fungi and oomycetes, based on presence or absence of key annotated pathways.

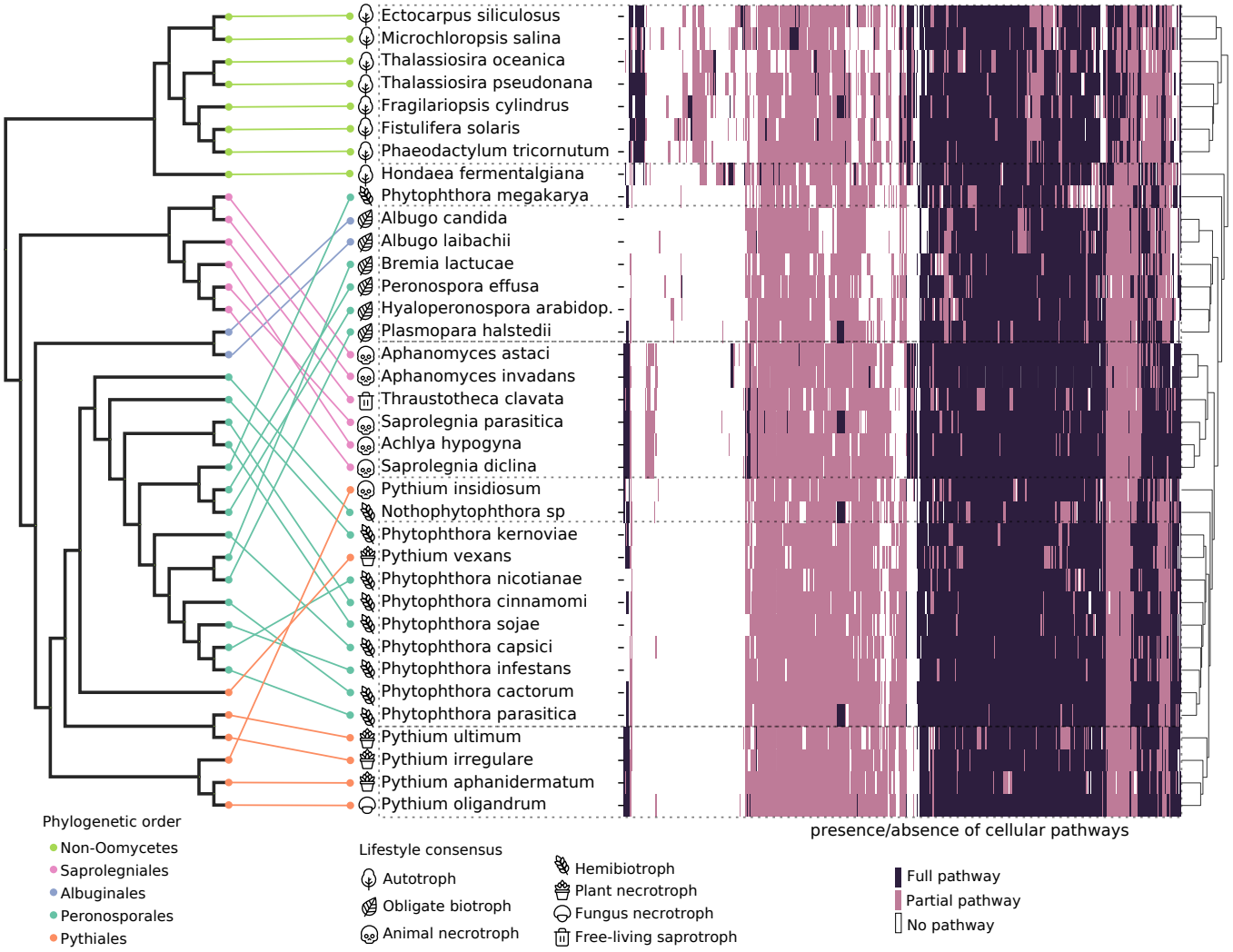

Figure 1. Presence/absence of cellular pathways in Stramenopiles correlated with phylogeny. Equal distance cladogram constructed from conserved families (present in at least 25 taxa) and clustering by WPGMA of genome properties of dataset. Dark purple corresponds to presence of the full pathway, light purple to partial presence of the pathway and white to no pathway. Dashed lines represent inferred clusters with a similar distribution of cellular pathways.

Figure 1–Figure supplement 1. Differences in annotated cellular pathways from the Stramenopile dataset. Shown are pathways which have up to 30 repeated values per taxa.

Figure 1–Figure supplement 2. Differences in annotated cellular pathways for the members of the Saprolegniales in the Stramenopile dataset.

Figure 1–Figure supplement 3. Differences in annotated cellular pathways for the members of the Pythiales order in the Stramenopile dataset.

Results

Proteome pathway annotation

We downloaded the genomes of 28 oomycete species and eight non-oomycete Stramenopiles from the NCBI and FungalDB databases and annotated their proteomes by the presence or absence of known cellular pathways to get insights into the divergence of the dataset (***Figure 1***). Weighted Pair Group Method with Arithmetic Mean (WPGMA) based on the Euclidean distance along with midpoint rooting, resulted in two main groups, one corresponding to the oomycetes and the other to the remaining Stramenopiles. The main difference among them was the lack of photosynthesis-related pathways in the oomycetes, such as chlorophyll biosynthesis (***Figure 1–Figure Supplement 1***). In the oomycetes, we defined four subclusters based on their distance (from top to bottom): obligate biotrophs, Saprolegniales, hemibiotrophs and Pythiales. The obligate biotroph cluster consists of the Albuginales and the downy mildews from the Peronosporales (*Bremia lactucae*, *Plasmopara halstedii*, *Peronospora effusa* and *Hyalopernospora arabidopsidis*). The most striking characteristic was an overall reduction of their metabolism, evident by the lack of some pathways in comparison with other oomycetes. Notable was the lack of core biosynthetic pathways of this group, including vitamin and cofactor biosynthesis, for which they presumably rely on their host (***Figure 1–Figure Supplement 1***). The hemibiotroph group, consisting of all the *Phytophthora* species in the dataset and the closely related *Pythium vexans*, recently reclassified as *Phytophythium vexans*, also showed significant metabolic reduction (***Cock et al., 2015***). The Saprolegniales differed from other oomycetes mainly in the presence of steroid biosynthesis pathways (***Figure 1–Figure Supplement 2***). The Pythiales group had also biosynthetic pathways that most other oomycetes lacked but that they often shared with the Saprolegniales, as a result most likely of their common facultative lifestyles.

The four clusters roughly reflected the lifestyles of the taxa in the dataset, mostly highlighted by the hemibiotrophs and obligate biotrophs. To a lesser extent this was evident in the other two groups as most Saprolegniales are facultative animal necrotrophs, and most Pythiales facultative plant necrotrophs. There were some exceptions to this, with some taxa clustering with a different lifestyle or failing to cluster with their own lifestyle. For example, *T. clavata*, the free-living organism in the dataset clustered with other phylogenetically close Saprolegniales. Nevertheless, it did show the greatest distance to its cluster neighbours. The most notable differences in the presence/absence of cellular functions when compared to other Saprolegniales were the presence of the xylose degradation pathway and the absence of the TFIIH basal transcription factor, the endopeptidase ClpXP complex and the RuvB-like helicase (***Figure 1–Figure Supplement 2***).

P. insidiosum, the only animal pathogen in the Pythiales, showed remarkably different genome properties than its peers, being placed as an outgroup of hemibiotrophs and Pythiales. It shared common pathways with the other animal pathogens in the dataset, namely, a methyltransferase that is part of the pterostilbene and serotonin/melatonin biosynthesis, which other plant-infecting *Pythiales* lacked. Of note, pterostilbene has been shown to have strong immunosuppressive properties in animals (***Yashiro et al., 2020***). Still, *P. insidiosum* retained some of the Pythiales nutrient assimilation pathways, including Leloir pathway for catabolism of D-galactose, methionine salvage and allantoin catabolism for, respectively, sulphur and nitrogen assimilation. Another outgroup of the same cluster was represented by *Nothophytophthora*, a hybrid species of the Peronosporales order of which little is known about. Most interesting was the presence of thiamine and particular thiazole biosynthesis genes for the synthesis of a key moiety of this cofactor, which all other *Phytophthora* have apparently lost. Based on this evidence, we speculate a facultative lifestyle for *Nothophytophthora*, in contrast to the hemitroph neighbour Peronosporales. It is not uncommon for hybridization to facilitate niche adaptation (***Samarasinghe et al., 2020***; ***Bertier et al., 2013***). In the Pythiales, the mycopathogen *Pythium oligandrum* clustered with plant pathogenic Phythiales. Notable is its lack of inositol degradation pathways and the presence of para-aminobenzoic acid biosynthesis from the chorismate pathway (***Figure 1–Figure Supplement 3***). In summary, our anal-

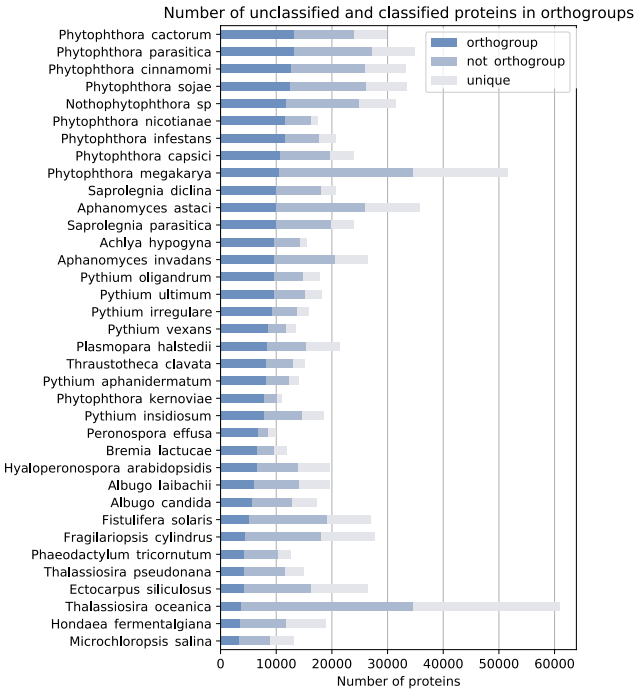

Figure 2. Protein-encoding genes from the Stramenopile dataset classified by taxa. Number of proteins classified into orthologs (five or more members) non-orthologs (less than five members) or unique (not in an ortholog group).

Figure 2–Figure supplement 1. Duplicates in protein families in the Stramenopile dataset. Number of families with five or more members that contain paralogs in the dataset.

Figure 2–Figure supplement 2. Distribution of protein family size in the dataset. Number of families with the same member size.

ysis indicates that loss and maintenance of metabolic and key regulatory genes in oomycetes is dependent to a larger extent on environmental and lifestyle factors than on phylogenetic evolutionary distance.

Ortholog group classification

To infer positive selection from the Stramenopile dataset of 36 genomes, we classified the proteomes into ortholog groups by taking sequence similarity and in addition gene order into account. We selected protein clusters of at least five members to get a good balance between a representative number of families and results that are statistically robust. This corresponded to 24,017 protein families, which cover 47.86% of the total proteins in the dataset (***Figure 2***). The selected orthogroups are mainly composed of one-to-one orthologs, however, we detected a significant number of paralogs for some oomycetes, particularly for *Nothophytophthora* sp., as well as for *Phytophthora nicotianae*, *Phytophthora parasitica* and *Phytophthora palmivora*. This might be related to the reported whole genome duplications in *Phytopthora* species (***Martens and Peer, 2010***), as well as the recent hybridization event that gave rise to *Nothophytophthora* (***Studholme et al., 2018***) (***Figure 2–Figure Supplement 1***). Additionally, the diatom *Fistulifera solaris*'s large presence of gene duplications highlights its recent whole genome duplication (***Tanaka et al., 2015***).

The more abundant orthogroups have between five and nine members (***Figure 2–Figure Supplement 1***). Orthogroups corresponding to all taxa were a minority. Instead, most orthogroups corresponded to closely related five to ten-member clades. When looking at the taxa with higher number of genes not classified into orthogroups from the oomycetes, the *Phytophthora* genus had the highest count (***Figure 2***). This may be related to the large arsenal of unique effectors with no

conserved domains or homologous outside of their own species that play a large role in host adaptation. *Aphanomyces astaci* also had a high amount, most likely because of the recent expansions in its genome (***Gaulin et al., 2018***). In summary, this highlights a patchy ortholog distribution in the dataset, with most protein families corresponding only to their close phylogenetic order. Despite this, a significant pool of ortholog protein families to test for positive selection could be inferred from the analysis.

Positive selection analyses

Positive selection screening for orthologous groups was performed by using first a site-specific codon model to detect families under selection. This was followed by a branch-site-specific codon model to detect the taxa experiencing positive selection on those genes. The number of genes under selection varied for the different phylogenetic orders. Members of the Saprolegniales and Pythiales had a higher count and therefore more genes under selection in orthogroups (mean = 1036.58, std = 110.75) than Peronosporales and Albugo (mean= 516.56, std = 225.29) (***Figure 3–Figure Supplement 1***). A special case was the hybrid *Nothophytophthora* sp., which had a comparable amount of positively selected genes to Pythiales and Saprolegniales, however composed in great part by predicted duplicated genes after speciation, 37.2% of the total (orange bar). When comparing hemibiotrophs and obligate biotrophs of the Peronosporales order (mean = 570 and 270, respectively), the trend was that of a decrease in the number of genes under positive selection for the biotrophs (***Figure 3–Figure Supplement 1***).

To infer potential biases in our analyses we tested for a correlation between the number of genes under positive selection and the amount of proteins classified into orthogroups for each taxa (Pearson's correlation, $r = 0.54$, p value < 0.01). A correlation of 0.54 suggested that there is a larger number of positives because of more extensive testing in the oomycete species, as they have in average more members in the ortholog dataset. This bias is more evident in the non-oomycetes (Pearson's correlation, $r = 0.29$, p value = 0.50) than when considering just the oomycetes (Pearson's correlation, $r = 0.17$, p value = 0.38). As the proteomes of the non-oomycetes are overall smaller compared to oomycetes (***Figure 4***), we hypothesize that less testing renders them more prone to this bias.

Out of the 23,254 detected genes under positive selection, 14,897 were successfully annotated with at least a gene ontology term (64.06%). We performed GO enrichment on the four clusters with similar metabolism. The results are discussed below. As a control for our pipeline, we performed the same analyses in a subset of 26 plant pathogenens from a dataset of 65 basidiomycete fungi (***Table 2***). Highly enriched terms include known virulence factors in such pathogenic fungi, including fatty acid and certain amino acid biosynthesis, ion transport and protein targeting and transport (***Table 4***) (***Stephenson et al., 1997***; ***Hallen et al., 2007***; ***Walley et al., 2013***). In summary, we could identify signatures of positive selection in 7.75% of all genes analyzed in the Stramenopile dataset. A significant number could be functionally annotated and potential functions assigned.

Enriched biological functions under selection

To gauge the selective pressures for adaptation to a parasitic lifestyle in the oomycetes, we explored the enriched GO terms that were pervasive in all oomycetes (***Figure 5***). Highly enriched term categories related to catabolic processes, response to stress, transmembrane transport, cellular component assembly and localization, protein phosphorylation as well as nitrogen, lipid, and carbohydrate metabolism. In the cellular compartment category, highly enriched terms include protein-containing complexes, nucleus, intracellular organelles and membranes ***Figure 5–Figure Supplement 1***.

Additionally, we performed similar enrichments on the groups defined in the presence/absence of cellular pathway analysis (***Figure 1***). We found the largest unique GO terms to belong to the Saprolegniales group. We observed the largest overlap coincides between the Pythiales and Saprolegniales group, followed by a smaller overlap of enriched functions in the plant pathogens of the

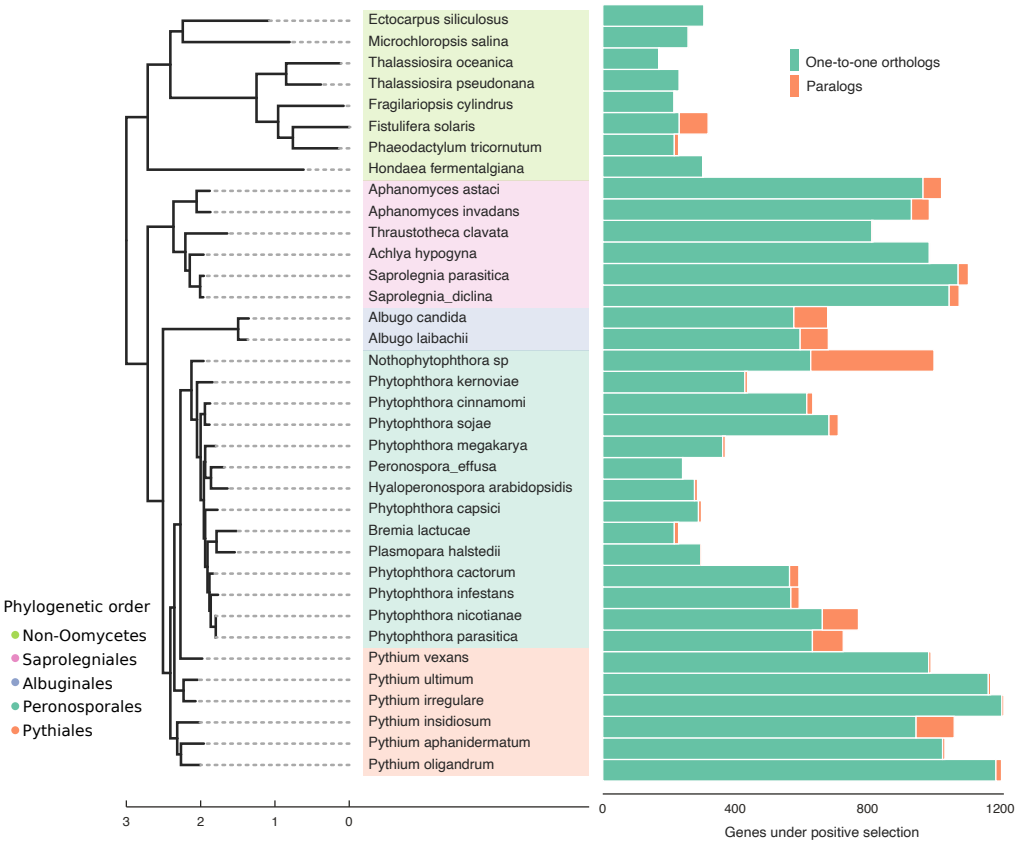

Figure 3. Positive selection in oomycetes. Maximum likelihood supertree constructed from inferred protein families in the Stramenopile dataset that are conserved in at least 25 taxa. Positively selected genes are represented as bars. One-to-one orthologs are represented as green, duplicated genes inside the same family under positive selection in orange.

Figure 3–Figure supplement 1. Comparison of the distribution of positive genes under selection for different lifestyles. Specialists refers to Albuginales and Peronosporales and generalists to Pythiales and Saprolegniales.

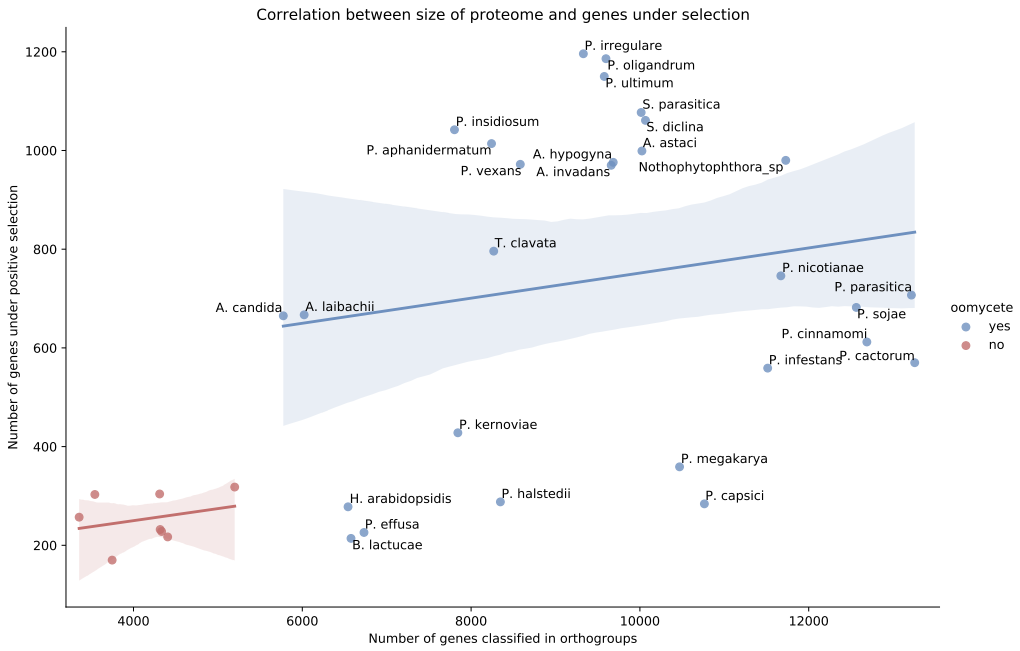

Figure 4. Correlation between genes under positive selection and proteome size in the Stramenopile dataset. Oomycetes are in blue and non-oomycetes in red.

dataset: Pythiales, obligate biotrophs and hemibiotrophs (***Figure 6***). The most significant terms are listed in ***Table 5***, ***Table 7***, ***Table 6***, ***Table 8***, and ***Table 9***. With respect to the group-specific terms, Saprolegniales had the largest number (74) amounting to almost half of the total terms. In the remaining groups this corresponded to 10 - 20 %.

We also studied the enrichment of biological functions in these expanded gene families and, in general, found that they reflected positive selection enrichment (***Table 10***). In the obligate biotrophs of the Stramenopile dataset, these related to phospholipid metabolism, cell wall biosynthesis, protein modification and transport. In the hemibiotrophs, they relate to transmembrane transport, cellular component organization and genetic interchange. In the Pythiales, carbohydrate metabolism and calcium transmembrane transport. Finally, in the Saprolegniales, molybdopterin cofactor biosynthesis, phosphatidylinositol metabolism, transcription regulation, and G receptor signaling pathway.

Lifestyle prediction with machine learning model

We visualized in a heatmap all protein annotations and added the information of positive selection by performing the same clustering as we did for pathway completeness (***Figure 7***). We find that there is a higher congruence to the species phylogeny than to the lifestyle clustering (Robison-Foulds cluster distance: 32.5 and 34.0 respectively). Based on this observation, we conclude that the presence/absence of model key pathways is a better indication of lifestyle in oomycetes than positively selected gene functions. Therefore, we constructed a model to predict lifestyle in plant pathogenic fungi and oomycetes. We assembled a dataset based on 107 plant pathogenic and saprotrophic fungi and oomycetes genomes (***Table 3***). Using this dataset, we built a deep neural network classifier with four output classes corresponding to their lifestyle consensus in the literature: saprotroph, necrotroph, hemibiotroph and biotroph. We found a high accuracy on the validation dataset for the optimized model (loss = 0.78, accuracy = 0.97), only failing to predict one genome (***Figure 8***).The model and the steps to reproduce it together with the entire dataset can be found at https://github.com/danielzmbp/lspred.

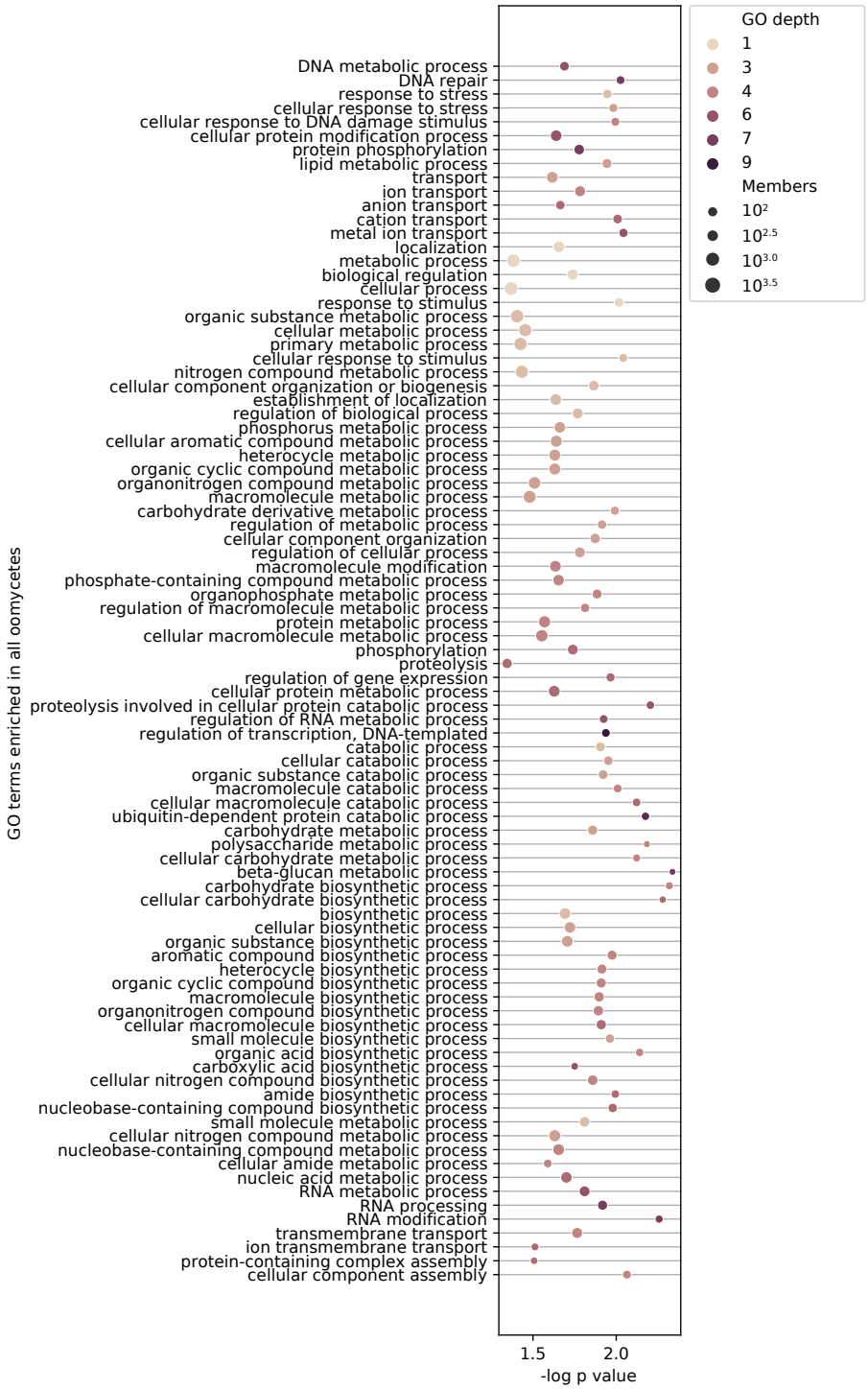

Figure 5. Significantly enriched terms with a corrected p-value of less than 0.05 relating to biological processes in all oomycetes in the Stramenopile dataset. GO terms were categorized using GO slim database. The color represents the GO depth. GO depth is a measure of the number of parent nodes in the GO tree. That is, the more specific the GO term the higher its depth. The size of the dots corresponds to the total number of proteins under selection in the Stramenopile dataset that belong to said term.

Figure 5–Figure supplement 1. Cellular compartment enrichment under selection in all oomycetes in the Stramenopile dataset.

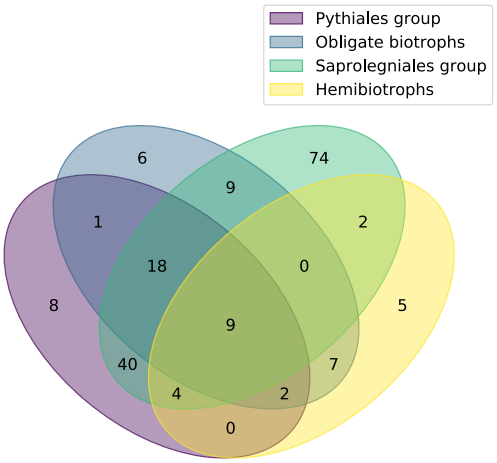

Figure 6. Venn diagram of unique overlapping biological functions under selection in the oomycetes groups clustered by convergence of presence/absence of cellular pathways. The four groups represented correspond to the defined clusters in ***Figure 1***.

Discussion

Cellular pathways largely correlate with lifestyle

Convergence of the presence/absence of key pathways in distant species that have similar lifestyles has been shown before for different sets of organisms (***McCutcheon and Moran, 2010***; ***Shang et al., 2016***). Different lifestyles require different biological functional processes, resulting in similar selective pressures that shape their genome in an analogous way. Comparable to the study by ***Rodenburg et al.*** (***2020***), we have shown the tight clustering of some groups with a similar lifestyle, most strikingly for the obligate biotrophs and hemibiotrophs. Conversely, there is one exception, the hemibiotroph *Phytophthora megakarya*, which does not clearly cluster with any of the other oomycetes. We hypothesize this may be due to the quality of its gene annotation, which has significantly lower number of key orthologs from Stramenopiles as compared to other *Phytophthora* spp. in the dataset (***Table 1***).

Generalists have more genes under positive selection

A higher number of genes under selection was found for the more generalist orders of Saprolegniales and Pythiales when compared to the more specialists Peronosporales and Albuginales (Mann-Whitney test, $p < 0.01$). Furthermore, obligate biotrophs in the Peronosporales have lower number than the hemibiotrophs (Mann-Whitney test, $p < 0.01$) (***Figure 3–Figure Supplement 1***). This cannot be explained alone by the different sizes of the proteomes in the dataset or by their phylogenetic closeness (***Figure 4***). However, we hypothesize that both of these factors confound our results to a large extent. Smaller proteomes in the dataset, as is the case of the non-oomycetes, show a larger correlation of their size to the number of genes under positive selection. The phylogeny influence is highlighted by the similar number of genes under positive selection of taxa within the same genera as shown in ***Figure 4***.

The large overrepresentation of paralogs as positive selected genes is evident in many of the taxa. After a gene duplication event takes place, there is usually an increase in the selective pressure on one of the copies that maintains the function. Meanwhile, in the other one, these constraints are relaxed, freeing it for potential divergent evolution (***Zhang et al., 1998***). Interestingly, many of the enriched functions in the paralogs correlated with terms under positive selection ***Table 10***. In the *Phytophthora* lineages these include microtubule-based processes, biological reg-

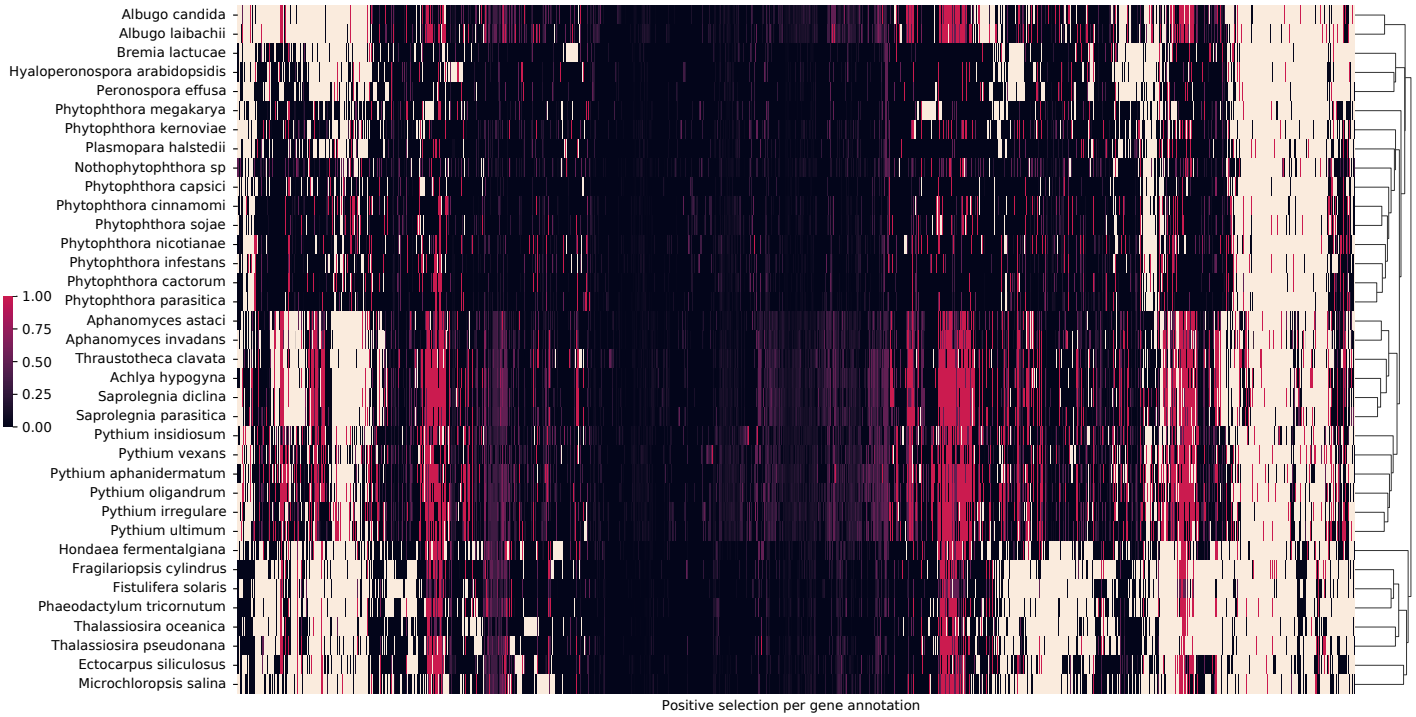

Figure 7. Heatmap of positive selection ratio of functional annotations in the Stramenopile dataset. The ratio represents the percentage of genes with this functional annotation that are under selection. Weighted-based clustering of this information is represented.

Figure 7–Figure supplement 1. Tetrahydrofolate salvage and biosynthesis-related enzymes in Stramenopiles. Heatmap of the presence and absence of the enzymes relating to tetrahydropholate metabolism in the Stramenopiles. The yellow gradient represents the ratio of predicted positive selection in genes with this annotation.

Figure 7–Figure supplement 2. Molybdopterin synthesis enzymes in Stramenopiles. Heatmap of the presence and absence of the enzymes relating to molybdopterin biosynthesis in the Stramenopiles. The yellow gradient represents the ratio of predicted positive selection in genes with this annotation.

Figure 7–Figure supplement 3. Glycolysis I, II and III-related enzymes in Stramenopiles. Heatmap of the presence and absence of the enzymes relating to glycolysis pathway in the Stramenopiles. The yellow gradient represents the normalized ratio of predicted positive selection in genes with this annotation.

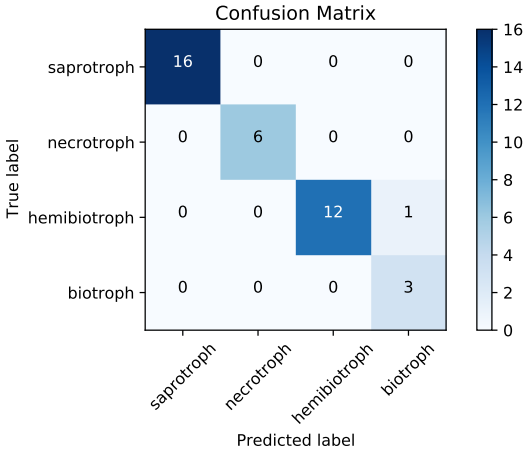

Figure 8. Confusion matrix of lifestyle predictor model. Results of predictions in the random validation set of 39 annotated proteomes.

ulation, and transmembrane transport. In Albugo and other obligate biotrophs, these are related to phosphatidylinositol, beta-glucan biosynthesis and protein modification. In the Saprolegniales, these are the more specific G protein-coupled receptor signaling and phosphatidylinositol metabolism for *Saprolegnia diclina* and molybdopterin cofactor biosynthesis and transcription regulation for *Aphanomyces invadans*.

While all Peronosporales are hemibiotrophs or obligate plant parasites, Saprolegniales and Pythiales show adaptation to a variety of lifestyles, including free-living (*T. clavata*) and facultative parasites of either animals (*P. insidiosum*, *Achlya hypogina*, *Aphanomyces* and *Saprolegnia* spp.) or other fungi and oomycetes (*P. oligandrum*). Facultative parasites can live as saprotrophs on decaying matter but also as opportunistic necrotrophs on a host (***LEWIS, 1973***). The higher number of potential niches they are able to occupy may be positively correlated with the amount of genes under diversifying selection. Additionally, when compared to the Peronosporales, which are highly adapted to infect a particular species, e.g., lettuce for *B. lactucae* and soybean for *Phytophthora sojae*, most of the Pythiales and Saprolegniales are able to infect a wide range of hosts. For instance, *A. astaci* is capable of infecting up to twelve genera of crayfish and is known for its ease of host jumping (***Svoboda et al., 2016***). Having a higher number of genes under positive selection could explain, at least partially, host flexibility.

Selective pressures in the ooomycetes help explain host adaptation

Biological functions under selection for all oomycetes, shown in ***Figure 5***, tell us which of these are important for the diversification in this clade. Many biosynthetic functions, particularly related to carbohydrates, are found to be enriched. Lipid metabolism, known to be important for host adaptation, is also enriched (***Wang et al., 2012***). Transport-related proteins, and in particular metal ion transport, are prominently enriched in these terms. The role of the expanded calcium transporter genes in the oomycetes has been extensively studied in the context of host interaction (***Zheng and Mackrill, 2016***). Overall, many of these terms allude to important virulence factors known for the oomycetes: transmembrane transport, effector protein processing and secretion, cell wall synthesis and remodeling, and lipid localization (***Judelson, 2012***).

Plant parasites with different lifestyles have similar selective pressures

The enriched terms common to the Albuginales and obligate biotrophs in the Peronosporales greatly relate to known virulence factors for these plant pathogens, including carbohydrate metabolism, protein modification, transport, and cell organization. This suggests that these biological functions are under selection and played a big role in the adaptation of oomycetes to an obligate biotrophic lifestyle. Some of these, particularly carbohydrate metabolism, transport, and protein modification, are common to the other plant pathogens in the hemibiotrophs and Pythiales group (***Table 6*** and ***Table 7***), highlighting a broader mechanism of adaptation to a plant-parasitic lifestyle.

The most significant terms enriched in the obligate biotrophs of the dataset mainly relate to the biosynthesis and metabolism of carbohydrates. These underscore the requirements for rapid growth during the hyphal stage, which depends largely on the synthesis of carbohydrates with the resources obtained from the plant host. Beta-glucan biostynthesis in particular, an important component of the oomycete's cell wall, which is also an elicitor of the plant immune response, features prominently in the enriched terms (***Klarzynski et al., 2000***).

Secretion of small proteins, known as effectors, is key for host adaptation in plant pathogenic oomycetes, as in other fungal filamentous pathogens (***Rocafort et al., 2020***). Many unique effector proteins have been characterized in the oomycetes that contribute to virulence by modulating the immune response of the plant (***Wang et al., 2019***). Therefore, this dependence on the secretion machinery of the cell for successful infection and thus survival has led to high selective pressures in their genome. Proteins predicted to be secreted in the Stramenopile dataset were found to be enriched in the positively selected terms in all taxa (hypergeometric test, $p = < 0.01$). When looking at the enrichment per species, the majority of the taxa in the Peronosporales order, which sig-

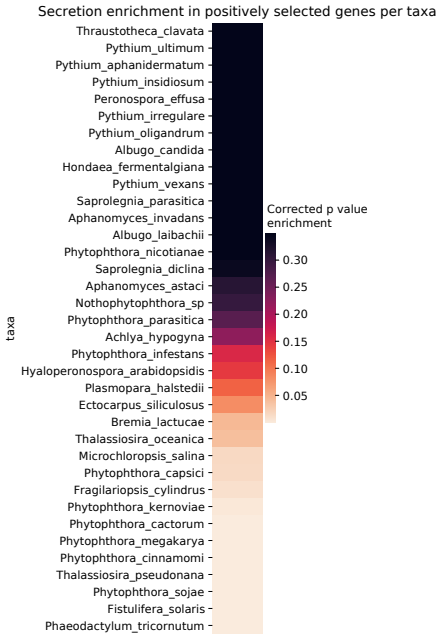

Figure 9. Enrichment of genes under positive selection predicted to be secreted in the Stramenopile dataset. Color gradient represents p values from hypergeometric tests per taxa corrected for multiple testing using bonferroni. A lighter color represents a more significant enrichment.

Figure 9–Figure supplement 1. Correlation between genes under selection and secretion. Oomycetes represented in blue (Pearson's correlation, r = 0.6, p = < 0.01) and non-oomycetes in red (Pearson's correlation, r = 0.5, p = 0.2). Pearson correlation represented as a straight line and the confidence interval represented as a lighter shade.

nificantly depend on effector proteins for host infection are also enriched (***Figure 9***). Albuginales, which also depend on secreted proteins, don't show enrichment in our analysis. This may be due to the lack of orthologs on host specific effectors and thus not found in the positive selection screen. There's a high correlation between number of positively selected genes compared to those with a predicted secretion signal ***Figure 9–Figure Supplement 1***. In the GO enrichment for all oomycetes there is an abundance of processes related to protein secretion under selection, including protein modification and vesicular transport. Other secretion-related terms, although more general, also show enrichment including those relating to microtubule assembly and movement, cellular organization, and transmembrane transport in the *Phytophthora*.

The hemibiotrophs present some unique terms. Interesting are those related to response to stress and base excision repair, a mechanism for DNA repair. These may be related to the defense against plant oxidative stress during the necrotrophic phase (***Torres, 2010***). Another virulence-related term, sulfur amino acid biosynthesis, features as an enriched term. This may be associated with the abundance of cysteine-rich proteins are abundant in the effector arsenal of the *Phytophthora* genus (***Raffaele et al., 2010***). The disulfide bonds help maintain the structure of the proteins released into the apoplast, a slightly acidic environment rich in plant proteases (***Idänheimo et al., 2014***).

Selective pressures in Saprolegniales and Pythiales differ from other oomycetes

Many terms in the Pythiales group overlap with the Saprolegniales, most likely relating to their facultative saprobe lifestyle. These include nucleic acid metabolism, citrate metabolism, lipid metabolism (in particular phosphatidylinositol), NADP biosynthesis, and heterocyclic compound biosynthesis, most likely as a result of their autotrophic and more developed secondary metabolism compared to that of other oomycetes (***Dahlin et al., 2017***). Unique functional terms present in the Saproleg-

niales include: signal transduction, including G protein-coupled receptor signaling pathways (also enriched in the paralogs from *S. diclina*), ncRNA processing, protein insertion in the membrane, regulation of gene expression and DNA repair and recombination. Those relating to metabolism include: vitamin and cofactor biosynthesis, such as folic acid, gluconeogenesis, amino acid biosynthesis, including serine and threonine, tRNA metabolism, glycolipid biosynthesis. Here, we discuss some of these functions important for adaptation.

Vitamin biosynthesis plays a big role in the evolution of pathogen adaptation to its host. Vitamins are expensive to produce and often require dedicated pathways. Heterotrophs that have adapted to obligate biotrophic lifestyles, such as Albugo and the downy mildews, circumvent this by losing their biosynthetic capability and developing ways of utilizating their host's vitamin supply (***Ah-Fong et al., 2019***). Meanwhile, those that live without a host at any point in their lifecycle must maintain these pathways under strong purifying selection. In our dataset we have found signatures of positive selection in several enzymes relating to tedrahydrofolate (THF) salvage and biosynthesis, namely dihydrofolate synthase and phosphoribosylglycinamide formyltransferase (***Figure 7–Figure Supplement 1***). As THF is a derivative of Vitamin B9 or folic acid, it is crucial for the synthesis of several amino acids such as serine and methionine as well as for purines and thiamine (***Huennekens, 1969***). It is therefore likely that oomycetes that are not able to get THF from a living host have strong selection to maintain THF metabolism in order to ensure their own amino acid biosynthesis.

Molybdopterin cofactor is important for the production of certain detoxification enzymes (***Schwarz and Mendel, 2006***). In oomycete obligate biotrophs, molybdopterin-requiring biosynthetic pathways have been lost several times, due to host adaptation (***Kemen and Jones, 2012***). As it would be expected, molybdopterin metabolism was found under high selective pressure in the *Saprolegniales* order and *Phytophthora* genus (***Figure 7–Figure Supplement 2***). The transport of molybdate ions also features as a positively selected term in Saprolegniales (***Table 8*** and ***Table 9***).

Proteins relating to the glycolysis pathway and amino acid biosynthesis are a special case in oomycetes (***Judelson, 2016***). Many of these enzymes have originated from horizontal gene transfer from plants or bacteria. This might explain their high rate of positive selection, which is usually the case for genes recently acquired by horizontal transfer, which need to be adapted for the new host. In the glycolysis pathway, we detected signatures of positive selection for most oomycetes in the Stramenopile dataset. Particularly in the enzymes phosphoglycerate mutase, glyceraldehyde-3-phosphate dehydrogenase, and fructose 1,6 biphosphatase (***Figure 7–Figure Supplement 3***).

A model based on presence/absence of cellular pathways accurately predicts lifestyle

The genome convergence of phylogenetically diverse fungi and oomycetes allowed us to create a model that can predict plant pathogenic lifestyle based on annotations from both eukaryotes. To our knowledge, there is only another published model that attempts to predict lifestyle (***Hane et al., 2020***). In this model they predict trophic categories based on principal component analysis centroid distance of carbohydrate-active enzyme annotations. We find that a model such as ours, based on the entire genome annotations, allows for a better overall accuracy. Furthermore, having a larger training dataset allows for the prediction of incompletely annotated specimens that may result from environmental sampling. Given the availability of more and more proteomic and transcriptomic data of unknown fungal and oomycete origin such prediction tools will become crucial to identify pathogenic potentials of facultative and obligate parasites.

Conclusions

The presence/absence of metabolism-related genes is known to converge for phylogenetically distant organisms that follow the same lifestyle (***Torruella et al., 2015***; ***Rodenburg et al., 2020***). Here, we report a similar case for our dataset of Stramenopiles. In addition, we describe a pipeline for seamless throughput analysis of positive selective pressures using genome data as input. We employ it to show that patterns of selective pressure also converge on hosts that cannot be explained

by phylogeny alone. We have identified a number of GOs that are commonly found under selection for all oomycetes of different lifestyles. We explored lifestyle-specific adaptive genes that correspond to biological regulation, transport, protein modification and metabolite biosynthesis. Our results help explain the selective pressures of closely related organisms that have adapted to different lifestyles. Finally, we described a model based on genome properties that is able to accurate predict plant pathogenic lifestyle on filamentous fungi and oomycetes.

Methods

Data selection and functional annotation

Stramenopile genetic data were downloaded from the NCBI and FungiDB databases setting as cut-off assemblies with reported gene annotation, resulting in a dataset of 36 total proteomes. The genomes were screened using BUSCO for high abundance of key orthologs in the Stramenopile dataset as a form of quality control (***Seppey et al., 2019***). Annotation of the proteomes was performed using InterProScan version 5.39-77.0 (***Jones et al., 2014***). Secretion signal of the proteins was annotated using the prediction tool SignalP 4.1 (***Petersen et al., 2011***). Annotation of the cellular pathways from each genome was accomplished with the Genome Properties database and visualized with the Python package Seaborn (***Richardson et al., 2018***). Comparison of midpoint rooted cladograms from the clustering analysis and the phylogenetic tree was performed using the Robison-Foulds metric based on clusters with the application TreeCmp (***Robinson and Foulds, 1981***; ***Goluch et al., 2020***).

Orthogroup classification and positive selection analyses

We developed a pipeline for whole genome positive selection analysis in Python using the Snakemake modular workflow framework (***Köster and Rahmann, 2012***). It uses as input the coding nucleotide sequences as well as their corresponding predicted proteins from each proteome. The code, documentation and examples are freely available at https://github.com/danielzmbp/wsgups. The steps of the pipeline include: grouping of sequences into families, their alignment with MAFFT, phylogenetic tree inference using FastTree, codon alignment using PAL2NAL, and finally two positive selection algorithms from the HYPHY package (***Katoh et al., 2009***; ***Price et al., 2010***; ***Suyama et al., 2006***). The first step, consisting of the classification of these proteomes into ortholog groups was performed with the software Proteinortho version 6, using the synteny parameter and the Diamond algorithm for homology search (***Lechner et al., 2011***). The first HYPHY algorithm used in the pipeline is FUBAR, a site-based program that scans the alignment for pervasive positive selection (***Smith et al., 2015***). Families with at least one codon position under positive selection were subsequently analyzed on all branches with the aBSREL algorithm to relate selective pressures to specific lineages (***Murrell et al., 2013***).

Enrichment analyses

We used the Gene Ontology (GO) released in 2020-06-01 (***Ashburner et al., 2000***; ***Carbon et al., 2018***). GO enrichment and visualization were performed using the Python package Goatools based on the InterPro database annotations (***Klopfenstein et al., 2018***; ***Hunter et al., 2009***). The background dataset corresponded to the sum of all proteome annotations for the corresponding taxa and the study dataset to the genes found to be under selection. Terms that did not have representative sequences in all analyzed taxa were filtered out. The negative log10 of Holm-Bonferroni corrected p values were used to assess the significance of the terms. Broad and non-informative GO terms like biological or cellular processes were not included in the enrichment tables.

Machine learning model

The multilayered deep learning model was constructed using the Tensorflow version 2.3 library with the Keras API (***Abadi et al., 2015***). The training dataset consisted of 367 unique proteomes

from fungi and oomycete plant pathogens. We labeled each proteome as one of the four respective plant pathogenic classes based on literature consensus: sapotroph, necrotroph, hemibiotroph and biotroph. We extracted the features of each genome and encoded them based on the presence or absence of all the identified pathways, which resulted in an array of 5024 binary features each. We randomly split the dataset into training dataset, corresponding to 80 % of the total, and optimization and validation datasets, each corresponding to half of the remaining 20 %. Hyperparameter optimization, namely learning rate, activating functions and dense layer units, was carried out using Keras Tuner and its implementation of the Hyperband algorithm (***O'Malley et al., 2019***; ***Li et al., 2018***).

Acknowledgments

The authors would like to thank Libera Lo Presti for her comments and suggestions. The authors acknowledge support by the High Performance and Cloud Computing Group at the Zentrum für Datenverarbeitung of the University of Tübingen, the state of Baden-Württemberg through bwHPC and the German Research Foundation (DFG) through grant no INST 37/935-1 FUGG. In addition, the authors are grateful to the DFG-funded research training group RTG 1708 'Molecular principles of bacterial survival strategies' (grant # 174858087; Y.H) for providing the funding for the development of this manuscript.

Competing interests

The authors declare that they have no competing interests.

References

Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado GS, Davis A, Dean J, Devin M, Ghemawat S, Goodfellow I, Harp A, Irving G, Isard M, Jia Y, Jozefowicz R, Kaiser L, Kudlur M, Levenberg J, et al., TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems; 2015. http://tensorflow.org/, software available from tensorflow.org.

Adhikari BN, Hamilton JP, Zerillo MM, Tisserat N, Lévesque CA, Buell CR. Comparative Genomics Reveals Insight into Virulence Strategies of Plant Pathogenic Oomycetes. PLoS ONE. 2013; 8(10). doi: 10.1371/journal.pone.0075072.

Ah-Fong AMV, Kagda MS, Abrahamian M, Judelson HS. Niche-specific metabolic adaptation in biotrophic and necrotrophic oomycetes is manifested in differential use of nutrients, variation in gene content, and enzyme evolution. PLOS Pathogens. 2019; 15(4):e1007729. doi: 10.1371/journal.ppat.1007729.

Ali SS, Shao J, Lary DJ, Kronmiller BA, Shen D, Strem MD, Amoako-Attah I, Akrofi AY, Begoude BAD, Hoopen GMt, Coulibaly K, Kebe BI, Melnick RL, Guiltinan MJ, Tyler BM, Meinhardt LW, Bailey BA. 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 Biology and Evolution. 2017; 9(3):536–557. doi: 10.1093/gbe/evx021.

Armitage AD, Lysøe E, Nellist CF, Lewis LA, Cano LM, Harrison RJ, Brurberg MB. Bioinformatic characterisation of the effector repertoire of the strawberry pathogen Phytophthora cactorum. PLOS ONE. 2018; 13(10):e0202305. doi: 10.1371/journal.pone.0202305.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene Ontology: tool for the unification of biology. Nature Genetics. 2000; 25(1):25–29. doi: 10.1038/75556.

Baxter L, Tripathy S, Ishaque N, Boot N, Cabral A, Kemen E, Thines M, Ah-Fong A, Anderson R, Badejoko W, Bittner-Eddy P, Boore JL, Chibucos MC, Coates M, Dehal P, Delehaunty K, Dong S, Downton P, Dumas B, Fabro G, et al. Signatures of Adaptation to Obligate Biotrophy in the Hyaloperonospora arabidopsidis Genome. Science. 2010; 330(6010):1549–1551. doi: 10.1126/science.1195203.

Bebber DP, Gurr SJ. Crop-destroying fungal and oomycete pathogens challenge food security. Fungal Genetics and Biology. 2015; 74:62–64. doi: 10.1016/j.fgb.2014.10.012.

Behe M. Experimental evolution, loss-of-function mutations, and "the first rule of adaptive evolution". The Quarterly Review of Biology. 2010; 85(4):419–445. doi: 10.1086/656902.

Bertier L, Leus L, D'hondt L, Cock AWAMd, Höfte M. Host Adaptation and Speciation through Hybridization and Polyploidy in Phytophthora. PLoS ONE. 2013; 8(12):e85385. doi: 10.1371/journal.pone.0085385.

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, et al. The Phaeodactylum genome reveals the evolutionary history of diatom genomes. Nature. 2008; 456(7219):239–244. doi: 10.1038/nature07410.

Carbon S, Douglass E, Dunn N, Good B, Harris NL, Lewis SE, Mungall CJ, Basu S, Chisholm RL, Dodson RJ, Hartline E, Fey P, Thomas PD, Albou LP, Ebert D, Kesling MJ, Mi H, Muruganujan A, Huang X, Poudel S, et al. The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Research. 2018; 47(D1):gky1055–. doi: 10.1093/nar/gky1055.

Cavalier-Smith T, Chao EEY. Phylogeny and Megasystematics of Phagotrophic Heterokonts (Kingdom Chromista). Journal of Molecular Evolution. 2006; 62(4):388–420. doi: 10.1007/s00239-004-0353-8.

Cock AWAMd, Lodhi AM, Rintoul TL, Bala K, Robideau GP, Abad ZG, Coffey MD, Shahzad S, Lévesque CA. Phytopythium: molecular phylogeny and systematics. Persoonia - Molecular Phylogeny and Evolution of Fungi. 2015; 34(1):25–39. doi: 10.3767/003158515x685382.

Cock JM, Sterck L, Rouzé P, Scornet D, Allen AE, Amoutzias G, Anthouard V, Artiguenave F, Aury JM, Badger JH, Beszteri B, Billiau K, Bonnet E, Bothwell JH, Bowler C, Boyen C, Brownlee C, Carrano CJ, Charrier B, Cho GY, et al. The Ectocarpus genome and the independent evolution of multicellularity in brown algae. Nature. 2010; 465(7298):617–621. doi: 10.1038/nature09016.

Dahlin P, Srivastava V, Ekengren S, McKee LS, Bulone V. Comparative analysis of sterol acquisition in the oomycetes Saprolegnia parasitica and Phytophthora infestans. PLOS ONE. 2017; 12(2):e0170873. doi: 10.1371/journal.pone.0170873.

Derevnina L, Petre B, Kellner R, Dagdas YF, Sarowar MN, Giannakopoulou A, Concepcion JCDl, Chaparro-Garcia A, Pennington HG, West Pv, Kamoun S. Emerging oomycete threats to plants and animals. Philosophical Transactions of the Royal Society B: Biological Sciences. 2016; 371(1709):20150459. doi: 10.1098/rstb.2015.0459.

Faure C, Veyssière M, Boëlle B, San Clemente H, Bouchez O, Lopez-Roques C, Chaubet A, Martinez Y, Bezouška K, Suchánek M, et al. Long-Read Genome Sequence of the Sugar Beet Rhizosphere Mycoparasite Pythium oligandrum. G3: Genes, Genomes, Genetics. 2020; 10(2):431–436.

Fletcher K, Gil J, Bertier LD, Kenefick A, Wood KJ, Zhang L, Reyes-Chin-Wo S, Cavanaugh K, Tsuchida C, Wong J, et al. Genomic signatures of somatic hybrid vigor due to heterokaryosis in the oomycete pathogen, Bremia lactucae. bioRxiv. 2019; p. 516526.

Gaulin E, Pel MJC, Camborde L, San-Clemente H, Courbier S, Dupouy MA, Lengellé J, Veyssiere M, Ru AL, Grandjean F, Cordaux R, Moumen B, Gilbert C, Cano LM, Aury JM, Guy J, Wincker P, Bouchez O, Klopp C, Dumas B. Genomics analysis of Aphanomyces spp. identifies a new class of oomycete effector associated with host adaptation. BMC Biology. 2018; 16(1):43. doi: 10.1186/s12915-018-0508-5.

Goluch T, Bogdanowicz D, Giaro K. Visual TreeCmp: Comprehensive Comparison of Phylogenetic Trees on the Web. Methods in Ecology and Evolution. 2020; 11(4):494–499. doi: 10.1111/2041-210x.13358.

Haas BJ, Kamoun S, Zody MC, Jiang RHY, Handsaker RE, Cano LM, Grabherr M, Kodira CD, Raffaele S, Torto-Alalibo T, Bozkurt TO, Ah-Fong AMV, Alvarado L, Anderson VL, Armstrong MR, Avrova A, Baxter L, Beynon J, Boevink PC, Bollmann SR, et al. Genome sequence and analysis of the Irish potato famine pathogen Phytophthora infestans. Nature. 2009; 461(7262):393–398. doi: 10.1038/nature08358.

Hallen HE, Huebner M, Shiu SH, Güldener U, Trail F. Gene expression shifts during perithecium development in Gibberella zeae (anamorph Fusarium graminearum), with particular emphasis on ion transport proteins. Fungal Genetics and Biology. 2007; 44(11):1146–1156. doi: 10.1016/j.fgb.2007.04.007.

Hane JK, Paxman J, Jones DAB, Oliver RP, Wit Pd. "CATAStrophy," a Genome-Informed Trophic Classification of Filamentous Plant Pathogens – How Many Different Types of Filamentous Plant Pathogens Are There? Frontiers in Microbiology. 2020; 10:3088. doi: 10.3389/fmicb.2019.03088.

Huennekens FM. Folic Acid Coenzymes in the Biosynthesis of Purines and Pyrimidines. Vitamins & Hormones. 1969; 26:375–394. doi: 10.1016/s0083-6729(08)60762-1.

Hughes AL, Packer B, Welch R, Bergen AW, Chanock SJ, Yeager M. Widespread purifying selection at polymorphic sites in human protein-coding loci. Proceedings of the National Academy of Sciences. 2003; 100(26):15754–15757. doi: 10.1073/pnas.2536718100.

Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L, Finn RD, Gough J, Haft D, Hulo N, Kahn D, Kelly E, Laugraud A, Letunic I, Lonsdale D, Lopez R, et al. InterPro: the integrative protein signature database. Nucleic Acids Research. 2009; 37(suppl_1):D211–D215. doi: 10.1093/nar/gkn785.

Idänheimo N, Gauthier A, Salojärvi J, Siligato R, Brosché M, Kollist H, Mähönen AP, Kangasjärvi J, Wrzaczek M. The Arabidopsis thaliana cysteine-rich receptor-like kinases CRK6 and CRK7 protect against apoplastic oxidative stress. Biochemical and Biophysical Research Communications. 2014; 445(2):457–462. doi: 10.1016/j.bbrc.2014.02.013.

Jiang RHY, Bruijn Id, Haas BJ, Belmonte R, Löbach L, Christie J, Ackerveken Gvd, Bottin A, Bulone V, Díaz-Moreno SM, Dumas B, Fan L, Gaulin E, Govers F, Grenville-Briggs LJ, Horner NR, Levin JZ, Mammella M, Meijer HJG, Morris P, et al. Distinctive Expansion of Potential Virulence Genes in the Genome of the Oomycete Fish Pathogen Saprolegnia parasitica. PLoS Genetics. 2013; 9(6):e1003272. doi: 10.1371/journal.pgen.1003272.

Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014; 30(9):1236–1240. doi: 10.1093/bioinformatics/btu031.

Judelson HS. Dynamics and Innovations within Oomycete Genomes: Insights into Biology, Pathology, and Evolution. Eukaryotic Cell. 2012; 11(11):1304–1312. doi: 10.1128/ec.00155-12.

Judelson HS. Metabolic Diversity and Novelties in the Oomycetes. Annual Review of Microbiology. 2016; 71(1):1–19. doi: 10.1146/annurev-micro-090816-093609.

Katoh K, Asimenos G, Toh H. Multiple alignment of DNA sequences with MAFFT. Methods in molecular biology (Clifton, NJ). 2009; 537:39–64. doi: 10.1007/978-1-59745-251-9_3.

Kemen E, Gardiner A, Schultz-Larsen T, Kemen AC, Balmuth AL, Robert-Seilaniantz A, Bailey K, Holub E, Studholme DJ, MacLean D, Jones JD. Gene Gain and Loss during Evolution of Obligate Parasitism in the White Rust Pathogen of Arabidopsis thaliana. PLoS Biology. 2011; 9(7):e1001094. http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001094, doi: 10.1371/journal.pbio.1001094.

Kemen E, Jones JDG. Obligate biotroph parasitism: can we link genomes to lifestyles? Trends in Plant Science. 2012; 17(8):448–457. doi: 10.1016/j.tplants.2012.04.005.

Klarzynski O, Plesse B, Joubert JM, Yvin JC, Kopp M, Kloareg B, Fritig B. Linear β-1,3 Glucans Are Elicitors of Defense Responses in Tobacco. Plant Physiology. 2000; 124(3):1027–1038. doi: 10.1104/pp.124.3.1027.

Klopfenstein DV, Zhang L, Pedersen BS, Ramírez F, Vesztrocy AW, Naldi A, Mungall CJ, Yunes JM, Botvinnik O, Weigel M, Dampier W, Dessimoz C, Flick P, Tang H. GOATOOLS: A Python library for Gene Ontology analyses. Scientific Reports. 2018; 8(1):10872. doi: 10.1038/s41598-018-28948-z.

Kosiol C, Vinař T, Fonseca RRd, Hubisz MJ, Bustamante CD, Nielsen R, Siepel A. Patterns of Positive Selection in Six Mammalian Genomes. PLoS Genetics. 2008; 4(8):e1000144. doi: 10.1371/journal.pgen.1000144.

Köster J, Rahmann S. Snakemake—a scalable bioinformatics workflow engine. Bioinformatics. 2012; 28(19):2520–2522. doi: 10.1093/bioinformatics/bts480.

Lamour KH, Mudge J, Gobena D, Hurtado-Gonzales OP, Schmutz J, Kuo A, Miller NA, Rice BJ, Raffaele S, Cano LM, Bharti AK, Donahoo RS, Finley S, Huitema E, Hulvey J, Platt D, Salamov A, Savidor A, Sharma R, Stam R, et al. Genome Sequencing and Mapping Reveal Loss of Heterozygosity as a Mechanism for Rapid Adaptation in the Vegetable Pathogen Phytophthora capsici. Molecular Plant-Microbe Interactions®. 2012; 25(10):1350–1360. doi: 10.1094/mpmi-02-12-0028-r.

Latijnhouwers M, Wit PJGMd, Govers F. Oomycetes and fungi: similar weaponry to attack plants. Trends in Microbiology. 2003; 11(10):462–469. doi: 10.1016/j.tim.2003.08.002.

Lechner M, Findeiß S, Steiner L, Marz M, Stadler PF, Prohaska SJ. Proteinortho: Detection of (Co-)orthologs in large-scale analysis. BMC Bioinformatics. 2011; 12(1):124. doi: 10.1186/1471-2105-12-124.

Lee SJ, Rose JKC. Mediation of the transition from biotrophy to necrotrophy in hemibiotrophic plant pathogens by secreted effector proteins. Plant Signaling & Behavior. 2010; 5(6):769–772. doi: 10.4161/psb.5.6.11778.

LEWIS DH. CONCEPTS IN FUNGAL NUTRITION AND THE ORIGIN OF BIOTROPHY. Biological Reviews. 1973; 48(2):261–277. doi: 10.1111/j.1469-185x.1973.tb00982.x.

Li L, Jamieson K, DeSalvo G, Rostamizadeh A, Talwalkar A. Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization. Journal of Machine Learning Research. 2018; 18(185):1–52. http://jmlr.org/papers/v18/16-558.html.

Liu H, Ma X, Yu H, Fang D, Li Y, Wang X, Wang W, Dong Y, Xiao B. Genomes and virulence difference between two physiological races of Phytophthora nicotianae. GigaScience. 2016; 5(1):3. doi: 10.1186/s13742-016-0108-7.

Lommer M, Specht M, Roy AS, Kraemer L, Andreson R, Gutowska MA, Wolf J, Bergner SV, Schilhabel MB, Klostermeier UC, Beiko RG, Rosenstiel P, Hippler M, LaRoche J. Genome and low-iron response of an oceanic diatom adapted to chronic iron limitation. Genome Biology. 2012; 13(7):R66. doi: 10.1186/gb-2012-13-7-r66.

Lévesque CA, Brouwer H, Cano L, Hamilton JP, Holt C, Huitema E, Raffaele S, Robideau GP, Thines M, Win J, Zerillo MM, Beakes GW, Boore JL, Busam D, Dumas B, Ferriera S, Fuerstenberg SI, Gachon CM, Gaulin E, Govers F, et al. Genome sequence of the necrotrophic plant pathogen Pythium ultimum reveals original pathogenicity mechanisms and effector repertoire. Genome Biology. 2010; 11(7):R73. doi: 10.1186/gb-2010-11-7-r73.

Martens C, Peer YVd. The hidden duplication past of the plant pathogen Phytophthora and its consequences for infection. BMC Genomics. 2010; 11(1):353. doi: 10.1186/1471-2164-11-353.

Matari NH, Blair JE. A multilocus timescale for oomycete evolution estimated under three distinct molecular clock models. BMC Evolutionary Biology. 2014; 14(1):101. doi: 10.1186/1471-2148-14-101.

McCutcheon JP, Moran NA. Functional Convergence in Reduced Genomes of Bacterial Symbionts Spanning 200 My of Evolution. Genome Biology and Evolution. 2010; 2:708–718. doi: 10.1093/gbe/evq055.

McGowan J, O'Hanlon R, Owens RA, Fitzpatrick DA. Comparative Genomic and Proteomic Analyses of Three Widespread Phytophthora Species: Phytophthora chlamydospora, Phytophthora gonapodyides and Phytophthora pseudosyringae. Microorganisms. 2020; 8(5):653. doi: 10.3390/microorganisms8050653.

Misner I, Blouin N, Leonard G, Richards TA, Lane CE. The Secreted Proteins of Achlya hypogyna and Thraustotheca clavata Identify the Ancestral Oomycete Secretome and Reveal Gene Acquisitions by Horizontal Gene Transfer. Genome Biology and Evolution. 2015; 7(1):120–135. https://academic.oup.com/gbe/article/7/1/120/601821, doi: 10.1093/gbe/evu276.

Mock T. Extensive genetic diversity and differential bi-allelic expression in a Southern Ocean diatom. In: *EUROPEAN JOURNAL OF PHYCOLOGY*, vol. 50 TAYLOR & FRANCIS LTD 4 PARK SQUARE, MILTON PARK, ABINGDON OX14 4RN, OXON . . . ; 2015. p. 75–75.

Money NP, Davis CM, Ravishankar JP. Biomechanical evidence for convergent evolution of the invasive growth process among fungi and oomycete water molds. Fungal Genetics and Biology. 2004; 41(9):872–876. doi: 10.1016/j.fgb.2004.06.001.

Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Pond SLK, Scheffler K. FUBAR: A Fast, Unconstrained Bayesian AppRoximation for Inferring Selection. Molecular Biology and Evolution. 2013; 30(5):1196–1205. doi: 10.1093/molbev/mst030.

Ohan JA, Hovde BT, Zhang XL, Davenport KW, Chertkov O, Han C, Twary SN, Starkenburg SR. Nuclear Genome Assembly of the Microalga Nannochloropsis salina CCMP1776. Microbiology Resource Announcements. 2019; 8(44). doi: 10.1128/mra.00750-19.

O'Malley T, Bursztein E, Long J, Chollet F, Jin H, Invernizzi L, et al., Keras Tuner; 2019. https://github.com/keras-team/keras-tuner.

Petersen TN, Brunak S, Heijne Gv, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature Methods. 2011; 8(10):785–786. doi: 10.1038/nmeth.1701.

Price MN, Dehal PS, Arkin AP. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLoS ONE. 2010; 5(3):e9490. doi: 10.1371/journal.pone.0009490.

Props R, Monsieurs P, Vandamme P, Leys N, Denef VJ, Boon N. Gene Expansion and Positive Selection as Bacterial Adaptations to Oligotrophic Conditions. mSphere. 2019; 4(1). doi: 10.1128/mspheredirect.00011-19.

Raffaele S, Win J, Cano LM, Kamoun S. Analyses of genome architecture and gene expression reveal novel candidate virulence factors in the secretome of Phytophthora infestans. BMC Genomics. 2010; 11(1):637. doi: 10.1186/1471-2164-11-637.

Richards TA, Dacks JB, Jenkinson JM, Thornton CR, Talbot NJ. Evolution of Filamentous Plant Pathogens: Gene Exchange across Eukaryotic Kingdoms. Current Biology. 2006; 16(18):1857–1864. doi: 10.1016/j.cub.2006.07.052.

Richardson LJ, Rawlings ND, Salazar GA, Almeida A, Haft DR, Ducq G, Sutton GG, Finn RD. Genome properties in 2019: a new companion database to InterPro for the inference of complete functional attributes. Nucleic Acids Research. 2018; 47(D1):D564–D572. doi: 10.1093/nar/gky1013.

Robinson DF, Foulds LR. Comparison of phylogenetic trees. Mathematical Biosciences. 1981; 53(1-2):131–147. doi: 10.1016/0025-5564(81)90043-2.

Rocafort M, Fudal I, Mesarich CH. Apoplastic effector proteins of plant-associated fungi and oomycetes. Current Opinion in Plant Biology. 2020; 56:9–19. doi: 10.1016/j.pbi.2020.02.004.

Rodenburg SYA, Ridder Dd, Govers F, Seidl MF. Oomycete metabolism is highly dynamic and reflects lifestyle adaptations. bioRxiv. 2020; p. 2020.02.12.941195. doi: 10.1101/2020.02.12.941195.

Rujirawat T, Patumcharoenpol P, Lohnoo T, Yingyong W, Lerksuthirat T, Tangphatsornruang S, Suriyaphol P, Grenville-Briggs LJ, Garg G, Kittichotirat W, Krajaejun T. Draft Genome Sequence of the Pathogenic Oomycete Pythium insidiosum Strain Pi-S, Isolated from a Patient with Pythiosis. Genome Announcements. 2015; 3(3):e00574–15. doi: 10.1128/genomea.00574-15.

Samarasinghe H, You M, Jenkinson TS, Xu J, James TY. Hybridization Facilitates Adaptive Evolution in Two Major Fungal Pathogens. Genes. 2020; 11(1):101. doi: 10.3390/genes11010101.

Savory F, Leonard G, Richards TA. The Role of Horizontal Gene Transfer in the Evolution of the Oomycetes. PLOS Pathogens. 2015; 11(5):e1004805. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4447370/, doi: 10.1371/journal.ppat.1004805.

Scannell DR, Byrne KP, Gordon JL, Wong S, Wolfe KH. Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. Nature. 2006; 440(7082):341–345. doi: 10.1038/nature04562.

Schwarz G, Mendel RR. MOLYBDENUM COFACTOR BIOSYNTHESIS AND MOLYBDENUM ENZYMES. Plant Biology. 2006; 57(1):623–647. doi: 10.1146/annurev.arplant.57.032905.105437.

Seddiki K, Godart F, Cigliano RA, Sanseverino W, Barakat M, Ortet P, Rébeillé F, Maréchal E, Cagnac O, Amato A. Sequencing, De Novo Assembly, and Annotation of the Complete Genome of a New Thraustochytrid Species, Strain CCAP_4062/3. Genome Announcements. 2018; 6(11):e01335–17. doi: 10.1128/genomea.01335-17.

Seppey M, Manni M, Zdobnov EM. Gene Prediction, Methods and Protocols. Methods in Molecular Biology. 2019; 1962:227–245. doi: 10.1007/978-1-4939-9173-0_14.

Shang Y, Xiao G, Zheng P, Cen K, Zhan S, Wang C. Divergent and Convergent Evolution of Fungal Pathogenicity. Genome Biology and Evolution. 2016; 8(5):1374–1387. doi: 10.1093/gbe/evw082.

Sheridan PO, Raguideau S, Quince C, Holden J, Zhang L, Gaze WH, Holden J, Mead A, Raguideau S, Quince C, Singer AC, Wellington EMH, Zhang L, Williams TA, Gubry-Rangin C. Gene duplication drives genome expansion in a major lineage of Thaumarchaeota. Nature Communications. 2020; 11(1):5494. doi: 10.1038/s41467-020-19132-x.

Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Pond SLK. Less Is More: An Adaptive Branch-Site Random Effects Model for Efficient Detection of Episodic Diversifying Selection. Molecular Biology and Evolution. 2015; 32(5):1342–1353. doi: 10.1093/molbev/msv022.

Steciow MM, Lara E, Paul C, Pillonel A, Belbahri L. Multiple barcode assessment within the Saprolegnia-Achlya clade (Saprolegniales, Oomycota, Straminipila) brings order in a neglected group of pathogens. IMA Fungus. 2014; 5(2):439–448. doi: 10.5598/imafungus.2014.05.02.08.

Stephenson SA, Green JR, Manners JM, Maclean DJ. Cloning and characterisation of glutamine synthetase from Colletotrichum gloeosporioides and demonstration of elevated expression during pathogenesis on Stylosanthes guianensis. Current Genetics. 1997; 31(5):447–454. doi: 10.1007/s002940050228.

Studholme DJ, Panda P, Stowasser ESV, González M, Hill R, Sambles C, Grant M, Williams NM, McDougal RL. Genome sequencing of oomycete isolates from Chile supports the New Zealand origin of Phytophthora kernoviae and makes available the first Nothophytophthora sp. genome: Comparative genomics of Chilean oomycete isolates. Molecular Plant Pathology. 2018; 20(3):423–431. doi: 10.1111/mpp.12765.

Studholme DJ, McDougal RL, Sambles C, Hansen E, Hardy G, Grant M, Ganley RJ, Williams NM. Genome sequences of six Phytophthora species associated with forests in New Zealand. Genomics Data. 2016; 7:54–56. doi: 10.1016/j.gdata.2015.11.015.

Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Research. 2006; 34(suppl_2):W609–W612. doi: 10.1093/nar/gkl315.

Svoboda J, Mrugała A, Kozubíková-Balcarová E, Petrusek A. Hosts and transmission of the crayfish plague pathogen Aphanomyces astaci : a review. Journal of Fish Diseases. 2016; 40(1):127–140. doi: 10.1111/jfd.12472.

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. Oil accumulation by the oleaginous diatom Fistulifera solaris as revealed by the genome and transcriptome. The Plant cell. 2015; 27(1):162–76. doi: 10.1105/tpc.114.135194.

Torres MA. ROS in biotic interactions. Physiologia Plantarum. 2010; 138(4):414–429. doi: 10.1111/j.1399-3054.2009.01326.x.

Torruella G, de Mendoza A, Grau-Bové X, Antó M, Chaplin M, del Campo J, Eme L, Pérez-Cordón G, Whipps C, Nichols K, Paley R, Roger A, Sitjà-Bobadilla A, Donachie S, Ruiz-Trillo I. Phylogenomics Reveals Convergent Evolution of Lifestyles in Close Relatives of Animals and Fungi. Current Biology. 2015; 25(18):2404–2410. doi: 10.1016/j.cub.2015.07.053.

Treangen TJ, Rocha EPC. Horizontal Transfer, Not Duplication, Drives the Expansion of Protein Families in Prokaryotes. PLoS Genetics. 2011; 7(1):e1001284. doi: 10.1371/journal.pgen.1001284.

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, et al. Phytophthora Genome Sequences Uncover Evolutionary Origins and Mechanisms of Pathogenesis. Science. 2006; 313(5791):1261–1266. doi: 10.1126/science.1128796.

Walley JW, Kliebenstein DJ, Bostock RM, Dehesh K. Fatty acids and early detection of pathogens. Current Opinion in Plant Biology. 2013; 16(4):520–526. doi: 10.1016/j.pbi.2013.06.011.

Wang E, Schornack S, Marsh J, Gobbato E, Schwessinger B, Eastmond P, Schultze M, Kamoun S, Oldroyd GD. A Common Signaling Process that Promotes Mycorrhizal and Oomycete Colonization of Plants. Current Biology. 2012; 22(23):2242–2246. doi: 10.1016/j.cub.2012.09.043.

Wang Y, Tyler BM, Wang Y. Defense and Counterdefense During Plant-Pathogenic Oomycete Infection. Annual Review of Microbiology. 2019; 73(1):1–30. doi: 10.1146/annurev-micro-020518-120022.

Weiblen C, Robe LJ, Azevedo MId, Ianiski LB, Stibbe PC, Ribeiro TC, Zanette RA, Pereira DIB, Santurio JM, Botton SdA. New insights on evolutionary aspects of Pythium insidiosum and other peronosporaleans. Mycoses. 2020; doi: 10.1111/myc.13057.

Yang Z. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Molecular Biology and Evolution. 1998 05; 15(5):568–573. https://doi.org/10.1093/oxfordjournals.molbev.a025957, doi: 10.1093/oxfordjournals.molbev.a025957.

Yashiro T, Yura S, Tobita A, Toyoda Y, Kasakura K, Nishiyama C. Pterostilbene reduces colonic inflammation by suppressing dendritic cell activation and promoting regulatory T cell development. The FASEB Journal. 2020; 34(11):14810–14819. doi: 10.1096/fj.202001502r.

Zhang J, Rosenberg HF, Nei M. Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proceedings of the National Academy of Sciences. 1998; 95(7):3708–3713. doi: 10.1073/pnas.95.7.3708.

Appendix 0 Table 1. Stramenopile genomes dataset used for positive selective analyses.

Phylogenetic group	Species name	Accession	Lifestyle	Complete BUSCOs	Complete and single-copy BUSCOs	Complete and duplicated BUSCOs	Reference
Non-Oomycete	Ectocarpus siliculosus	GCA_000310025.1	Autotroph	97	97	0	*Cock et al.* (*2010*)
	Fistulifera solaris	GCA_002217885.1	Autotroph	97	14	83	*Tanaka et al.* (*2015*)
	Fragilariopsis cylindrus	GCA_001750085.1	Autotroph	95	95	0	*Mock* (*2015*)
	Hondaea fermentalgiana	GCA_002897355.1	Autotroph	95	90	5	*Seddiki et al.* (*2018*)
	Microchloropsis salina	GCA_004565275.1	Autotroph	92	90	2	*Ohan et al.* (*2019*)
	Phaeodactylum tricornutum	GCA_000150955.2	Autotroph	97	95	2	*Bowler et al.* (*2008*)
	Thalassiosira oceanica	GCA_000296195.2	Autotroph	90	90	0	*Lommer et al.* (*2012*)
	Thalassiosira pseudonana	GCA_000149405.2	Autotroph	97	95	2	*Bowler et al.* (*2008*)
Saprolegniales	Achlya hypogyna	GCA_002081595.1	Animal necrotroph	99	98	1	*Misner et al.* (*2015*)
	Aphanomyces astaci	GCA_000520075.1	Animal necrotroph	100	82	18	
	Aphanomyces invadans	GCA_000520115.1	Animal necrotroph	100	83	17	
	Saprolegnia diclina	GCA_000281045.1	Animal necrotroph	99	98	1	
	Saprolegnia parasitica	GCA_000151545.2	Animal necrotroph	99	99	0	*Jiang et al.* (*2013*)
	Thraustotheca clavata	GCA_002081575.1	Free-living saprotroph	99	98	1	*Misner et al.* (*2015*)
Albuginales	Albugo candida	GCA_001078535.1	Obligate biotroph	98	86	12	
	Albugo laibachii	PRJEA53219	Obligate biotroph	95	82	13	*Kemen et al.* (*2011*)
Peronosporales	Bremia lactucae	GCA_004359215.1	Obligate biotroph	96	90	6	*Fletcher et al.* (*2019*)
	Hyaloperonospora arabidopsidis	GCA_000173235.2	Obligate biotroph	89	82	7	*Baxter et al.* (*2010*)
	Nothophytophthora sp.	GCA_001712635.2	Hemibiotroph	90	28	62	*Studholme et al.* (*2018*)
	Peronospora effusa	GCA_003843895.1	Obligate biotroph	94	93	1	
	Phytophthora cactorum	GCA_003287315.1	Hemibiotroph	100	98	2	*Armitage et al.* (*2018*)
	Phytophthora capsici	GCA_000325885.1	Hemibiotroph	98	97	1	*Lamour et al.* (*2012*)
	Phytophthora cinnamomi	GCA_001314365.1	Hemibiotroph	96	94	2	*Studholme et al.* (*2016*)
	Phytophthora infestans	GCA_000142945.1	Hemibiotroph	100	99	1	
	Phytophthora kernoviae	GCA_001712645.2	Hemibiotroph	96	96	0	*Studholme et al.* (*2018*)
	Phytophthora megakarya	GCA_002215365.1	Hemibiotroph	91	90	1	*Ali et al.* (*2017*)
	Phytophthora nicotianae	GCA_001483015.1	Hemibiotroph	99	86	13	*Liu et al.* (*2016*)
	Phytophthora parasitica	GCA_000247585.2	Hemibiotroph	98	87	11	
	Phytophthora sojae	GCA_000149755.2	Hemibiotroph	99	98	1	*Tyler et al.* (*2006*)
	Plasmopara halstedii	GCA_900000015.1	Obligate biotroph	100	100	0	
Pythiales	Pythium aphanidermatum	GCA_000387445.2	Plant necrotroph	94	93	1	*Adhikari et al.* (*2013*)
	Pythium insidiosum	GCA_001029375.1	Animal necrotroph	99	87	12	*Rujirawat et al.* (*2015*)
	Pythium_irregulare	GCA_000387425.2	Plant necrotroph	98	96	2	*Lévesque et al.* (*2010*)
	Pythium_oligandrum	GCA_005966545.1	Fungus necrotroph	100	100	0	*Faure et al.* (*2020*)
	Pythium_ultimum	GCA_000143045.1	Plant necrotroph	94	93	1	*Lévesque et al.* (*2010*)
	Pythium_vexans	GCA_000387545.2	Plant necrotroph	94	92	2	*Adhikari et al.* (*2013*)

Zhang W, Zhang X, Li K, Wang C, Cai L, Zhuang W, Xiang M, Liu X. Introgression and gene family contraction drive the evolution of lifestyle and host shifts of hypocrealean fungi. Mycology. 2018; 9(3):1–13. doi: 10.1080/21501203.2018.1478333.

Zheng L, Mackrill JJ. Calcium Signaling in Oomycetes: An Evolutionary Perspective. Frontiers in Physiology. 2016; 7:123. doi: 10.3389/fphys.2016.00123.

Appendix 0 Table 2. Summary of basidiomycete dataset.

Species name	Plant pathogen	Accession
Acaromyces ingoldii	no	GCA_003144295.1
Anthracocystis flocculosa	no	GCA_000417875.1
Apiotrichum porosum	no	GCA_003942205.1
Ceraceosorus bombacis	yes	GCA_900000165.1
Ceraceosorus guamensis	no	GCA_003144195.1
Ceratobasidium theobromae	yes	GCA_009078325.1
Cryptococcus amylolentus	no	GCA_001720205.1
Cryptococcus gattii	no	GCA_000855695.1
Cryptococcus neoformans	no	GCA_000149245.3
Cryptococcus wingfieldii	no	GCA_001720155.1
Cutaneotrichosporon oleaginosum	no	GCA_001027345.1
Fomitiporia mediterranea	yes	GCA_000271605.1
Jaapia argillacea	no	GCA_000697665.1
Jaminaea rosea	no	GCA_003144245.1
Kalmanozyma brasiliensis	no	GCA_000497045.1
Kockovaella imperatae	no	GCA_002102565.1
Kwoniella bestiolae	no	GCA_000512585.2
Kwoniella dejecticola	no	GCA_000512565.2
Kwoniella pini	no	GCA_000512605.2
Leucosporidium creatinivorum	no	GCA_002105055.1
Malassezia globosa	no	GCA_000181695.1
Malassezia restricta	no	GCA_003290485.1
Malassezia sympodialis	no	GCA_000349305.2
Meira miltonrushii	no	GCA_003144205.1
Melampsora larici-populina	yes	GCA_000204055.1
Microbotryum lychnidis-dioicae	yes	GCA_000166175.1
Mixia osmundae	yes	GCA_000708205.1
Moesziomyces antarcticus	no	GCA_000747765.1
Moesziomyces aphidis	no	GCA_000517465.1
Moniliophthora roreri	yes	GCA_001466705.1
Paxillus involutus	no	GCA_000827475.1
Peniophora sp	no	GCA_900536885.1
Piloderma croceum	no	GCA_000827315.1
Pseudomicrostroma glucosiphilum	no	GCA_003144135.1
Pseudozyma hubeiensis	no	GCA_000403515.1
Puccinia coronata	yes	GCA_002873125.1
Puccinia graminis	yes	GCA_000149925.1
Puccinia sorghi	yes	GCA_001263375.1
Puccinia striiformis	yes	GCA_002920065.1
Puccinia triticina	yes	GCA_000151525.2
Rhizoctonia solani	yes	GCA_000524645.1
Rhodotorula graminis	no	GCA_001329695.1
Rhodotorula toruloides	no	GCA_000320785.2
Saitozyma podzolica	no	GCA_003942215.1
Serendipita indica	no	GCA_000313545.1
Serendipita vermifera	no	GCA_000827415.1
Sporisorium graminicola	no	GCA_005498985.1
Sporisorium reilianum	yes	GCA_900162835.1
Sporisorium scitamineum	yes	GCA_001243155.1
Testicularia cyperi	yes	GCA_003144125.1
Tilletia controversa	yes	GCA_001645045.2
Tilletia laevis	yes	GCA_009428275.1
Tilletia walkeri	yes	GCA_009428295.1
Tilletiaria anomala	yes	GCA_000711695.1
Tilletiopsis washingtonensis	yes	GCA_003144115.1
Trichosporon asahii	no	GCA_000293215.1
Ustilago bromivora	yes	GCA_900080155.1
Ustilago hordei	yes	GCA_000286035.1
Ustilago maydis	yes	GCA_000328475.2
Ustilago trichophora	yes	GCA_900323505.1
Violaceomyces palustris	no	GCA_003144235.1
Wallemia hederae	no	GCA_004918325.1
Wallemia ichthyophaga	no	GCA_000400465.1
Wallemia mellicola	no	GCA_000263375.1
Xanthophyllomyces dendrorhous	no	GCA_001007165.2

Appendix 0 Table 3. Summary of genomes used for the lifestyle model construction.

Species name	Number of proteomes	Lifestyle
Agaricus bisporus	1	S
Albugo species	2	B
Alternaria species	14	N
Ascochyta rabiei	1	N
Aspergillus species	34	S
Bipolaris species	7	N/H
Blumeria graminis	4	B
Botrytis cinerea	3	N
Bremia lactucae	1	B
Colletotrichum species	14	H
Debaryomyces hansenii	1	S
Dothistroma septosporum	1	H
Erysiphe necator	1	B
Eutypa lata	1	N
Fusarium species	6	H
Gigaspora margarita	1	B
Globisporangium ultimum	1	N
Gloeophyllum trabeum	1	S
Hyaloperonospora arabidopsidis	1	B
Komagataella phaffii	5	S
Leptosphaeria maculans	1	H
Macrophomina phaseolina	1	H
Marssonina brunnea	1	H
Melampsora laricis-populina	1	B
Microbotryum violaceum	1	B
Monilinia laxa	1	N
Moniliophthora species	3	H
Neurospora crassa	2	S
Oidium neolycopersici	1	B
Parastagonospora nodorum	1	N
Peronospora effusa	2	B
Phytophthora species	34	H
Plasmodiophora brassicae	2	B
Plasmopara halstedii	1	B
Pleurotus ostreatu	1	S
Pseudocercospora fijiensis	1	H
Puccinia species	10	B
Pyrenophora species	18	N
Pyricularia oryzae	4	H
Ramularia collo-cygni	1	H
Rhizoctonia solani	7	N
Rhizopus delemar	1	S
Saccharomyces cerevisiae	131	S
Schizosaccharomyces pombe	1	S
Sclerotinia species	3	N
Serpula lacrymans	2	S
Setosphaeria turcica	1	H
Sphaerobolus stellatus	1	S
Sporisorium reilianum	2	B
Stereum hirsutum	1	S
Synchytrium endobioticum	2	B
Taphrina deformans	1	B
Thraustotheca clavata	1	S
Tilletia indica	3	H
Trametes versicolor	1	S
Tremella mesenterica	2	B
Trichoderma species	7	S
Uncinocarpus reesii	1	S
Ustilaginoidea virens	2	B
Ustilago species	2	B
Venturia inaequalis	4	H
Verticillium dahliae	10	H
Yarrowia lipolytica	13	S
Zymoseptoria species	6	H

S: saptroph, N: necrotroph, H: hemibiotroph, B: biotroph

Table 3–source data 1. https://github.com/danielzmbp/lspred/blob/main/genome_list.tsv

Appendix 0 Table 4. Significant GO terms with a depth higher than 7 found enriched in the positively selected proteins in plant fungal pathogens.

GO number	Name	Ratio in study	Ratio in population	Depth	-log10 of p value
GO:0009064	glutamine family amino acid metabolic process	140/13729	458/237259	8	57.33
GO:0006165	nucleoside diphosphate phosphorylation	99/13729	320/237259	8	40.37
GO:0006096	glycolytic process	80/13729	266/237259	12	31.2
GO:0006399	tRNA metabolic process	239/13729	1881/237259	8	25.41
GO:1901607	alpha-amino acid biosynthetic process	138/13729	830/237259	8	24.68
GO:0006525	arginine metabolic process	54/13729	157/237259	9	23.51
GO:0006546	glycine catabolic process	40/13729	86/237259	10	22.7
GO:0001510	RNA methylation	56/13729	211/237259	8	18.2
GO:0006750	glutathione biosynthetic process	29/13729	56/237259	8	17.46
GO:0034470	ncRNA processing	186/13729	1549/237259	8	16.55
GO:0008033	tRNA processing	130/13729	991/237259	9	13.87
GO:1901606	alpha-amino acid catabolic process	62/13729	359/237259	8	10.58
GO:0006418	tRNA aminoacylation for protein translation	109/13729	880/237259	10	9.59
GO:0009435	NAD biosynthetic process	34/13729	145/237259	11	8.46
GO:0016579	protein deubiquitination	59/13729	393/237259	9	7.36
GO:0009150	purine ribonucleotide metabolic process	119/13729	1092/237259	9	6.92
GO:0015693	magnesium ion transport	28/13729	123/237259	8	6.23
GO:0006633	fatty acid biosynthetic process	35/13729	219/237259	8	4.08
GO:0015031	protein transport	179/13729	2051/237259	8	3.95
GO:0006355	regulation of transcription, DNA-templated	337/13729	4376/237259	9	3.6
GO:0009165	nucleotide biosynthetic process	123/13729	1368/237259	8	2.48
GO:0006605	protein targeting	38/13729	288/237259	10	2.38
GO:0001522	pseudouridine synthesis	35/13729	260/237259	8	2.28
GO:0006364	rRNA processing	57/13729	515/237259	9	2.22
GO:0006511	ubiquitin-dependent protein catabolic process	86/13729	934/237259	8	1.37

Appendix 0 Table 5. Significant enriched terms relating to biological processes in the positively selected obligate biotroph proteins.

GO number	Name	Ratio in study	Ratio in population	Depth	-log10 of p value
GO:0051274	beta-glucan biosynthetic process	15/2417	93/75580	8	3.28
GO:0051273	beta-glucan metabolic process	17/2417	124/75580	7	3.02
GO:0016579	protein deubiquitination	23/2417	199/75580	9	2.99
GO:0071103	DNA conformation change	14/2417	75/75580	6	2.86
GO:0016051	carbohydrate biosynthetic process	28/2417	203/75580	4	2.85
GO:0070647	protein modification by small protein conjugation or removal	37/2417	359/75580	7	2.83
GO:0006511	ubiquitin-dependent protein catabolic process	32/2417	359/75580	8	2.83
GO:0030001	metal ion transport	32/2417	344/75580	6	2.75
GO:0022402	cell cycle process	22/2417	206/75580	2	2.74
GO:0044262	cellular carbohydrate metabolic process	28/2417	256/75580	4	2.74
GO:0034637	cellular carbohydrate biosynthetic process	22/2417	160/75580	5	2.7
GO:0007018	microtubule-based movement	33/2417	402/75580	3	2.69
GO:0006812	cation transport	46/2417	597/75580	5	2.48
GO:0055085	transmembrane transport	88/2417	1402/75580	4	2.33
GO:0006996	organelle organization	47/2417	588/75580	4	2.31
GO:0005975	carbohydrate metabolic process	56/2417	897/75580	3	2.2
GO:0051603	proteolysis involved in cellular protein catabolic process	34/2417	413/75580	6	2.19
GO:1901576	organic substance biosynthetic process	167/2417	3213/75580	3	2.15
GO:0051179	localization	159/2417	3144/75580	1	2.14
GO:0006810	transport	156/2417	3083/75580	3	2.12
GO:0051234	establishment of localization	157/2417	3094/75580	2	2.11
GO:0044249	cellular biosynthetic process	154/2417	3097/75580	3	2.1
GO:0044267	cellular protein metabolic process	177/2417	3711/75580	5	2.1
GO:0009058	biosynthetic process	171/2417	3416/75580	2	2.07
GO:0043412	macromolecule modification	172/2417	3359/75580	4	2.05
GO:0006464	cellular protein modification process	153/2417	2997/75580	6	2.04
GO:0019538	protein metabolic process	220/2417	4882/75580	4	1.99
GO:0044260	cellular macromolecule metabolic process	280/2417	5736/75580	4	1.96
GO:0048523	negative regulation of cellular process	20/2417	196/75580	4	1.96
GO:0043170	macromolecule metabolic process	381/2417	8302/75580	3	1.89
GO:0044238	primary metabolic process	489/2417	10774/75580	2	1.88
GO:0006807	nitrogen compound metabolic process	422/2417	9685/75580	2	1.86
GO:1901360	organic cyclic compound metabolic process	192/2417	4279/75580	3	1.85
GO:0006793	phosphorus metabolic process	146/2417	3105/75580	3	1.77
GO:0006725	cellular aromatic compound metabolic process	188/2417	4193/75580	3	1.76
GO:0006811	ion transport	93/2417	1806/75580	4	1.74
GO:1901564	organonitrogen compound metabolic process	279/2417	6661/75580	3	1.73
GO:0006078	(1->6)-beta-D-glucan biosynthetic process	9/2417	46/75580	9	1.61
GO:0006796	phosphate-containing compound metabolic process	144/2417	3088/75580	4	1.59
GO:0034645	cellular macromolecule biosynthetic process	75/2417	1388/75580	5	1.58
GO:0046483	heterocycle metabolic process	185/2417	4207/75580	3	1.48
GO:1903047	mitotic cell cycle process	15/2417	130/75580	3	1.44
GO:0016310	phosphorylation	93/2417	1835/75580	5	1.41
GO:0019438	aromatic compound biosynthetic process	63/2417	1122/75580	4	1.36
GO:0048519	negative regulation of biological process	25/2417	308/75580	3	1.34
GO:0071840	cellular component organization or biogenesis	78/2417	1472/75580	2	1.34
GO:0006139	nucleobase-containing compound metabolic process	173/2417	3927/75580	4	1.31

Appendix 0 Table 6. Significant enriched terms relating to biological processes in the positively selected *Phytophthora* (and *P. vexans*) proteins.

GO number	Name	Ratio in study	Ratio in population	Depth	-log10 of p value
GO:0006284	base-excision repair	11/5162	41/157599	8	2.9
GO:0015698	inorganic anion transport	26/5162	262/157599	6	2.87
GO:0070838	divalent metal ion transport	23/5162	200/157599	7	2.77
GO:0006813	potassium ion transport	45/5162	300/157599	7	2.75
GO:0044262	cellular carbohydrate metabolic process	45/5162	536/157599	4	2.68
GO:0034637	cellular carbohydrate biosynthetic process	32/5162	356/157599	5	2.67
GO:0030001	metal ion transport	76/5162	913/157599	6	2.57
GO:0016051	carbohydrate biosynthetic process	41/5162	449/157599	4	2.55
GO:0015672	monovalent inorganic cation transport	48/5162	667/157599	6	2.48
GO:0005976	polysaccharide metabolic process	34/5162	414/157599	4	2.34
GO:0051273	beta-glucan metabolic process	28/5162	315/157599	7	2.33
GO:0006812	cation transport	88/5162	1364/157599	5	2.28
GO:0006811	ion transport	195/5162	3485/157599	4	2.05
GO:0005975	carbohydrate metabolic process	149/5162	2464/157599	3	2.05
GO:0051179	localization	321/5162	7164/157599	1	1.92
GO:0006810	transport	318/5162	7068/157599	3	1.91
GO:0051234	establishment of localization	319/5162	7086/157599	2	1.9
GO:0055085	transmembrane transport	200/5162	4363/157599	4	1.87
GO:0000097	sulfur amino acid biosynthetic process	14/5162	105/157599	5	1.75
GO:0050794	regulation of cellular process	156/5162	3286/157599	3	1.72
GO:0044238	primary metabolic process	806/5162	20276/157599	2	1.71
GO:0006950	response to stress	74/5162	1321/157599	2	1.63
GO:0050896	response to stimulus	77/5162	1394/157599	1	1.56
GO:0050789	regulation of biological process	163/5162	3513/157599	2	1.53

Appendix 0 Table 7. Significant enriched terms relating to biological processes in the positively selected *Pythium* proteins.

GO number	Name	Ratio in study	Ratio in population	Depth	-log10 of p value
GO:0006741	NADP biosynthetic process	8/4619	8/56393	11	3.44
GO:0006101	citrate metabolic process	8/4619	8/56393	7	3.44
GO:0006075	(1->3)-beta-D-glucan biosynthetic process	16/4619	30/56393	9	3.34
GO:0043547	positive regulation of GTPase activity	8/4619	11/56393	6	3.28
GO:0051336	regulation of hydrolase activity	21/4619	68/56393	4	3.22
GO:0006813	potassium ion transport	40/4619	138/56393	7	2.88
GO:0009073	aromatic amino acid family biosynthetic process	16/4619	48/56393	5	2.86
GO:0006564	L-serine biosynthetic process	8/4619	12/56393	10	2.84
GO:0006851	mitochondrial calcium ion transmembrane transport	7/4619	9/56393	10	2.82
GO:0046488	phosphatidylinositol metabolic process	43/4619	206/56393	7	2.65
GO:0090481	pyrimidine nucleotide-sugar transmembrane transport	12/4619	29/56393	9	2.61
GO:0044262	cellular carbohydrate metabolic process	41/4619	218/56393	4	2.57
GO:0034637	cellular carbohydrate biosynthetic process	33/4619	143/56393	5	2.54
GO:0016051	carbohydrate biosynthetic process	40/4619	180/56393	4	2.51
GO:0006650	glycerophospholipid metabolic process	51/4619	233/56393	6	2.51
GO:0048519	negative regulation of biological process	42/4619	211/56393	3	2.5
GO:0006144	purine nucleobase metabolic process	11/4619	25/56393	7	2.49
GO:0030001	metal ion transport	63/4619	377/56393	6	2.47
GO:0045116	protein neddylation	6/4619	7/56393	9	2.42
GO:0006996	organelle organization	71/4619	398/56393	4	2.38
GO:0006644	phospholipid metabolic process	52/4619	302/56393	5	2.34
GO:0015672	monovalent inorganic cation transport	48/4619	276/56393	6	2.34
GO:0006629	lipid metabolic process	113/4619	755/56393	3	2.32
GO:0044283	small molecule biosynthetic process	85/4619	536/56393	3	2.31
GO:0044782	cilium organization	18/4619	65/56393	6	2.22
GO:0071840	cellular component organization or biogenesis	128/4619	984/56393	2	2.21
GO:0006396	RNA processing	109/4619	797/56393	7	2.19
GO:0019438	aromatic compound biosynthetic process	120/4619	799/56393	4	2.18
GO:1901362	organic cyclic compound biosynthetic process	126/4619	893/56393	4	2.16
GO:0016043	cellular component organization	127/4619	919/56393	3	2.16
GO:0018130	heterocycle biosynthetic process	112/4619	850/56393	4	2.15
GO:0034645	cellular macromolecule biosynthetic process	119/4619	880/56393	5	2.11
GO:1901566	organonitrogen compound biosynthetic process	148/4619	1138/56393	4	2.1
GO:0044271	cellular nitrogen compound biosynthetic process	157/4619	1288/56393	4	2.07
GO:1901137	carbohydrate derivative biosynthetic process	45/4619	263/56393	4	2.06
GO:0009059	macromolecule biosynthetic process	129/4619	989/56393	4	2.06
GO:0090304	nucleic acid metabolic process	237/4619	1993/56393	5	2.06
GO:0044255	cellular lipid metabolic process	76/4619	529/56393	4	2.06
GO:0016053	organic acid biosynthetic process	53/4619	335/56393	4	2.03
GO:0009058	biosynthetic process	316/4619	2375/56393	2	2.0
GO:0006725	cellular aromatic compound metabolic process	362/4619	2819/56393	3	1.99
GO:0016070	RNA metabolic process	169/4619	1403/56393	6	1.99
GO:1901576	organic substance biosynthetic process	308/4619	2221/56393	3	1.98
GO:1901360	organic cyclic compound metabolic process	369/4619	2872/56393	3	1.98
GO:0034654	nucleobase-containing compound biosynthetic process	86/4619	636/56393	5	1.98
GO:0050790	regulation of catalytic activity	25/4619	115/56393	3	1.97
GO:0019637	organophosphate metabolic process	104/4619	790/56393	4	1.96
GO:0044249	cellular biosynthetic process	282/4619	2116/56393	3	1.96
GO:0007017	microtubule-based process	76/4619	540/56393	2	1.95
GO:0034641	cellular nitrogen compound metabolic process	397/4619	3286/56393	3	1.95
GO:0046483	heterocycle metabolic process	350/4619	2804/56393	3	1.94
GO:0006139	nucleobase-containing compound metabolic process	321/4619	2584/56393	4	1.94
GO:0009113	purine nucleobase biosynthetic process	7/4619	11/56393	8	1.94
GO:0034404	nucleobase-containing small molecule biosynthetic process	20/4619	81/56393	6	1.93
GO:0048523	negative regulation of cellular process	28/4619	138/56393	4	1.91
GO:0044260	cellular macromolecule metabolic process	422/4619	3984/56393	4	1.89
GO:0050789	regulation of biological process	189/4619	1639/56393	2	1.87
GO:0006627	protein processing involved in protein targeting to mitochondrion	6/4619	8/56393	8	1.86
GO:0046474	glycerophospholipid biosynthetic process	19/4619	76/56393	7	1.78
GO:0043170	macromolecule metabolic process	615/4619	6203/56393	3	1.78
GO:0044237	cellular metabolic process	839/4619	7324/56393	2	1.75
GO:0006807	nitrogen compound metabolic process	756/4619	7339/56393	2	1.75
GO:0006812	cation transport	78/4619	566/56393	5	1.73
GO:0070647	protein modification by small protein conjugation or removal	43/4619	258/56393	7	1.72
GO:0044238	primary metabolic process	867/4619	8282/56393	2	1.7
GO:0071704	organic substance metabolic process	939/4619	8804/56393	2	1.69
GO:0008152	metabolic process	1090/4619	10893/56393	1	1.68
GO:0046854	phosphatidylinositol phosphorylation	16/4619	58/56393	8	1.66
GO:0051274	beta-glucan biosynthetic process	21/4619	91/56393	8	1.66
GO:0007018	microtubule-based movement	61/4619	418/56393	3	1.59
GO:0046129	purine ribonucleoside biosynthetic process	11/4619	30/56393	9	1.58
GO:0070925	organelle assembly	22/4619	100/56393	5	1.52
GO:0001522	pseudouridine synthesis	20/4619	86/56393	8	1.51
GO:0050794	regulation of cellular process	174/4619	1533/56393	3	1.51
GO:0065007	biological regulation	195/4619	1755/56393	1	1.49
GO:0009311	oligosaccharide metabolic process	18/4619	73/56393	4	1.47
GO:0009451	RNA modification	39/4619	233/56393	7	1.44
GO:0008033	tRNA processing	35/4619	199/56393	9	1.4
GO:0044248	cellular catabolic process	84/4619	641/56393	3	1.36
GO:0009116	nucleoside metabolic process	20/4619	88/56393	6	1.36

Appendix 0 Table 8. Significant enriched terms relating to biological processes in the positively selected Saprolegniales proteins.

GO number	Name	Ratio in study	Ratio in population	Depth	-log10 of p value
GO:1905775	negative regulation of DNA helicase activity	6/5995	6/113005	9	4.35
GO:0010390	histone monoubiquitination	6/5995	6/113005	11	4.35
GO:0000463	maturation of LSU-rRNA from tricistronic rRNA transcript	6/5995	6/113005	11	4.35
GO:0009088	threonine biosynthetic process	6/5995	6/113005	10	4.35
GO:0015689	molybdate ion transport	12/5995	13/113005	7	3.9
GO:0072350	tricarboxylic acid metabolic process	12/5995	23/113005	6	3.68
GO:0090090	negative regulation of canonical Wnt signaling pathway	6/5995	7/113005	8	3.53
GO:0071596	ubiquitin-dependent protein catabolic process via the N-end rule pathway	15/5995	36/113005	10	3.43
GO:0006094	gluconeogenesis	11/5995	30/113005	7	3.41
GO:0009396	folic acid-containing compound biosynthetic process	18/5995	50/113005	6	3.38
GO:0090481	pyrimidine nucleotide-sugar transmembrane transport	14/5995	43/113005	9	3.38
GO:0009311	oligosaccharide metabolic process	20/5995	83/113005	4	3.24
GO:0006101	citrate metabolic process	11/5995	11/113005	7	3.2
GO:1905515	non-motile cilium assembly	12/5995	20/113005	8	3.19
GO:0006488	dolichol-linked oligosaccharide biosynthetic process	17/5995	52/113005	6	3.15
GO:0046129	purine ribonucleoside biosynthetic process	14/5995	45/113005	9	3.12
GO:0009070	serine family amino acid biosynthetic process	17/5995	77/113005	9	3.03
GO:0030488	tRNA methylation	14/5995	46/113005	11	3.0
GO:0006741	NADP biosynthetic process	11/5995	21/113005	11	2.97
GO:0043547	positive regulation of GTPase activity	12/5995	16/113005	6	2.97
GO:0009113	purine nucleobase biosynthetic process	10/5995	16/113005	8	2.97
GO:0006360	transcription by RNA polymerase I	9/5995	16/113005	10	2.97
GO:0007186	G protein-coupled receptor signaling pathway	25/5995	114/113005	5	2.95
GO:0051056	regulation of small GTPase mediated signal transduction	21/5995	79/113005	7	2.87
GO:0006011	UDP-glucose metabolic process	10/5995	12/113005	7	2.84
GO:0006661	phosphatidylinositol biosynthetic process	26/5995	126/113005	8	2.8
GO:0046474	glycerophospholipid biosynthetic process	31/5995	167/113005	7	2.78
GO:0006506	GPI anchor biosynthetic process	24/5995	117/113005	9	2.76
GO:0043647	inositol phosphate metabolic process	12/5995	41/113005	6	2.75
GO:0009247	glycolipid biosynthetic process	30/5995	147/113005	7	2.74
GO:0048519	negative regulation of biological process	54/5995	385/113005	3	2.74
GO:0045116	protein neddylation	11/5995	17/113005	9	2.72
GO:0070925	organelle assembly	40/5995	223/113005	5	2.7
GO:0001522	pseudouridine synthesis	31/5995	148/113005	8	2.69
GO:0060271	cilium assembly	30/5995	128/113005	7	2.69
GO:0072522	purine-containing compound biosynthetic process	35/5995	213/113005	5	2.65
GO:0072330	monocarboxylic acid biosynthetic process	21/5995	100/113005	7	2.64
GO:0008654	phospholipid biosynthetic process	37/5995	263/113005	6	2.63
GO:0051603	proteolysis involved in cellular protein catabolic process	72/5995	528/113005	6	2.63
GO:0051336	regulation of hydrolase activity	26/5995	129/113005	4	2.63
GO:0030031	cell projection assembly	31/5995	129/113005	5	2.63
GO:0046488	phosphatidylinositol metabolic process	63/5995	449/113005	7	2.62
GO:0006400	tRNA modification	35/5995	225/113005	10	2.62
GO:0009152	purine ribonucleotide biosynthetic process	25/5995	160/113005	10	2.61
GO:0042398	cellular modified amino acid biosynthetic process	26/5995	110/113005	5	2.6
GO:0048523	negative regulation of cellular process	31/5995	226/113005	4	2.58
GO:0006310	DNA recombination	49/5995	295/113005	7	2.57
GO:0006643	membrane lipid metabolic process	35/5995	193/113005	5	2.56
GO:0006260	DNA replication	39/5995	270/113005	6	2.53
GO:0015672	monovalent inorganic cation transport	71/5995	681/113005	6	2.47
GO:0007166	cell surface receptor signaling pathway	15/5995	67/113005	5	2.46
GO:0016053	organic acid biosynthetic process	68/5995	589/113005	4	2.45
GO:0006511	ubiquitin-dependent protein catabolic process	66/5995	483/113005	8	2.45
GO:0044265	cellular macromolecule catabolic process	86/5995	710/113005	5	2.43
GO:0006650	glycerophospholipid metabolic process	72/5995	502/113005	6	2.43
GO:1901137	carbohydrate derivative biosynthetic process	78/5995	504/113005	4	2.43
GO:0009451	RNA modification	69/5995	461/113005	7	2.42
GO:0006813	potassium ion transport	51/5995	409/113005	7	2.4
GO:0007017	microtubule-based process	100/5995	1021/113005	2	2.39
GO:0030036	actin cytoskeleton organization	11/5995	37/113005	6	2.38
GO:0044283	small molecule biosynthetic process	105/5995	934/113005	3	2.38
GO:0006996	organelle organization	106/5995	823/113005	4	2.37
GO:0034470	ncRNA processing	93/5995	669/113005	8	2.37
GO:0008033	tRNA processing	76/5995	428/113005	9	2.37
GO:0030001	metal ion transport	94/5995	965/113005	6	2.37
GO:0006302	double-strand break repair	21/5995	124/113005	8	2.35
GO:0006974	cellular response to DNA damage stimulus	88/5995	911/113005	4	2.34
GO:0034654	nucleobase-containing compound biosynthetic process	116/5995	1127/113005	5	2.33
GO:0009057	macromolecule catabolic process	92/5995	885/113005	4	2.32
GO:0006399	tRNA metabolic process	95/5995	726/113005	8	2.32
GO:0010629	negative regulation of gene expression	28/5995	199/113005	6	2.3
GO:0006281	DNA repair	86/5995	859/113005	7	2.3
GO:0044255	cellular lipid metabolic process	115/5995	978/113005	4	2.29
GO:1901135	carbohydrate derivative metabolic process	117/5995	949/113005	3	2.29
GO:0006644	phospholipid metabolic process	80/5995	633/113005	5	2.28
GO:0022607	cellular component assembly	90/5995	838/113005	4	2.28
GO:0007018	microtubule-based movement	78/5995	785/113005	3	2.28
GO:0019438	aromatic compound biosynthetic process	158/5995	1413/113005	4	2.28
GO:0090407	organophosphate biosynthetic process	91/5995	812/113005	5	2.28
GO:0070647	protein modification by small protein conjugation or removal	73/5995	634/113005	7	2.28
GO:0044248	cellular catabolic process	131/5995	1206/113005	3	2.27
GO:0008610	lipid biosynthetic process	60/5995	566/113005	4	2.26
GO:0006082	organic acid metabolic process	139/5995	1609/113005	3	2.25
GO:0006812	cation transport	119/5995	1348/113005	5	2.25
GO:0043086	negative regulation of catalytic activity	9/5995	25/113005	4	2.24
GO:1901575	organic substance catabolic process	136/5995	1433/113005	3	2.21

Appendix 0 Table 9. Significant enriched terms relating to biological processes in the positively selected Saprolegniales proteins (continued).

GO number	Name	Ratio in study	Ratio in population	Depth	-log10 of p value
GO:0034645	cellular macromolecule biosynthetic process	152/5995	1546/113005	5	2.21
GO:0016043	cellular component organization	188/5995	1870/113005	3	2.21
GO:0009056	catabolic process	151/5995	1551/113005	2	2.2
GO:0018130	heterocycle biosynthetic process	158/5995	1513/113005	4	2.18
GO:0034660	ncRNA metabolic process	112/5995	973/113005	7	2.18
GO:0002098	tRNA wobble uridine modification	14/5995	62/113005	12	2.18
GO:0006396	RNA processing	181/5995	1482/113005	7	2.17
GO:0006259	DNA metabolic process	133/5995	1301/113005	6	2.17
GO:1901566	organonitrogen compound biosynthetic process	190/5995	2018/113005	4	2.16
GO:0071840	cellular component organization or biogenesis	193/5995	1982/113005	2	2.14
GO:0031122	cytoplasmic microtubule organization	6/5995	10/113005	7	2.14
GO:0006629	lipid metabolic process	166/5995	1491/113005	3	2.14
GO:0071897	DNA biosynthetic process	11/5995	39/113005	7	2.14
GO:0051205	protein insertion into membrane	11/5995	39/113005	7	2.14
GO:0019637	organophosphate metabolic process	154/5995	1529/113005	4	2.14
GO:0009059	macromolecule biosynthetic process	165/5995	1774/113005	4	2.13
GO:1901362	organic cyclic compound biosynthetic process	170/5995	1614/113005	4	2.12
GO:0019752	carboxylic acid metabolic process	123/5995	1464/113005	5	2.11
GO:0044271	cellular nitrogen compound biosynthetic process	210/5995	2228/113005	4	2.1
GO:0050896	response to stimulus	105/5995	1185/113005	1	2.09
GO:0032446	protein modification by small protein conjugation	40/5995	342/113005	8	2.08
GO:0016051	carbohydrate biosynthetic process	34/5995	274/113005	4	2.07
GO:0016070	RNA metabolic process	285/5995	2598/113005	6	2.05
GO:0046394	carboxylic acid biosynthetic process	52/5995	500/113005	6	2.05
GO:0090304	nucleic acid metabolic process	412/5995	3827/113005	5	2.02
GO:0044281	small molecule metabolic process	228/5995	2727/113005	2	2.01
GO:1901576	organic substance biosynthetic process	409/5995	4042/113005	3	2.0
GO:0065007	biological regulation	260/5995	3535/113005	1	2.0
GO:0044249	cellular biosynthetic process	364/5995	3740/113005	3	1.97
GO:0009058	biosynthetic process	422/5995	4297/113005	2	1.94
GO:0006139	nucleobase-containing compound metabolic process	517/5995	4883/113005	4	1.93
GO:0000724	double-strand break repair via homologous recombination	14/5995	65/113005	9	1.93
GO:0050789	regulation of biological process	242/5995	3297/113005	2	1.92
GO:0032502	developmental process	16/5995	83/113005	1	1.91
GO:0046483	heterocycle metabolic process	564/5995	5285/113005	3	1.91
GO:0006228	UTP biosynthetic process	8/5995	21/113005	11	1.91
GO:1901360	organic cyclic compound metabolic process	581/5995	5461/113005	3	1.89
GO:0006725	cellular aromatic compound metabolic process	564/5995	5305/113005	3	1.89
GO:0006793	phosphorus metabolic process	414/5995	6035/113005	3	1.89
GO:0034641	cellular nitrogen compound metabolic process	616/5995	6052/113005	3	1.88
GO:0006950	response to stress	98/5995	1139/113005	2	1.87
GO:0043412	macromolecule modification	477/5995	6689/113005	4	1.86
GO:0006796	phosphate-containing compound metabolic process	404/5995	6006/113005	4	1.85
GO:0044267	cellular protein metabolic process	466/5995	6974/113005	5	1.85
GO:0006421	asparaginyl-tRNA aminoacylation	6/5995	11/113005	11	1.83
GO:0044260	cellular macromolecule metabolic process	724/5995	9543/113005	4	1.8
GO:1901564	organonitrogen compound metabolic process	792/5995	12133/113005	3	1.75
GO:0046854	phosphatidylinositol phosphorylation	21/5995	136/113005	8	1.72
GO:0006807	nitrogen compound metabolic process	1232/5995	16008/113005	2	1.71
GO:0044237	cellular metabolic process	1339/5995	15528/113005	2	1.71
GO:0043170	macromolecule metabolic process	1040/5995	13998/113005	3	1.7
GO:0009206	purine ribonucleoside triphosphate biosynthetic process	14/5995	68/113005	10	1.7
GO:0044238	primary metabolic process	1373/5995	17978/113005	2	1.66
GO:0009116	nucleoside metabolic process	22/5995	148/113005	6	1.66
GO:0071704	organic substance metabolic process	1481/5995	19014/113005	2	1.64
GO:0008152	metabolic process	1664/5995	21941/113005	1	1.64
GO:0033013	tetrapyrrole metabolic process	18/5995	107/113005	4	1.62
GO:0009102	biotin biosynthetic process	6/5995	12/113005	8	1.56
GO:0007062	sister chromatid cohesion	10/5995	37/113005	6	1.54
GO:0016042	lipid catabolic process	16/5995	89/113005	4	1.53
GO:0006464	cellular protein modification process	408/5995	6220/113005	6	1.48
GO:0016579	protein deubiquitination	33/5995	280/113005	9	1.42
GO:0032787	monocarboxylic acid metabolic process	39/5995	355/113005	6	1.4
GO:0050794	regulation of cellular process	219/5995	3093/113005	3	1.4
GO:0050790	regulation of catalytic activity	30/5995	246/113005	3	1.39
GO:0061024	membrane organization	16/5995	92/113005	4	1.34
GO:0007165	signal transduction	85/5995	1001/113005	4	1.34
GO:0008612	peptidyl-lysine modification to peptidyl-hypusine	6/5995	13/113005	9	1.31

Appendix 0 Table 10. Significant enriched terms relating to biological processes in the Stramenopile dataset's paralogs.

GO number	Name	Ratio in study	Ratio in population	Depth	-log10 p value	Species
GO:0006629	lipid metabolic process	40/861	195/10646	3	2.85	Albugo candida
GO:0050789	regulation of biological process	75/861	467/10646	2	2.45	Albugo candida
GO:0050794	regulation of cellular process	71/861	437/10646	3	2.38	Albugo candida
GO:0006796	phosphate-containing compound metabolic process	92/861	661/10646	4	2.31	Albugo candida
GO:0044238	primary metabolic process	249/861	2192/10646	2	2.11	Albugo candida
GO:0016310	phosphorylation	62/861	414/10646	5	2.07	Albugo candida
GO:0044255	cellular lipid metabolic process	30/861	153/10646	4	1.96	Albugo candida
GO:0043412	macromolecule modification	93/861	736/10646	4	1.68	Albugo candida
GO:0008299	isoprenoid biosynthetic process	8/861	16/10646	6	1.67	Albugo candida
GO:0006468	protein phosphorylation	54/861	364/10646	7	1.67	Albugo candida
GO:0006650	glycerophospholipid metabolic process	20/861	89/10646	6	1.44	Albugo candida
GO:0051726	regulation of cell cycle	13/861	43/10646	4	1.44	Albugo candida
GO:0046488	phosphatidylinositol metabolic process	18/861	76/10646	7	1.36	Albugo candida
GO:0016570	histone modification	14/861	50/10646	7	1.33	Albugo candida
GO:0034645	cellular macromolecule biosynthetic process	37/714	243/10801	5	2.38	Albugo laibachii
GO:0006075	(1->3)-beta-D-glucan biosynthetic process	6/714	9/10801	9	1.99	Albugo laibachii
GO:0044262	cellular carbohydrate metabolic process	12/714	44/10801	4	1.48	Albugo laibachii
GO:0051560	mitochondrial calcium ion homeostasis	4/714	4/10801	10	1.48	Albugo laibachii
GO:0032324	molybdopterin cofactor biosynthetic process	6/541	19/18668	6	1.67	Aphanomyces invadans
GO:0045893	positive regulation of transcription, DNA-templated	4/541	7/18668	10	1.37	Aphanomyces invadans
GO:0055114	oxidation-reduction process	25/272	277/8136	2	2.04	Bremia lactucae
GO:0018216	peptidyl-arginine methylation	5/272	10/8136	9	1.82	Bremia lactucae
GO:0006811	ion transport	20/218	280/10489	4	2.6	Hyaloperonospora arabidopsidis
GO:0051179	localization	28/218	505/10489	1	2.49	Hyaloperonospora arabidopsidis
GO:0006810	transport	26/218	495/10489	3	1.69	Hyaloperonospora arabidopsidis
GO:0006820	anion transport	14/218	182/10489	5	1.36	Hyaloperonospora arabidopsidis
GO:0022607	cellular component assembly	59/4204	192/25544	4	2.23	Nothophytophthora sp.
GO:0043933	protein-containing complex subunit organization	46/4204	143/25544	4	2.16	Nothophytophthora sp.
GO:0016042	lipid catabolic process	18/4204	38/25544	4	1.79	Nothophytophthora sp.
GO:0003341	cilium movement	11/4204	18/25544	4	1.35	Nothophytophthora sp.
GO:0016042	lipid catabolic process	6/212	25/17026	4	3.03	Phytophthora capsici
GO:0051016	barbed-end actin filament capping	4/245	5/20567	13	3.74	Phytophthora cinnamomi
GO:0048523	negative regulation of cellular process	6/245	38/20567	4	2.0	Phytophthora cinnamomi
GO:0034219	carbohydrate transmembrane transport	2/40	2/9567	8	1.52	Phytophthora kernoviae
GO:0098656	anion transmembrane transport	18/1969	47/16203	6	2.11	Phytophthora nicotianae
GO:0006996	organelle organization	28/1130	193/23486	4	2.92	Phytophthora parasitica
GO:0007017	microtubule-based process	34/1130	262/23486	2	2.59	Phytophthora parasitica
GO:0016043	cellular component organization	45/1130	404/23486	3	2.55	Phytophthora parasitica
GO:0090304	nucleic acid metabolic process	89/1130	929/23486	5	2.39	Phytophthora parasitica
GO:0016070	RNA metabolic process	54/1130	552/23486	6	2.33	Phytophthora parasitica
GO:0006139	nucleobase-containing compound metabolic process	94/1130	1128/23486	4	2.25	Phytophthora parasitica
GO:0046483	heterocycle metabolic process	98/1130	1202/23486	3	2.23	Phytophthora parasitica
GO:0006725	cellular aromatic compound metabolic process	96/1130	1208/23486	3	2.07	Phytophthora parasitica
GO:0065003	protein-containing complex assembly	17/1130	99/23486	5	2.05	Phytophthora parasitica
GO:0006397	mRNA processing	18/1130	111/23486	8	1.97	Phytophthora parasitica
GO:0007018	microtubule-based movement	26/1130	210/23486	3	1.72	Phytophthora parasitica
GO:0034641	cellular nitrogen compound metabolic process	100/1130	1337/23486	3	1.63	Phytophthora parasitica
GO:0022607	cellular component assembly	21/1130	153/23486	4	1.56	Phytophthora parasitica
GO:0006396	RNA processing	33/1130	306/23486	7	1.47	Phytophthora parasitica
GO:0051276	chromosome organization	14/1130	79/23486	5	1.38	Phytophthora parasitica
GO:0065007	biological regulation	62/1130	739/23486	1	1.36	Phytophthora parasitica
GO:0042592	homeostatic process	16/1130	101/23486	3	1.32	Phytophthora parasitica
GO:0043484	regulation of RNA splicing	6/1130	14/23486	7	1.3	Phytophthora parasitica
GO:0015074	DNA integration	12/428	111/21542	7	2.39	Phytophthora sojae
GO:0006310	DNA recombination	12/428	133/21542	7	1.57	Phytophthora sojae
GO:0046168	glycerol-3-phosphate catabolic process	4/770	4/13559	7	1.73	Pythium insidiosum
GO:0044237	cellular metabolic process	6/180	2053/14021	2	2.44	Pythium oligandrum
GO:0070588	calcium ion transmembrane transport	4/180	12/14021	9	1.65	Pythium oligandrum
GO:0005975	carbohydrate metabolic process	10/111	214/13625	3	1.75	Pythium ultimum
GO:0007186	G protein-coupled receptor signaling pathway	6/284	18/16252	5	3.1	Saprolegnia diclina
GO:0046488	phosphatidylinositol metabolic process	8/284	67/16252	7	1.4	Saprolegnia diclina
GO:0055114	oxidation-reduction process	12/60	433/11715	2	2.54	Thraustotheca clavata

Figure 1–Figure supplement 1. Differences in annotated cellular pathways from the Stramenopile dataset. Shown are pathways which have up to 30 repeated values per taxa.

Figure 1–Figure supplement 2. Differences in annotated cellular pathways for the members of the Saprolegniales in the Stramenopile dataset.

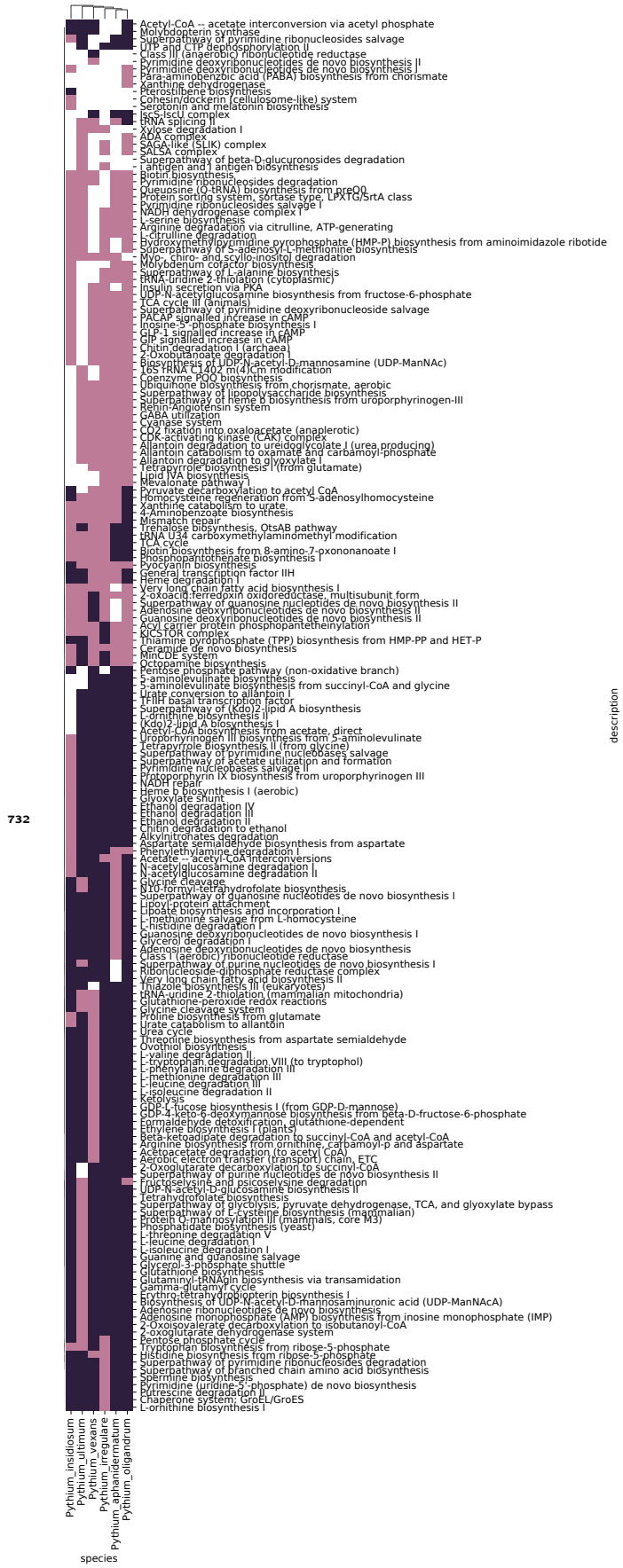

Figure 1–Figure supplement 3. Differences in annotated cellular pathways for the members of the Pythiales order in the Stramenopile dataset.

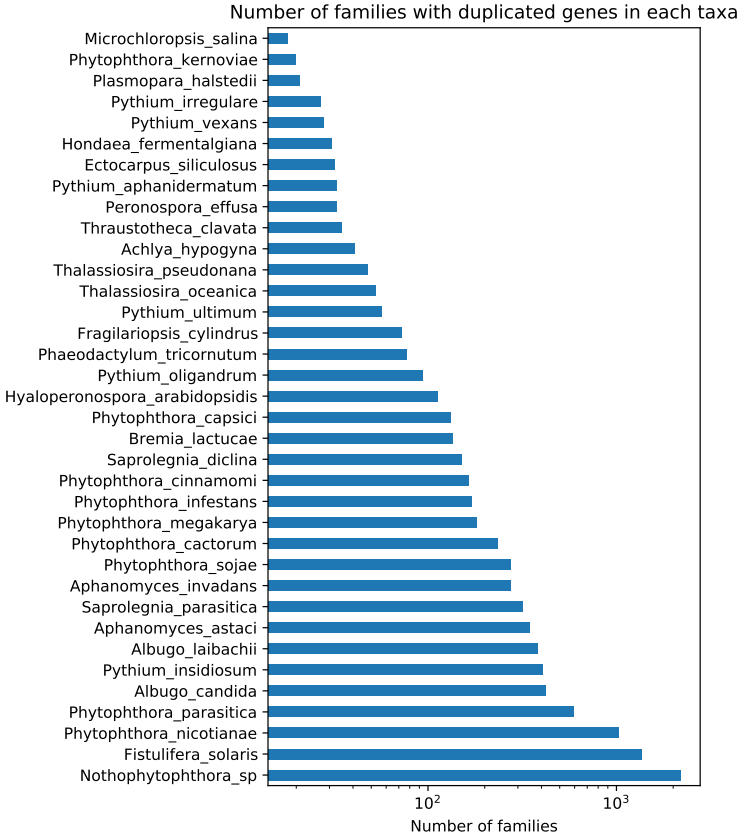

Figure 2–Figure supplement 1. Duplicates in protein families in the Stramenopile dataset. Number of families with five or more members that contain paralogs in the dataset.

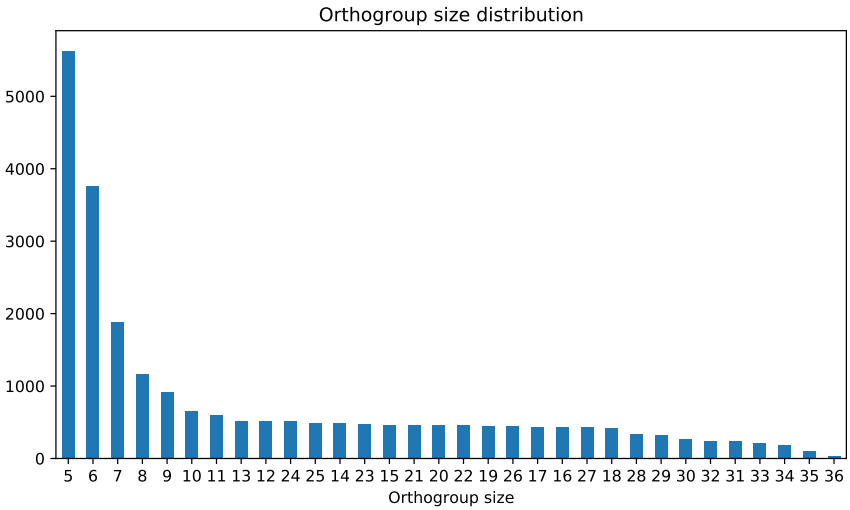

Figure 2–Figure supplement 2. Distribution of protein family size in the dataset. Number of families with the same member size.

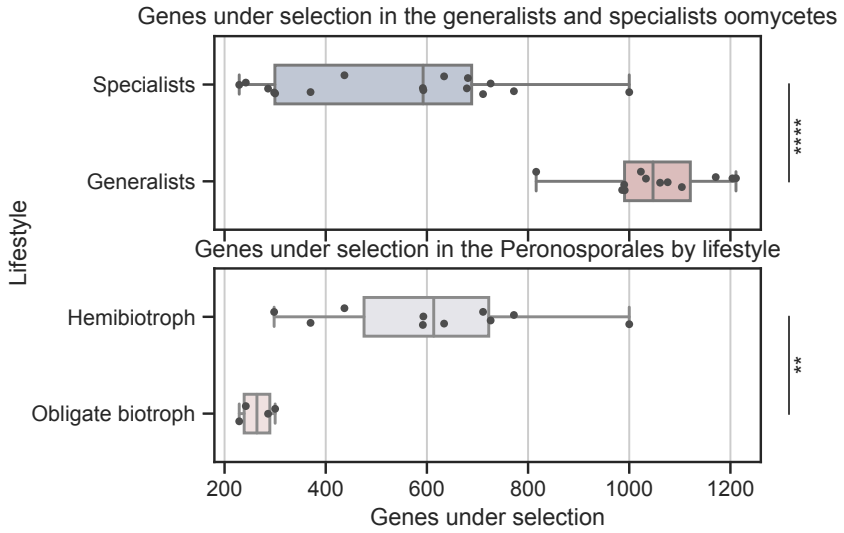

Figure 3–Figure supplement 1. Comparison of the distribution of positive genes under selection for different lifestyles. Specialists refers to Albuginales and Peronosporales and generalists to Pythiales and Saprolegniales.

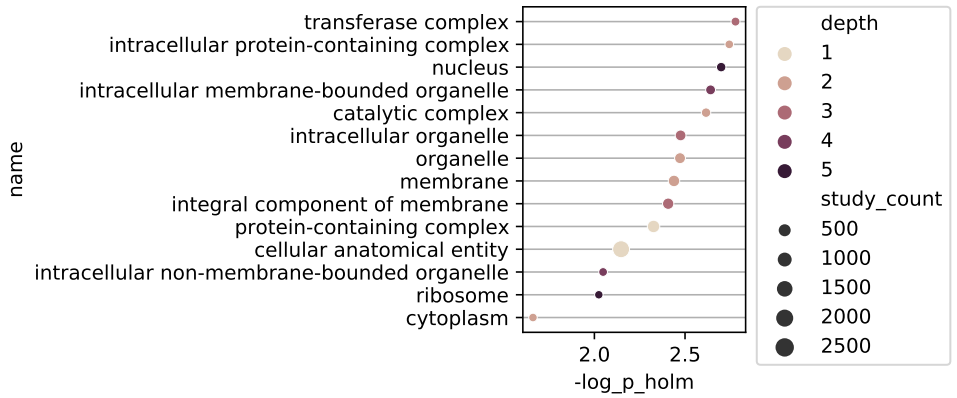

Figure 5–Figure supplement 1. Cellular compartment enrichment under selection in all oomycetes in the Stramenopile dataset.

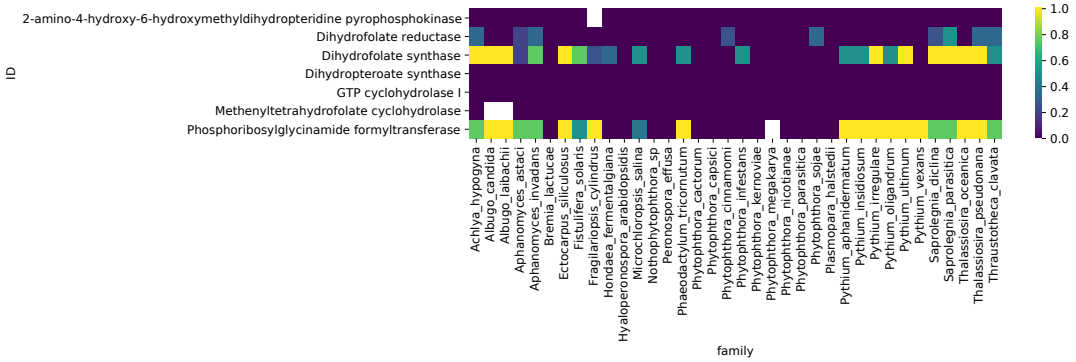

Figure 7–Figure supplement 1. Tetrahydrofolate salvage and biosynthesis-related enzymes in Stramenopiles. Heatmap of the presence and absence of the enzymes relating to tetrahydrophol-late metabolism in the Stramenopiles. The yellow gradient represents the ratio of predicted positive selection in genes with this annotation.

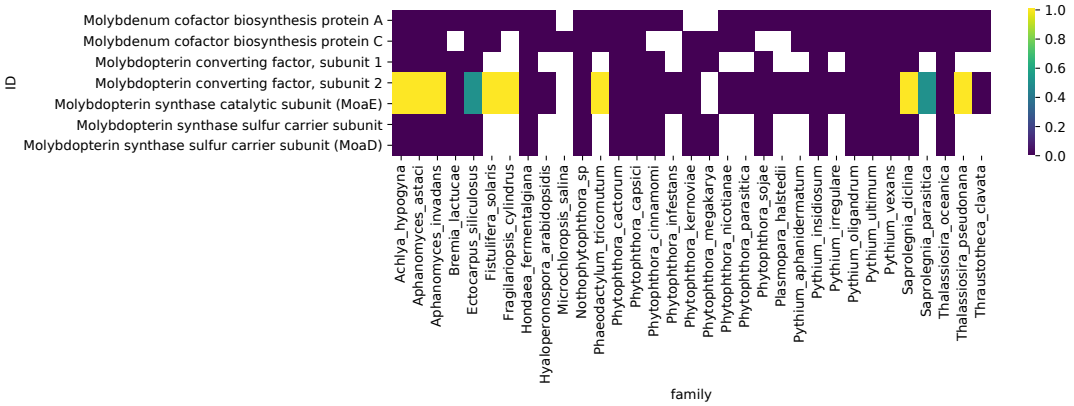

Figure 7–Figure supplement 2. Molybdopterin synthesis enzymes in Stramenopiles. Heatmap of the presence and absence of the enzymes relating to molybdopterin biosynthesis in the Stramenopiles. The yellow gradient represents the ratio of predicted positive selection in genes with this annotation.

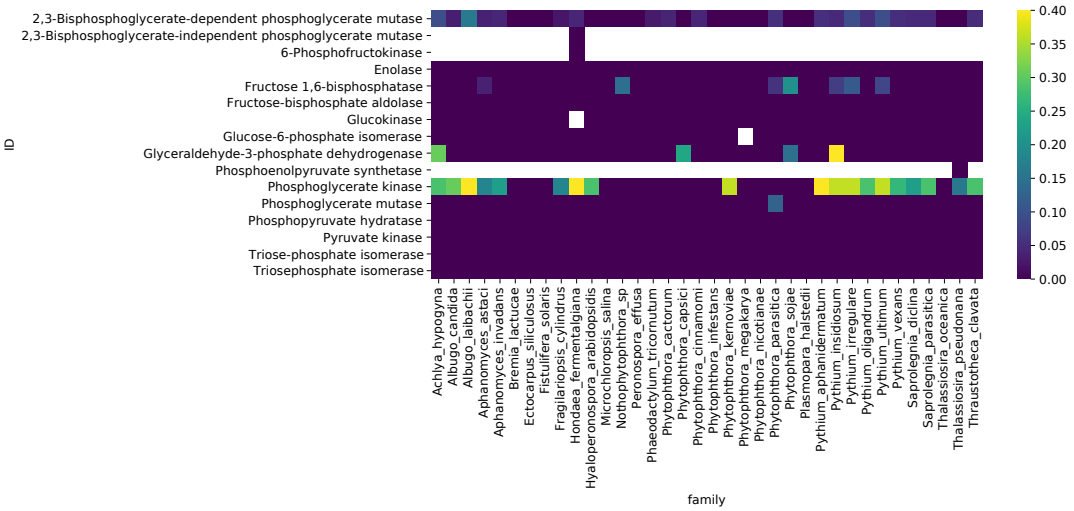

Figure 7–Figure supplement 3. Glycolysis I, II and III-related enzymes in Stramenopiles. Heatmap of the presence and absence of the enzymes relating to glycolysis pathway in the Stramenopiles. The yellow gradient represents the normalized ratio of predicted positive selection in genes with this annotation.

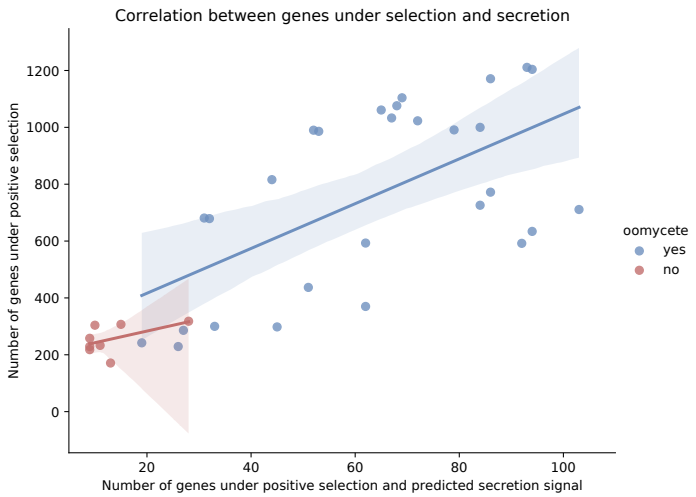

Figure 9–Figure supplement 1. Correlation between genes under selection and secretion. Oomycetes represented in blue (Pearson's correlation, r = 0.6, p = < 0.01) and non-oomycetes in red (Pearson's correlation, r = 0.5, p = 0.2). Pearson correlation represented as a straight line and the confidence interval represented as a lighter shade.