RESEARCH

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

Daniel Gómez-Pérez* and Eric Kemen

*Correspondence: daniel.gomez-perez@uni-tuebingen.de
Zentrum für Molekularbiologie der Pflanzen, Auf der Morgenstelle 32, 72076 Tübingen, Germany
Full list of author information is available at the end of the article

Abstract

Background: Host and niche shifts are a source of genomic and phenotypic diversification as evidenced in parasitism. Exemplary is core metabolism reduction as parasites adapt to a particular host, while the accessory genome often maintains a high degree of diversification. However, selective pressures acting on the genome of organisms that have undergone lifestyle or host change have not been fully investigated.

Results: Here, we developed a comparative genomics approach to study underlying adaptive trends in oomycetes, a eukaryotic phylum with a broad range of economically important plant and animal parasitic lifestyles. Our analysis reveals converging evolution on biological processes for oomycetes that have similar lifestyle. Besides, we find that certain functions, in particular carbohydrate metabolism, transport, and signaling, are important for host and environmental adaption in oomycetes.

Discussion: Given the high correlation between lifestyle and genome properties in our oomycete dataset and the convergent evolution of fungal and oomycete genomes, we have developed a model that predicts plant pathogen lifestyles with high accuracy based on functional annotations. Understanding how genomes and selective pressures correlate with lifestyle may be crucial to identify new emerging diseases and pandemic threats.

Keywords: oomycetes; lifestyle; evolution

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 [1, 2]. 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 [3–5]. 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 [6]. 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 [7]. 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 [8]. However, a small portion of them, those that have been subject to recent diversification, show signs of an increased nonsynonymous mutation rate. Codon models that take into account statistical rate variations are commonly used in comparative genomic studies [9]. When performed on related organisms that have different lifestyles and hosts the study of positively selected genes together with their functional annotation illustrates which gene functions played important roles in the adaptation process.

Oomycetes are eukaryotic organisms belonging, together with diatoms and brown algae, to the group of Stramenopiles [10, 11]. Since their origin from a marine autotrophic common ancestor around 400 million years ago, oomycetes have adapted to multiple environments and lifestyles, and many of them are economically impactful plant and animal parasites [12–14]. Therefore, they represent a relevant and appropriate system to study the genetic impact of lifestyle and host adaptation on genetically close genomes. Four phylogenetic families, representative of oomycete's large diversity, are the target of most current research efforts: Albuginaceae, Peronosporaceae, Saprolegniaceae, and Pythiaceae. The Albuginaceae and some Peronosporaceae independently evolved the ability to survive exclusively on living host material, also known as obligate biotrophy [15]. Most Peronosporaceae 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 [16]. Additionally in the Peronosporaceae, the early divergent clade of *Globisporangium* consists of plant necrotrophs previously classified as *Pythiaceae*. All Albuginaceae, Peronosporaceae, and most Pythiaceae are plant parasitic organisms [17]. On the contrary, most Saprolegniaceae are capable of infecting animals, with few exceptions including plant-causing root rot *Aphanomyces* and the free-living saprophyte *Thraustotheca clavata*, which does not need a host at any point in its life cycle [18–20].

Obligate biotrophs have a considerably reduced primary metabolism. Comparative genome studies have reported a significant and convergent loss of the enzymatic arsenal in independent lineages of the oomycetes following this lifestyle [21]. The picture is not so clear for the heterotrophs and their adaptation 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 [22]. There are many theories that explain how such drastic host shifts can occur in a small evolutionary timescale [23]. Particularly in oomycetes, large reservoirs of noncoding DNA material can readily evolve into small secreted proteins, known as effectors, facilitating new oomycete-host interactions [24]. Additionally, the 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 [25–27]. However, the impact of host shifts on genomic selective pressures has not been extensively studied.

There is a high degree of convergent evolution between oomycetes and fungi [28]. Both share many of the niches mentioned, including pathogenic niches of animals and plants, as well as lifestyles, including saprotrophy, hemibiotrophy, and obligate biotrophy. Oomycetes and fungi have developed similar strategies to overcome the same challenges, including comparable filamentous and reproductive morphology, as

well as akin infection strategies [29]. As mentioned above, convergence is probably promoted by genetic exchange, as the source of many oomycete genes with a role in host adaptation can be traced back to pathogenic fungi [30]. Because of the parallels between the adaptive strategies of these two eukaryotic phyla, we can infer underlying mechanistic principles in oomycetes on the basis of 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 34 representative members and an outgroup of eight non-oomycete Stramenopiles (Table 2). We compared the genes inferred as being under positive selection to the background annotated genes to identify 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 functions.

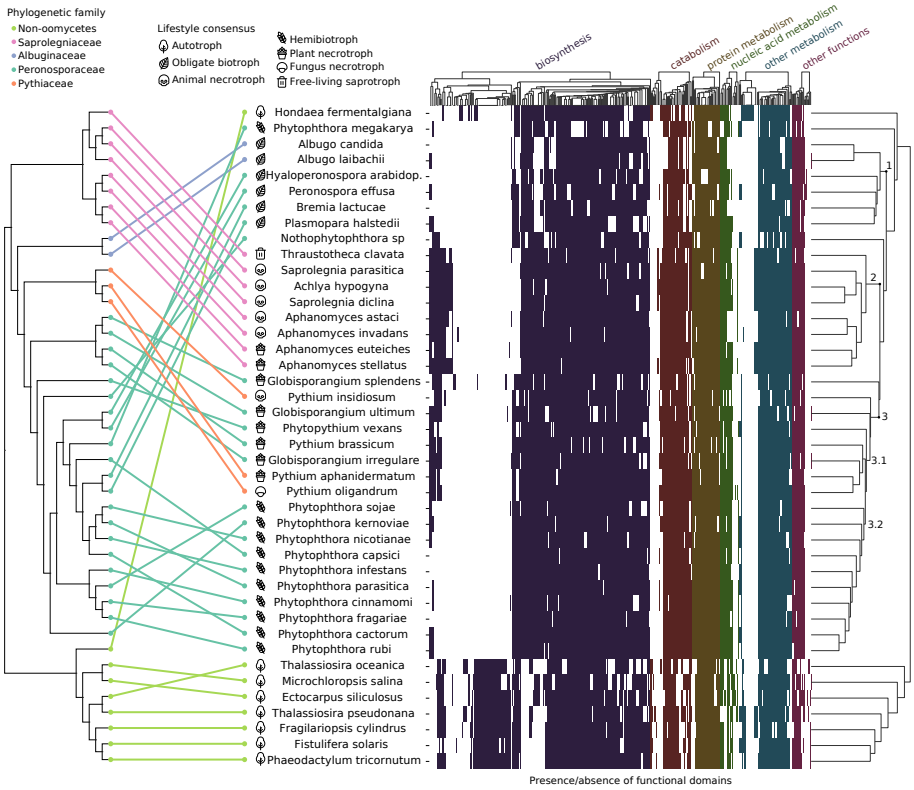

Figure 1 Presence/absence of functional attributes in the genome of the Stramenopiles dataset correlated with phylogeny. Equal distance cladogram constructed from conserved families on the left and clustering by UPGMA of genome properties of the dataset on the right. In the equal-distance phylogenetic tree, colored lines match phylogeny to the clustered taxa with annotated lifestyles. In the heatmap, different colors represent the presence or absence of particular functional groups belonging to the specified categories.

Results

Proteome annotation and clustering

We downloaded the genomes of 34 oomycete species and eight non-oomycete Stramenopiles from the NCBI and FungiDB databases and annotated their proteomes by the presence or absence of known functional signatures to get insights into the divergence of the dataset (Figure 1) [31, 32]. Unweighted Pair Group Method with Arithmetic Mean (UPGMA) 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 annotations in the oomycetes, such as chlorophyll biosynthesis (Figure 9). In the oomycetes, we defined three clusters based on their distance (1-3 in Figure 1): obligate biotrophs, Saprolegniaceae, and a final one grouping most of the Perosporanaceae and Pythiaceae of the dataset. The obligate biotroph cluster consisted of the Albuginaceae and the downy mildews from the Peronosporaceae (*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 many functional annotations in comparison with other oomycetes. A notable feature of this group was the lack of core biosynthetic pathways, including vitamin and cofactor biosynthesis, for which they presumably rely on their host (Figure 9). The Saprolegniaceae differed from other oomycetes mainly in the presence of steroid biosynthesis pathways (Figure 10). In the third cluster, we defined two subclusters, labeled as 3.1 and 3.2 in Figure 1. The first contained four of the *Pythium* and *Globisporangium* species of the dataset, and the second one included exclusively all *Phythophthora* in the dataset (except for *Phytophthora megakarya*. The *Pythium* and *Globisporangium* species in the dataset also had biosynthetic pathways that most other oomycetes lacked and that they often shared with the Saprolegniaceae, as a result most likely of their common facultative lifestyles. The hemibiotroph group, consisting of most of the *Phytophthora* species in the dataset, showed significant metabolic reduction, but not as extensive as in the obligate biotrophs [33].

These clusters and subclusters 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 Saprolegniaceae in the dataset are facultative animal necrotrophs, and most *Pythium* and *Globisporangium* species facultative plant necrotrophs. Interestingly, *T. clavata*, the free-living organism in the dataset, clustered as an outgroup of the other phylogenetically close Saprolegniaceae, showing the greatest distance to its animal and plant-infecting neighbours. The most notable differences in the presence/absence of cellular pathways of this *T. clavata* assembly when compared to other Saprolegniaceae were the absence of the endopeptidase ClpXP complex and RuvB-like helicase I (Figure 10). However, there were some exceptions to this arrangement, with some taxa clustering with a different lifestyle or failing to cluster with their own lifestyle. For example, the clustering of the two plant infecting necrotrophs of the Saprolegniaceae follows the phylogeny of the *Aphanomyces* genus.

P. insidiosum, the only animal pathogen in the Pythiaceae, showed remarkably different genome properties from its peers, being placed as an outgroup of

hemibiotrophs and Pythiaceae. 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 Pythiaceae lacked. Of note, pterostilbene has been shown to have strong immunosuppressive properties in animals [34]. Still, *P. insidiosum* retained some of the Pythiaceae nutrient assimilation pathways, including the Leloir pathway for the catabolism of D-galactose, as well as the methionine salvage and allantoin catabolic pathways for, respectively, sulphur and nitrogen assimilation. Another outgroup of the same cluster was represented by *Nothophytophthora*, a hybrid species of the Peronosporaceae family 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 but are retained in the facultative necrotroph oomycetes (except *Phytopythium vexans*). Based on this evidence and the prediction of necrotrophic lifestyle with the model we describe below, we speculate a facultative necrotroph lifestyle for *Nothophytophthora*, in contrast to the hemibiotroph neighbouring Peronosporaceae. It is not uncommon for hybridization to facilitate niche or lifestyle adaptation [35, 36]. In the Pythiaceae, the mycopathogen *Pythium oligandrum* clustered with plant pathogenic Pythiaceae. Notable was its lack of inositol degradation pathways and the partial presence of xanthine dehydrogenase and para-aminobenzoic acid biosynthesis from the chorismate pathway (Figure 11). In summary, our analysis 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 42 genomes, we classified the proteomes into ortholog groups by taking sequence similarity and in addition gene order into account. We selected protein clusters that had at least five members from different taxa to get a good balance between a representative number of families and results that are statistically robust. This corresponded to 29,123 protein families, which cover about half (49.02%) of the total proteins in the dataset (Figure 2). The orthogroups were mainly composed of one-to-one orthologs (78.70% of families), however, we detected a significant number of paralogs in some oomycetes, particularly for *Nothophytophthora* sp., as well as for *Phytophthora nicotianae*, *Globisporangium splendens* and *Phytophthora parasitica* (Figure 12). This might be related to the reported whole genome duplications in *Phytopthora* species [37], as well as the recent hybridization event that gave rise to *Nothophytophthora* [38]. Additionally, the diatom *Fistulifera solaris*'s large presence of gene duplications highlights its recent whole genome duplication [39].

The most abundant orthogroups had between five and nine members (Figure 13). Orthogroups corresponding to all taxa were a minority. Instead, most orthogroups were present in closely related five to ten-member clades. When looking at the number of genes not assigned to orthogroups in the oomycetes, the *Phytophthora* genus had the highest count (Figure 2). This may be related to the large arsenal of unique effectors that lack no conserved domains or homologs outside of their own species and play a large role in host adaptation. *Aphanomyces astaci* also

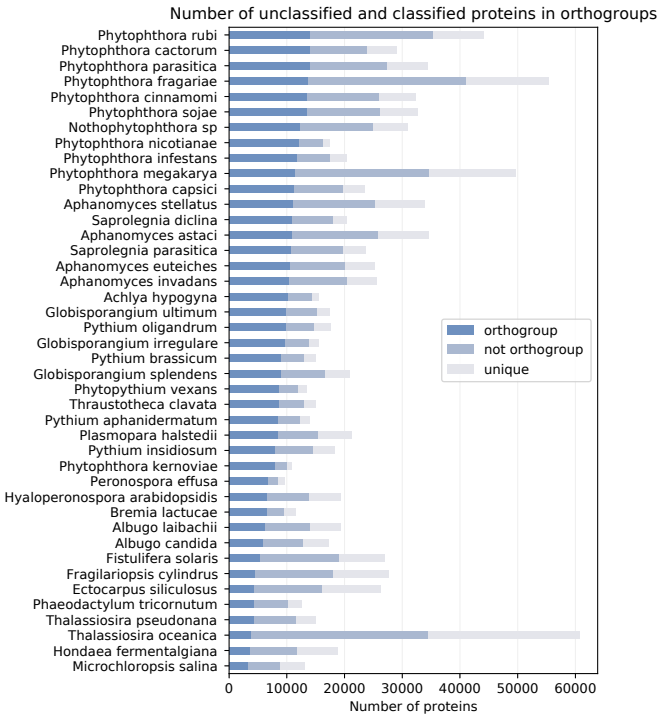

Figure 2 Protein-encoding genes from the Stramenopile dataset classified by taxa. Number of proteins classified into orthogroups (protein families of five or more members), not in an orthogroup (protein families of less than five members), or unique (not in a protein family).

had a high amount of genes outside of the orthogroups, most likely because of the recent expansions in its genome [40]. In summary, this highlights a patchy ortholog distribution in the dataset, with most protein families conserved only in phylogenetically close members of clades (Figure 13). Despite this, a significant pool of ortholog protein families representative of the Stramenopile genomes in the dataset could be inferred from the analysis as further discussed below.

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 clades. Members of the Saprolegniaceae and Pythiaceae, together with the necrotrophic *Globisporangium* had a higher count and therefore more genes under selection in orthogroups (mean = 1222, std = 152) than the remaining Peronosporaceae and the Albuginaceae (mean= 577, std = 245) (Figure 14). A special case was the hybrid *Nothophytophthora* sp., which had a comparable amount of positively selected genes to Pythiaceae and Saprolegniaceae, however composed in great part by duplicated genes after speciation, 44.45% of the total (orange bar). When comparing necrotrophs, hemibiotrophs, and obligate biotrophs within the Peronosporaceae family (mean = 1344, 663, and 269, respectively), the trend was

that of a decrease in the number of genes under positive selection with the increase of biotrophic potential (Figure 14).

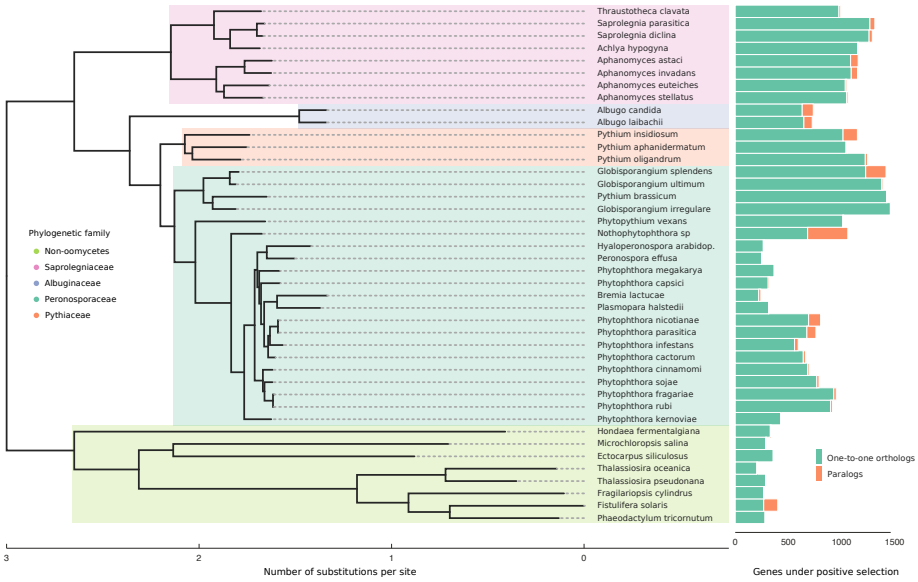

Figure 3 Number of genes under positive selection in the Stramenopile dataset. Maximum likelihood supertree constructed from inferred protein families in the Stramenopile dataset that are conserved in at least 25 taxa, corresponding to 3013 families of orthologs. Positively selected genes are represented as bars. One-to-one orthologs are represented in green, duplicated genes inside the same family under positive selection in orange.

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 = 0.50, p value < 0.01). A correlation of 0.5 suggested that there may be a larger number of positives because of more extensive testing in the oomycete species, as they have on average more members in the ortholog dataset. This bias is more evident in the non-oomycetes (Pearson's correlation, r = 0.52, p value = 0.18) than when considering just the oomycetes (Pearson's correlation, r = 0.15, p value = 0.39). As the proteomes of the non-oomycetes are overall smaller compared to oomycetes (Figure 4), we hypothesize that less extensive testing renders them more prone to this bias.

Out of the 32,661 detected genes under positive selection, 21,247 were successfully annotated with at least a gene ontology term (65%). We performed GO enrichment on the four main oomycete lifestyles in the Stramenopile dataset. The results are discussed below. As a control for the reliability of the pipeline, we performed the same analyses in a subset of 26 plant pathogens from a dataset of 65 basidiomycete fungi (Table 3). Highly enriched terms included processes known to be associated with virulence in such pathogenic fungi, like fatty acid and certain amino acid biosynthesis, ion transport, and protein targeting and transport (Table 5) [41–43].

In summary, we could identify signatures of positive selection in 4.14% of all genes analyzed in the Stramenopile dataset. A significant number could be functionally annotated and potential functions assigned.

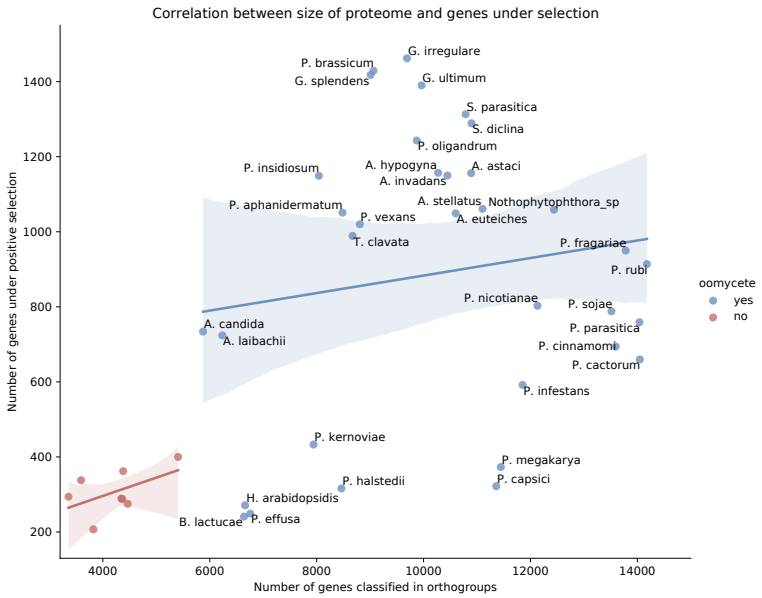

Figure 4 Correlation between genes under positive selection and proteome size in the Stramenopile dataset. Oomycetes are in blue (Pearson's correlation, r = 0.15, p value = 0.39) and non-oomycetes in red (Pearson's correlation, r = 0.52, p value = 0.18). Pearson correlation represented as a straight line and the confidence interval represented as a lighter shade.

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 response to stress, signal transduction, transmembrane transport, protein modification processes (phosphorylation, in particular), and localization, as well as numerous carbohydrate, lipid, nitrogen, and sulfur metabolism-related terms. Within the metabolism, abundant terms relating to biosynthesis are present. In the cellular compartment GO category, highly enriched terms include protein-containing complexes (for which transferase complexes show the larger significance), nucleus, intracellular organelles (for which ribosome shows the largest significance), and membranes (Figure 17).

Additionally, we performed similar enrichments on the oomycete groups defined by their lifestyle. We found the largest unique GO terms to belong to the plant and animal necrotrophs (36 and 21, respectively). In the plant necrotrophs, these included terms related to ion transport, carbohydrate biosynthesis, protein modification, and gene expression regulation. In the animal necrotrophs, unique terms had to do with vitamin biosynthesis, cilium movement, and protein localization. There were three unique terms in the hemibiotrophs related to response against stress and transmembrane transport while no unique terms were identified in the obligate biotrophs. We observed the largest overlap between animal and plant facultative necrotroph groups (59 common terms). These terms related to cell communication, glycolysis, organelle assembly, protein import, regulation of response to stimulus, translation, and numerous and diverse metabolic processes. This was followed by a smaller overlap of enriched functions in all four lifestyle groups, amounting to 33

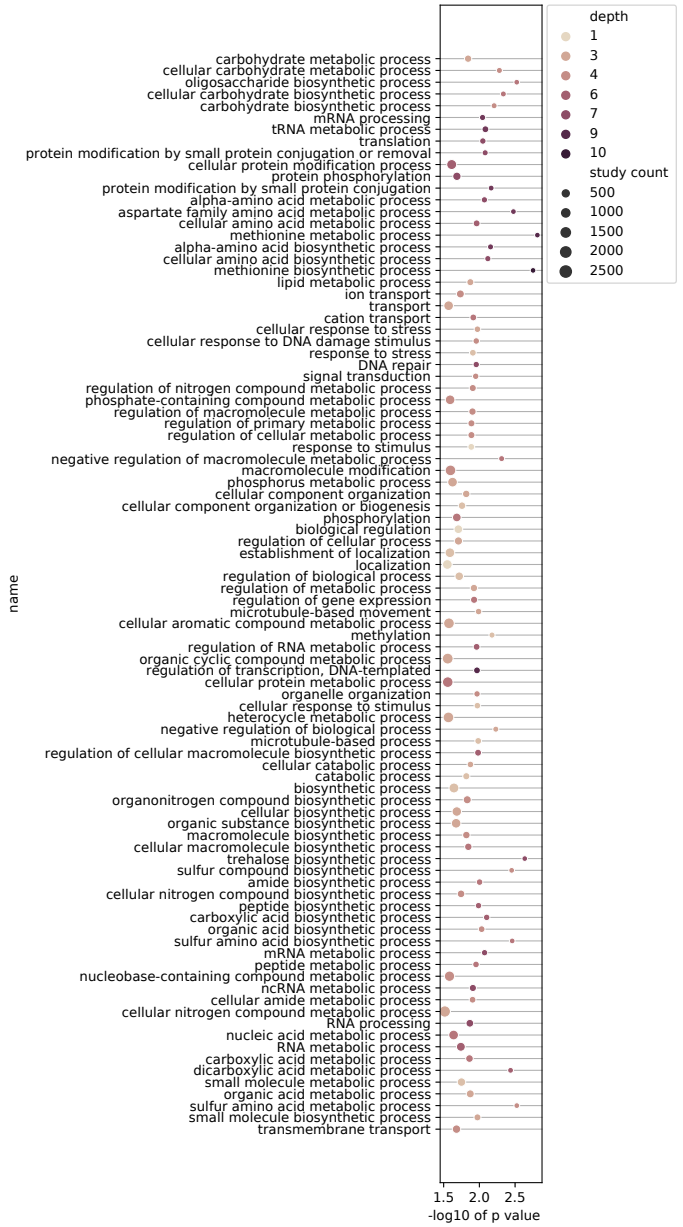

Figure 5 Significantly enriched biological processes in all oomycetes in the Stramenopile dataset. Included are GO terms with a corrected negative base 10 logarithm of the p value higher than 1.5 ordered by category using GO slim database. The color represents the GO depth. GO depth is a measure of the 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.

terms (Figure 6). The most significant terms for each lifestyle are listed in Tables 6-11.

We also studied the enrichment of biological functions in the expanded gene families of the dataset independently of whether the genes were under positive selection. In general, found that it reflected positive selection enrichment, however, the terms were highly variable when comparing different species (Table 12). In the ob-

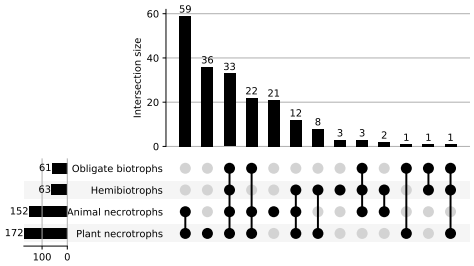

Figure 6 Upsetplot showing overlapping biological functions under selection in the oomycete lifestyle groups. The four groups correspond to the major lifestyles in the oomycetes of the Stramenopiles dataset.

ligate biotrophs, these related to phospholipid metabolism, cell wall biosynthesis, protein modification, biological regulation, and transmembrane transport. In the hemibiotrophs, they related to lipid metabolism, signaling, protein modification, and again to biological regulation, and transmembrane transport. Finally, in the plant necrotrophs, to DNA integration and localization.

Lifestyle prediction

We visualized in a heatmap all functional annotations with added information of positive selection by performing the same clustering as we did for the genome properties (Figure 16). We find that adding the positive selection data improves the clustering by lifestyle, particularly of the plant necrotrophs in the Pythiaceae and *Globisporangium*, which now form a single cluster that is closer to the other facultative necrotrophs of the dataset, the Saprolegniales, than to the obligate biotroph and hemibiotroph oomycetes in the dataset. Using the Robison-Foulds metric for clusters we find that there is a higher congruence between the phylogenetic tree and the genome properties clustering than to the positive selection one (Table 1).

Although we find that the positive selection information improves lifestyle prediction, we argue that it is impractical to implement as prediction method because it is computationally very intensive to calculate and not likely to be reproducible using different backgrounds for positive selection analyses. Therefore, we constructed a model to predict lifestyle in plant pathogenic fungi and oomycetes based on the genome properties alone. We assembled a dataset based on 115 plant pathogenic and saprotrophic fungi and oomycetes genomes (Table 4). 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 = 0.11, accuracy = 0.95), failing to predict two genomes in the hemibiotrophs and one in the biotrophs of the validation dataset (Figure 7). The model and the steps to reproduce it together with the entire dataset can be found at `https://github.com/danielzmbp/lspred`.

Table 1 Distance comparisons in the clusterings of the Stramenopile dataset. Phylogenetic and genome properties clustering is shown in Figure 1 and positive selection clustering in Figure 16

Clustering 1	Clustering 2	Robison-Foulds distance metric
Phylogenetic	Genome properties	28
Phylogenetic	Positive selection	30
Genome properties	Positive selection	24

Discussion

Functional genome annotations largely correlate with lifestyle

Convergence of the presence/absence of key functional annotations in species that do not share the same phylogenetic history but have similar lifestyle has been shown before for different sets of organisms [44, 45]. Distant species with the same lifestyle require similar functional biological processes, which results in similar selective pressures that analogously shape their genome, often leading to convergent evolution. Comparable to the study by Rodenburg *et al.* (2020) [46], we have shown the tight clustering of some groups with a similar lifestyle, most strikingly for the obligate biotrophs and hemibiotrophs. Conversely, there are a few exceptions, such as the hemibiotroph *P. megakarya* and the necrotroph *Globisporangium splendens*, which do not clearly cluster with any of the other oomycetes. We hypothesize this may be due to the quality of their gene annotation. Both have significantly lower number of key orthologs from the reference Stramenopile dataset as compared to other *Phytophthora* and *Globisporangium* species in the dataset (Table 2).

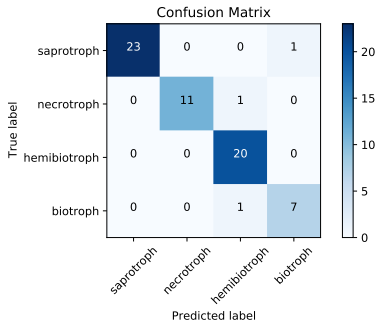

Figure 7 Confusion matrix of lifestyle predictor model. Results of predictions in the random validation set of 64 annotated proteomes. True values are represented in the x-axis and predicted values in the y-axis.

Generalists have more genes under positive selection

A higher number of genes under selection was found for the more generalist families of Saprolegniaceae, Pythiaceae, and necrotrophic Peronosporaceae, including the *Globisporangium* and *Phytopythium* clades, when compared to the more specialists remaining Peronosporaceae and Albuginaceae (Mann-Whitney test, $p < 0.01$). Within the Peronosporaceae, hemibiotrophs have a lower number of genes under selection than the facultative necrotrophs, and obligate biotrophs have in turn a lower number than hemibiotrophs (ANOVA one-tailed test, $p < 0.01$) (Figure 14). Thus, the number of genes under selection is inversely correlated to the biotrophic potential. With biotrophic potential we refer to the capability of survival exclusively

on a living host, such that no obligate biotroph can be cultured in vitro, while for some hemibiotrophs this is the case. On the opposite side of the spectrum, facultative plant necrotrophs thrive as saprotrophs without the need for a host. This correlation 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 genus as shown in Figure 4.

While all hemibiotrophs and biotrophs are obligate plant parasites, the necrotrophs in the Peronosporacea, Pythiaceae and Saprolegniaceae families show adaptation to a variety of lifestyles. They are facultative parasites of either animals, plants, or other fungi and oomycetes. Facultative parasites can live as saprotrophs on decaying matter but also as opportunistic necrotrophs on a suitable host [47]. The higher number of potential niches they are able to succesfully occupy may drive a larger number of genes to be under positive diversifying selection. Additionally, when compared to the obligate biotrophs and hemibiotrophs, which are highly adapted to infect a particular species, e.g., lettuce for *B. lactucae* and soybean for *Phytophthora sojae*, most of the necrotrophs 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 [48]. Having a higher number of genes under positive selection could be therefore correlated with this higher host flexibility.

Selective pressures in the ooomycetes help explain host adaptation

Biological functions under selection for all oomycetes in the Stramenopile dataset, shown in Figure 5, give insight into 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 in plant pathogenic fungi and oomycetes, is also enriched [49]. Transport-related proteins, and in particular cation transport, are also prominently enriched in these terms. As an example, the role of the expanded calcium transporter genes in the oomycetes has been extensively studied in the context of host interaction [50]. 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 [51].

Selective pressures relate to lifestyles in oomycetes

The enriched terms common to the Albuginaceae and downy mildews greatly relate to known virulence factors for these plant pathogens, including carbohydrate metabolism, protein modification, transport, negative regulation of gene expression, and response to stimuli. 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 plant necrotrophs (Table 7, 8 and 9), highlighting a broader mechanism of adaptation to a plant-parasitic lifestyle.

One of the most often found terms and among the most enriched in both the obligate biotrophs and the hemibiotrophs of the dataset corresponds to regulation of biosynthetic and metabolic processes, and particularly negative regulation. This may underscore the fitness advantage for rapid growth during the hyphal stage and its need for activation or deactivation according to the circumstances. When the hyphal stage takes place after colonization, the salvaging and biosynthesis of carbohydrates, nucleic acids, and lipids with the resources obtained from the plant host is key for a successful infection. Beta-glucan, for example, is an important component of the oomycete's cell wall and is also an elicitor of the plant immune response [52]. Its biosynthesis features prominently in the enriched terms for the hemibiotrophs.

Secretion of small effector proteins, as in other fungal filamentous pathogens, is key for host adaptation in plant pathogenic oomycetes [53]. Many unique effector proteins have been characterized in the oomycetes that contribute to virulence by modulating the immune response of the plant [54]. Therefore, this dependence on the secretion machinery of the cell for successful infection and thus survival has led to high selective pressures on their genome. We observed significant enrichment of the effectors in the positively selected terms in all oomycetes of the dataset (hypergeometric test, $p < 0.01$). When looking at the enrichment per species, the majority of the *Phytophthora* and plant necrotrophs, which significantly depend on effector proteins for host infection, were also enriched (Figure 8). The obligate biotrophs, which also depend greatly on secreted effectors, do not show enrichment in our analysis. This may be due to the lack of orthologs on host specific effectors and thus not analyzed in the positive selection screen. There is a moderate correlation between the number of positively selected genes compared to those with predicted to be effectors (Pearson's correlation, $r = 0.43$, $p < 0.01$), so these results may be skewed due to testing bias (Figure 15). Surprisingly, most non-oomycete autotrophs show high enrichment in their predicted secreted proteins. In the GO enrichment of all oomycetes, there are several processes directly related to protein secretion under selection, including protein modification. Other secretion-related terms, although more general, also show enrichment, including those relating to microtubule-based processes in the obligate biotrophs, and transmembrane transport in the hemibiotrophs.

Another interesting term indirectly related to effector proteins is sulfur amino acid biosynthesis. This term is highly enriched in the hemibiotrophs and the necrotrophs of the dataset. This may be associated with the abundance of cysteine-rich proteins in the effector arsenal of the plant pathogens with a necrotroph phase [55]. The disulfide bonds that link cysteine residues help maintain the structural integrity of the proteins released into the extracellular space called apoplast, a hostile environment that is slightly acidic and rich in plant proteases [56].

When looking exclusively at the necrotroph groups, many terms in the plant pathogens overlap with the animal pathogens, most likely relating to their facultative saprobe lifestyle. These include glycolysis, generation of energy, cell communication, as well as amino acid, tetrapyrrole, and amide biosynthetic processes. The latter group is most likely enriched as a result of their autotrophic and more developed secondary metabolism compared to that of other oomycetes, which makes

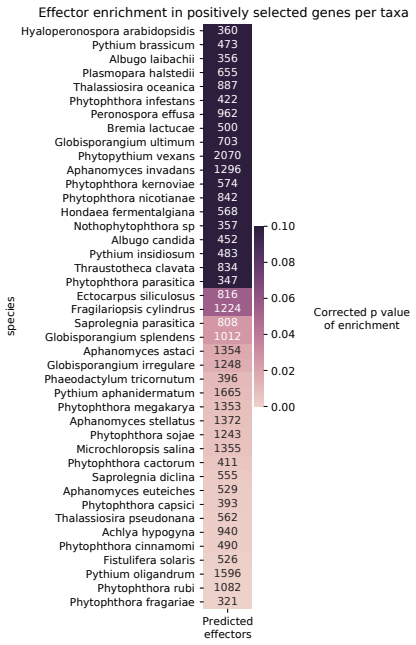

Figure 8 Enrichment of genes coding for effector proteins under positive selection. Color gradient represents significant p values from hypergeometric tests per taxa corrected for multiple testing using bonferroni. A lighter shade represents a more significant enrichment. The numbers within the cells represent the total identified effectors per proteome.

them suited for a free-living lifestyle [57]. Interesting is also the term DNA ligation involved in DNA repair, which may be related to the defense against oxidative stress that is key of the immune response in plants and animals against such pathogens [58]

Biosynthetic repertoire is important for lifestyle adaptation

As shown on Figure 1, the biosynthetic repertoire of each taxa plays a big role in defining the lifestyle of the organisms in the Stramenopile dataset. Particularly insteresting in oomycetes is the evolutionary history of sterol *de novo* biosynthesis. It is present in Saprolegniales and absent in other oomycete lineages due to their inability to synthesize oxidosqualene [59, 60]. The squalene synthase shows hints of positive selection in *Aphanomyces* (Figure 18). Furthermore, positive selection is pervasive in the enzymes that take part in sterol biosynthesis in the Stramenopile dataset.

Vitamin biosynthesis as well 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 capabilities and developing ways of utilizing host vitamin supply, also known as auxotrophy [61]. 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 19). 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 [62]. 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 [63]. In oomycete obligate biotrophs, molybdopterin-related biosynthetic pathways have been lost independently several times in the oomycetes lineage due to host adaptation [15]. Molybdopterin metabolism was found under high selective pressure in the facultative necrotrophs and autotrophs of the Stramenopile dataset, including *Saprolegniaceae* and *Pythiaceae* families, and *Phytophthora* genus (Figure 20). The biosynthesis of molybdopterin cofactor also features as an enriched GO term in the plant necrotrophs (Table 8 and 9).

Proteins relating to the glycolysis pathway and amino acid biosynthesis have a special evolutionary history in the oomycetes [64]. 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, as they need to be adapted to the new host. In the glycolysis pathway, we detected signatures of positive selection for most oomycetes in the Stramenopile dataset. Particularly in the enzymes glyceraldehyde-3-phosphate dehydrogenase and fructose-bisphosphate aldolase (Figure 21).

Protein family enrichment reflects lifestyle selective pressures

The large overrepresentation of paralogs as positively selected genes is evident in many of the taxa (Figure 3). After a gene duplication event occurs, 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 [65]. Interestingly, many of the enriched functions in the paralogs correlated with terms under positive selection for their specific lifestyle (Table 12). In the *Phytophthora* lineages these include biological regulation, glycolipid biosynthesis, and transmembrane transport. In *Albugo* and other obligate biotrophs, protein modification, carbohydrate metabolism, biological regulation, and glutamine metabolism.

A model based on genome properties 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. Assessment of lifestyle from genomic properties in plant pathogens has been traditionally done by characterizing cell wall-degrading enzyme annotations [66]. To our knowledge, there is only another published model that attempts to predict lifestyle from genomic features [67]. This model predicts trophic categories based on principal component analysis of carbohydrate-active enzyme annotations. We find that our model, which in contrast is based on the entire genome annotations, allows for a better overall accuracy. Furthermore, having trained the model on a larger number of features per sample allows for a more accurate prediction of incompletely annotated specimens that may result from environmental sampling. Given the availability of increasing proteomic and transcriptomic data of unknown fungal and oomycete origin, such prediction tools will become crucial to identify the pathogenic potential 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 [46, 68]. 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

We downloaded Stramenopile genetic data from the NCBI and FungiDB databases setting as cutoff assemblies with reported gene annotation, resulting in a dataset of 42 total proteomes. We screened the genomes using BUSCO for high abundance of key orthologs in the Stramenopile dataset as a form of quality control [69]. We performed functional annotation of the proteomes using InterProScan version 5.50-84.0 [70]. We annotated the effectors in the Stramenopile dataset by predicting secretion signal using the tool SignalP 5.0b followed by an annotation with the model EffectorO [71, 72]. We annotated the presence/absence of functional annotations from each genome with the Genome Properties database, performed the clustering with the Python package SciPy and visualized it with the package Seaborn [73, 74]. We compared UPGMA clusterings of the genome properties and genome properties with added positive selection information to the phylogenetic tree using the Robison-Foulds metric based on clusters with the application TreeCmp [75, 76].

Phylogeny inference

We constructed the concatenated Stramenopile tree using IQ-TREE 2 with automated partitioned model selection on inferred one-to-one orthogroups present in at least 25 of the taxa in the dataset [77]. We assessed full branch support in all nodes of the phylogenetic tree with 1,000 ultrafast bootstrap repetitions using the IQ-TREE 2 software and displayed it rooted on the outgroup of non-oomycetes.

Orthogroup classification and positive selection analyses

We developed a pipeline for whole genome positive selection analysis in Python using the Snakemake modular workflow framework [78]. It uses as input the coding nucleotide sequences as well as their corresponding predicted proteins from each proteome. The code and documentation are 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 [79–81]. 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 [82]. The first HYPHY algorithm used in the pipeline is FUBAR, a site-based program that scans the alignment for pervasive positive selection [83]. 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 [84]. Taxa downstream of nodes with a corrected p value of less than 0.05 were considered under positive selection for this particular gene.

Enrichment analyses

We used the Gene Ontology (GO) released in 2021-02-01 [85, 86]. We performed GO enrichment using the Python package Goatools based on the InterPro database annotations [87, 88]. 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. We used as a significance cutoff the negative base 10 logarithm of Holm-Bonferroni corrected p values that were higher than 1.3 (p value < 0.05). 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 [89]. The training dataset consisted of 319 unique proteomes from fungi and oomycete plant pathogens, and saprobes. 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 performed a stratified split of the dataset into training dataset, corresponding to 60% of the total, and optimization and validation datasets, each corresponding to half of the remaining 40%. Hyperparameter optimization, namely learning rate, activating functions and dense layer units, was carried out using Keras Tuner and its implementation of the Hyperband algorithm [90, 91].

Competing interests
The authors declare that they have no competing interests.

Author's contributions
D.G.P. performed the analyses, wrote the manuscript, and designed the figures. E.K contributed suggestions and reviewed the final manuscript.

Acknowledgments
The authors would like to thank Libera Lo Presti and the anonymous reviewers for their helpful 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 funding for the development of this manuscript.

Availability of Data and Materials
The datasets used and analysed during the current study are available from the corresponding author on reasonable request.

References

1. Zhang W, Zhang X, Li K, Wang C, Cai L, Zhuang W, et al. Introgression and gene family contraction drive the evolution of lifestyle and host shifts of hypocrealean fungi. Mycology. 2018;9(3):1–13.
2. 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).
3. 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.
4. Treangen TJ, Rocha EPC. Horizontal Transfer, Not Duplication, Drives the Expansion of Protein Families in Prokaryotes. PLoS Genetics. 2011;7(1):e1001284.
5. Sheridan PO, Raguideau S, Quince C, Holden J, Zhang L, Gaze WH, et al. Gene duplication drives genome expansion in a major lineage of Thaumarchaeota. Nature Communications. 2020;11(1):5494.
6. 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.
7. 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. Available from: `https://doi.org/10.1093/oxfordjournals.molbev.a025957`.
8. 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.
9. Kosiol C, Vinař T, Fonseca RRd, Hubisz MJ, Bustamante CD, Nielsen R, et al. Patterns of Positive Selection in Six Mammalian Genomes. PLoS Genetics. 2008;4(8):e1000144.
10. Cavalier-Smith T, Chao EEY. Phylogeny and Megasystematics of Phagotrophic Heterokonts (Kingdom Chromista). Journal of Molecular Evolution. 2006;62(4):388–420.
11. Beakes GW, Thines M. Handbook of the Protists; 2017.
12. Matari NH, Blair JE. A multilocus timescale for oomycete evolution estimated under three distinct molecular clock models. BMC Evolutionary Biology. 2014;14(1):101.
13. Bebber DP, Gurr SJ. Crop-destroying fungal and oomycete pathogens challenge food security. Fungal Genetics and Biology. 2015;74:62–64.
14. Derevnina L, Petre B, Kellner R, Dagdas YF, Sarowar MN, Giannakopoulou A, et al. Emerging oomycete threats to plants and animals. Philosophical Transactions of the Royal Society B: Biological Sciences. 2016;371(1709):20150459.
15. Kemen E, Jones JDG. Obligate biotroph parasitism: can we link genomes to lifestyles? Trends in Plant Science. 2012;17(8):448–457.
16. 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.
17. 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).
18. 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.
19. GAULIN E, JACQUET C, BOTTIN A, DUMAS B. Root rot disease of legumes caused by Aphanomyces euteiches. Molecular Plant Pathology. 2007;8(5):539–548.
20. 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. Available from: `https://academic.oup.com/gbe/article/7/1/120/601821`.
21. 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.
22. Weiblen C, Robe LJ, Azevedo MId, Ianiski LB, Stibbe PC, Ribeiro TC, et al. New insights on evolutionary aspects of Pythium insidiosum and other peronosporaleans. Mycoses. 2020;.
23. Thines M. An evolutionary framework for host shifts – jumping ships for survival. New Phytologist. 2019;224(2).
24. Haas BJ, Kamoun S, Zody MC, Jiang RHY, Handsaker RE, Cano LM, et al. Genome sequence and analysis of the Irish potato famine pathogen Phytophthora infestans. Nature. 2009;461(7262):393–398.
25. Richards TA, Soanes DM, Jones MDM, Vasieva O, Leonard G, Paszkiewicz K, et al. Horizontal gene transfer facilitated the evolution of plant parasitic mechanisms in the oomycetes. Proceedings of the National Academy of Sciences. 2011;108(37):15258–15263.
26. Savory F, Leonard G, Richards TA. The Role of Horizontal Gene Transfer in the Evolution of the Oomycetes. PLOS Pathogens. 2015;11(5):e1004805. Available from: `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4447370/`.
27. Savory EA, Fuller SL, Weisberg AJ, Thomas WJ, Gordon MI, Stevens DM, et al. Evolutionary transitions between beneficial and phytopathogenic Rhodococcus challenge disease management. eLife. 2017;6:e30925.
28. 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.
29. Latijnhouwers M, Wit PJGMd, Govers F. Oomycetes and fungi: similar weaponry to attack plants. Trends in Microbiology. 2003;11(10):462–469.
30. 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.
31. Basenko E, Pulman J, Shanmugasundram A, Harb O, Crouch K, Starns D, et al. FungiDB: An Integrated Bioinformatic Resource for Fungi and Oomycetes. Journal of Fungi. 2018;4(1):39.

32. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Research. 2016;44(D1):D733–D745.
33. Cock AWAMd, Lodhi AM, Rintoul TL, Bala K, Robideau GP, Abad ZG, et al. Phytopythium: molecular phylogeny and systematics. Persoonia - Molecular Phylogeny and Evolution of Fungi. 2015;34(1):25–39.
34. 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.
35. Samarasinghe H, You M, Jenkinson TS, Xu J, James TY. Hybridization Facilitates Adaptive Evolution in Two Major Fungal Pathogens. Genes. 2020;11(1):101.
36. 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.
37. Martens C, Peer YVd. The hidden duplication past of the plant pathogen Phytophthora and its consequences for infection. BMC Genomics. 2010;11(1):353.
38. Studholme DJ, Panda P, Stowasser ESV, González M, Hill R, Sambles C, et al. 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.
39. Tanaka T, Maeda Y, Veluchamy A, Tanaka M, Abida H, Maréchal E, et al. Oil accumulation by the oleaginous diatom Fistulifera solaris as revealed by the genome and transcriptome. The Plant cell. 2015;27(1):162–76.
40. Gaulin E, Pel MJC, Camborde L, San-Clemente H, Courbier S, Dupouy MA, et al. Genomics analysis of Aphanomyces spp. identifies a new class of oomycete effector associated with host adaptation. BMC Biology. 2018;16(1):43.
41. 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.
42. 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.
43. 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.
44. 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.
45. 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.
46. Rodenburg SYA, Ridder Dd, Govers F, Seidl MF. Oomycete metabolism is highly dynamic and reflects lifestyle adaptations. bioRxiv. 2020;p. 2020.02.12.941195.
47. LEWIS DH. CONCEPTS IN FUNGAL NUTRITION AND THE ORIGIN OF BIOTROPHY. Biological Reviews. 1973;48(2):261–277.
48. 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.
49. Wang E, Schornack S, Marsh J, Gobbato E, Schwessinger B, Eastmond P, et al. A Common Signaling Process that Promotes Mycorrhizal and Oomycete Colonization of Plants. Current Biology. 2012;22(23):2242–2246.
50. Zheng L, Mackrill JJ. Calcium Signaling in Oomycetes: An Evolutionary Perspective. Frontiers in Physiology. 2016;7:123.
51. Judelson HS. Dynamics and Innovations within Oomycete Genomes: Insights into Biology, Pathology, and Evolution. Eukaryotic Cell. 2012;11(11):1304–1312.
52. Klarzynski O, Plesse B, Joubert JM, Yvin JC, Kopp M, Kloareg B, et al. Linear β-1,3 Glucans Are Elicitors of Defense Responses in Tobacco. Plant Physiology. 2000;124(3):1027–1038.
53. Rocafort M, Fudal I, Mesarich CH. Apoplastic effector proteins of plant-associated fungi and oomycetes. Current Opinion in Plant Biology. 2020;56:9–19.
54. Wang Y, Tyler BM, Wang Y. Defense and Counterdefense During Plant-Pathogenic Oomycete Infection. Annual Review of Microbiology. 2019;73(1):1–30.
55. 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.
56. Idänheimo N, Gauthier A, Salojärvi J, Siligato R, Brosché M, Kollist H, et al. 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.
57. 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.
58. Fones H, Preston GM. Reactive oxygen and oxidative stress tolerance in plant pathogenic Pseudomonas. FEMS Microbiology Letters. 2012;327(1):1–8.
59. Gottlieb D. Differences in the Sterol Synthesizing Pathways of Sterol-Producing and Non-Sterol-Producing Fungi. Phytopathology. 1978;68(8):1168.
60. Gaulin E, Bottin A, Dumas B. Sterol biosynthesis in oomycete pathogens. Plant Signaling & Behavior. 2010;5(3):258–260.
61. 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.
62. Huennekens FM. Folic Acid Coenzymes in the Biosynthesis of Purines and Pyrimidines. Vitamins & Hormones. 1969;26:375–394.

63. Schwarz G, Mendel RR. MOLYBDENUM COFACTOR BIOSYNTHESIS AND MOLYBDENUM ENZYMES. Plant Biology. 2006;57(1):623–647.
64. Judelson HS. Metabolic Diversity and Novelties in the Oomycetes. Annual Review of Microbiology. 2016;71(1):1–19.
65. 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.
66. King BC, Waxman KD, Nenni NV, Walker LP, Bergstrom GC, Gibson DM. Arsenal of plant cell wall degrading enzymes reflects host preference among plant pathogenic fungi. Biotechnology for Biofuels. 2011;4(1):4.
67. 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.
68. Torruella G, de Mendoza A, Grau-Bové X, Antó M, Chaplin M, del Campo J, et al. Phylogenomics Reveals Convergent Evolution of Lifestyles in Close Relatives of Animals and Fungi. Current Biology. 2015;25(18):2404–2410.
69. Seppey M, Manni M, Zdobnov EM. Gene Prediction, Methods and Protocols. Methods in Molecular Biology. 2019;1962:227–245.
70. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30(9):1236–1240.
71. Petersen TN, Brunak S, Heijne Gv, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature Methods. 2011;8(10):785–786.
72. Nur M, Wood K, Michelmore R. EffectorO: motif-independent prediction of effectors in oomycete genomes using machine learning and lineage specificity. bioRxiv. 2021;Available from: `https://www.biorxiv.org/content/early/2021/03/20/2021.03.19.436227`.
73. Richardson LJ, Rawlings ND, Salazar GA, Almeida A, Haft DR, Ducq G, et al. 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.
74. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods. 2020;17:261–272.
75. Robinson DF, Foulds LR. Comparison of phylogenetic trees. Mathematical Biosciences. 1981;53(1-2):131–147.
76. 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.
77. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Haeseler Av, et al. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Molecular Biology and Evolution. 2020;37(5):1530–1534.
78. Köster J, Rahmann S. Snakemake—a scalable bioinformatics workflow engine. Bioinformatics. 2012;28(19):2520–2522.
79. Katoh K, Asimenos G, Toh H. Multiple alignment of DNA sequences with MAFFT. Methods in molecular biology (Clifton, NJ). 2009;537:39–64.
80. Price MN, Dehal PS, Arkin AP. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLoS ONE. 2010;5(3):e9490.
81. 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.
82. 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.
83. 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.
84. Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Pond SLK, et al. FUBAR: A Fast, Unconstrained Bayesian AppRoximation for Inferring Selection. Molecular Biology and Evolution. 2013;30(5):1196–1205.
85. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. Nature Genetics. 2000;25(1):25–29.
86. Carbon S, Douglass E, Dunn N, Good B, Harris NL, Lewis SE, et al. The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Research. 2018;47(D1):gky1055–.
87. Klopfenstein DV, Zhang L, Pedersen BS, Ramírez F, Vesztrocy AW, Naldi A, et al. GOATOOLS: A Python library for Gene Ontology analyses. Scientific Reports. 2018;8(1):10872.
88. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, et al. InterPro: the integrative protein signature database. Nucleic Acids Research. 2009;37(suppl_1):D211–D215.
89. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al.. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems; 2015. Software available from tensorflow.org. Available from: `http://tensorflow.org/`.
90. 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`.
91. 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. Available from: `http://jmlr.org/papers/v18/16-558.html`.
92. Cock JM, Sterck L, Rouzé P, Scornet D, Allen AE, Amoutzias G, et al. The Ectocarpus genome and the independent evolution of multicellularity in brown algae. Nature. 2010;465(7298):617–621.
93. 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.
94. Seddiki K, Godart F, Cigliano RA, Sanseverino W, Barakat M, Ortet P, et al. 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.
95. Ohan JA, Hovde BT, Zhang XL, Davenport KW, Chertkov O, Han C, et al. Nuclear Genome Assembly of the Microalga Nannochloropsis salina CCMP1776. Microbiology Resource Announcements. 2019;8(44).
96. Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A, et al. The Phaeodactylum genome reveals the evolutionary history of diatom genomes. Nature. 2008;456(7219):239–244.
97. Lommer M, Specht M, Roy AS, Kraemer L, Andreson R, Gutowska MA, et al. Genome and low-iron response of an oceanic diatom adapted to chronic iron limitation. Genome Biology. 2012;13(7):R66.
98. Jiang RHY, Bruijn Id, Haas BJ, Belmonte R, Löbach L, Christie J, et al. Distinctive Expansion of Potential Virulence Genes in the Genome of the Oomycete Fish Pathogen Saprolegnia parasitica. PLoS Genetics. 2013;9(6):e1003272.
99. Kemen E, Gardiner A, Schultz-Larsen T, Kemen AC, Balmuth AL, Robert-Seilaniantz A, et al. Gene Gain and Loss during Evolution of Obligate Parasitism in the White Rust Pathogen of Arabidopsis thaliana. PLoS Biology. 2011;9(7):e1001094. Available from: http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001094.
100. Fletcher K, Gil J, Bertier LD, Kenefick A, Wood KJ, Zhang L, et al. Genomic signatures of somatic hybrid vigor due to heterokaryosis in the oomycete pathogen, Bremia lactucae. bioRxiv. 2019;p. 516526.
101. Lévesque CA, Brouwer H, Cano L, Hamilton JP, Holt C, Huitema E, et al. Genome sequence of the necrotrophic plant pathogen Pythium ultimum reveals original pathogenicity mechanisms and effector repertoire. Genome Biology. 2010;11(7):R73.
102. Reghu RJ, Chellappan BV, Beena SH, Sasi A, Vasudevan SE, Nair AS. Draft Genome Sequence of the Oomycete Globisporangium splendens Strain rgcb-1. Microbiology Resource Announcements. 2020;9(16).
103. Baxter L, Tripathy S, Ishaque N, Boot N, Cabral A, Kemen E, et al. Signatures of Adaptation to Obligate Biotrophy in the Hyaloperonospora arabidopsidis Genome. Science. 2010;330(6010):1549–1551.
104. Armitage AD, Lysøe E, Nellist CF, Lewis LA, Cano LM, Harrison RJ, et al. Bioinformatic characterisation of the effector repertoire of the strawberry pathogen Phytophthora cactorum. PLOS ONE. 2018;13(10):e0202305.
105. Lamour KH, Mudge J, Gobena D, Hurtado-Gonzales OP, Schmutz J, Kuo A, 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.
106. Studholme DJ, McDougal RL, Sambles C, Hansen E, Hardy G, Grant M, et al. Genome sequences of six Phytophthora species associated with forests in New Zealand. Genomics Data. 2016;7:54–56.
107. Adams TM, Armitage AD, Sobczyk MK, Bates HJ, Tabima JF, Kronmiller BA, et al. Genomic Investigation of the Strawberry Pathogen Phytophthora fragariae Indicates Pathogenicity Is Associated With Transcriptional Variation in Three Key Races. Frontiers in Microbiology. 2020;11:490.
108. Ali SS, Shao J, Lary DJ, Kronmiller BA, Shen D, Strem MD, et al. Phytophthora megakarya and Phytophthora palmivora, Closely Related Causal Agents of Cacao Black Pod Rot, Underwent Increases in Genome Sizes and Gene Numbers by Different Mechanisms. Genome Biology and Evolution. 2017;9(3):536–557.
109. Liu H, Ma X, Yu H, Fang D, Li Y, Wang X, et al. Genomes and virulence difference between two physiological races of Phytophthora nicotianae. GigaScience. 2016;5(1):3.
110. Tyler BM, Tripathy S, Zhang X, Dehal P, Jiang RHY, Aerts A, et al. ¡i¿Phytophthora¡/i¿ Genome Sequences Uncover Evolutionary Origins and Mechanisms of Pathogenesis. Science. 2006;313(5791):1261–1266.
111. Rujirawat T, Patumcharoenpol P, Lohnoo T, Yingyong W, Lerksuthirat T, Tangphatsornruang S, et al. 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.
112. Faure C, Veyssière M, Boëlle B, San Clemente H, Bouchez O, Lopez-Roques C, et al. Long-Read Genome Sequence of the Sugar Beet Rhizosphere Mycoparasite Pythium oligandrum. G3: Genes, Genomes, Genetics. 2020;10(2):431–436.

Supplementary Figures

Supplementary tables

Table 2 Stramenopile genomes dataset used for positive selective analyses.

Phylogenetic family	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	[92]
	Fistulifera solaris	GCA_002217885.1	Autotroph	97	14	83	[39]
	Fragilariopsis cylindrus	GCA_001750085.1	Autotroph	95	95	0	[93]
	Hondaea fermentalgiana	GCA_002897355.1	Autotroph	95	90	5	[94]
	Microchloropsis salina	GCA_004565275.1	Autotroph	92	90	2	[95]
	Phaeodactylum tricornutum	GCA_000150955.2	Autotroph	97	95	2	[96]
	Thalassiosira oceanica	GCA_000296195.2	Autotroph	90	90	0	[97]
	Thalassiosira pseudonana	GCA_000149405.2	Autotroph	97	95	2	[96]
Saprolegniaceae	Achlya hypogyna	GCA_002081595.1	Animal necrotroph	99	98	1	[20]
	Aphanomyces astaci	GCA_000520075.1	Animal necrotroph	100	82	18	
	Aphanomyces euteiches	GCA_009835175.1	Plant necrotroph	99	99	0	
	Aphanomyces invadans	GCA_000520115.1	Animal necrotroph	100	83	17	
	Aphanomyces stellatus	GCA_009835185.1	Plant necrotroph	97	96	1	
	Saprolegnia diclina	GCA_000281045.1	Animal necrotroph	99	98	1	
	Saprolegnia parasitica	GCA_000151545.2	Animal necrotroph	99	99	0	[98]
	Thraustotheca clavata	GCA_002081575.1	Free-living saprotroph	99	98	1	[20]
Albuginaceae	Albugo candida	GCA_001078535.1	Obligate biotroph	98	86	12	
	Albugo laibachii	PRJEA53219	Obligate biotroph	95	82	13	[99]
Peronosporaceae	Bremia lactucae	GCA_004359215.1	Obligate biotroph	96	90	6	[100]
	Globisporangium irregulare	GCA_000387425.2	Plant necrotroph	98	96	2	[101]
	Globisporangium splendens	GCA_006386115.1	Plant necrotroph	91	74	17	[102]
	Globisporangium ultimum	GCA_000143045.1	Plant necrotroph	94	93	1	[101]
	Hyaloperonospora arabidopsidis	GCA_000173235.2	Obligate biotroph	89	82	7	[103]
	Nothophytophthora sp.	GCA_001712635.2		90	28	62	[38]
	Peronospora effusa	GCA_003843895.1	Obligate biotroph	94	93	1	
	Phytophthora cactorum	GCA_003287315.1	Hemibiotroph	100	98	2	[104]
	Phytophthora capsici	GCA_000325885.1	Hemibiotroph	98	97	1	[105]
	Phytophthora cinnamomi	GCA_001314365.1	Hemibiotroph	96	94	2	[106]
	Phytophthora fragariae	GCA_009729455.1	Hemibiotroph	94	93	1	[107]
	Phytophthora infestans	GCA_000142945.1	Hemibiotroph	100	99	1	
	Phytophthora kernoviae	GCA_001712645.2	Hemibiotroph	96	96	0	[38]
	Phytophthora megakarya	GCA_002215365.1	Hemibiotroph	91	90	1	[108]
	Phytophthora nicotianae	GCA_001483015.1	Hemibiotroph	99	86	13	[109]
	Phytophthora parasitica	GCA_000247585.2	Hemibiotroph	98	87	11	
	Phytophthora rubi	GCA_009733145.1	Hemibiotroph	100	98	2	[107]
	Phytophthora sojae	GCA_000149755.2	Hemibiotroph	99	98	1	[110]
	Phytopythium vexans	GCA_000387545.2	Plant necrotroph	94	92	2	[17]
	Pythium brassicum	GCA_008271595.1	Plant necrotroph	100	99	1	
	Plasmopara halstedii	GCA_900000015.1	Obligate biotroph	100	100	0	
Pythiaceae	Pythium aphanidermatum	GCA_000387445.2	Plant necrotroph	94	93	1	[17]
	Pythium insidiosum	GCA_001029375.1	Animal necrotroph	99	87	12	[111]
	Pythium oligandrum	GCA_005966545.1	Fungal necrotroph	100	100	0	[112]

Table 3 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

Table 4 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
Aphanomyces species	2	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 species	3	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	38	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
Pythium species	2	N
Ramularia collo-cygni	1	H
Rhizoctonia solani	7	N
Rhizopus delemar	1	S
Saccharomyces cerevisiae	60	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
Tilletiaria anomala	1	B
Trametes versicolor	1	S
Tremella mesenterica	2	B
Trichoderma species	7	S
Uncinocarpus reesii	1	S
Ustilaginoidea virens	2	B
Ustilago species	3	B
Venturia inaequalis	4	H
Verticillium dahliae	10	H
Yarrowia lipolytica	13	S
Zymoseptoria species	6	H

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

Table 5 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

Table 6 Significantly 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:2000113	negative regulation of cellular macromolecule biosynthetic process	12/2535	54/75580	7	3.44
GO:0043648	dicarboxylic acid metabolic process	16/2535	77/75580	6	3.3
GO:0051253	negative regulation of RNA metabolic process	11/2535	47/75580	7	3.18
GO:0031324	negative regulation of cellular metabolic process	14/2535	70/75580	5	2.98
GO:0008033	tRNA processing	33/2535	289/75580	9	2.89
GO:0016053	organic acid biosynthetic process	38/2535	453/75580	4	2.69
GO:0018193	peptidyl-amino acid modification	29/2535	315/75580	7	2.63
GO:0010605	negative regulation of macromolecule metabolic process	21/2535	172/75580	5	2.52
GO:0006399	tRNA metabolic process	49/2535	597/75580	8	2.49
GO:0009064	glutamine family amino acid metabolic process	16/2535	118/75580	8	2.35
GO:0007018	microtubule-based movement	35/2535	407/75580	3	2.35
GO:0006082	organic acid metabolic process	79/2535	1249/75580	3	2.34
GO:0006396	RNA processing	80/2535	1217/75580	7	2.3
GO:0006468	protein phosphorylation	97/2535	1582/75580	7	2.26
GO:0034637	cellular carbohydrate biosynthetic process	16/2535	120/75580	5	2.26
GO:0016310	phosphorylation	110/2535	1819/75580	5	2.17
GO:0043412	macromolecule modification	206/2535	3434/75580	4	2.13
GO:0019752	carboxylic acid metabolic process	72/2535	1134/75580	5	2.13
GO:0060255	regulation of macromolecule metabolic process	67/2535	1054/75580	4	2.13
GO:0007017	microtubule-based process	41/2535	556/75580	2	2.1
GO:0006725	cellular aromatic compound metabolic process	212/2535	4200/75580	3	2.08
GO:1901360	organic cyclic compound metabolic process	220/2535	4285/75580	3	2.08
GO:0006464	cellular protein modification process	178/2535	3032/75580	6	2.08
GO:0009058	biosynthetic process	168/2535	3331/75580	2	2.07
GO:0044267	cellular protein metabolic process	208/2535	3721/75580	5	2.07
GO:0046483	heterocycle metabolic process	215/2535	4214/75580	3	2.07
GO:0034470	ncRNA processing	41/2535	560/75580	8	2.05
GO:0034660	ncRNA metabolic process	57/2535	880/75580	7	2.05
GO:0090304	nucleic acid metabolic process	168/2535	3199/75580	5	2.05
GO:0019538	protein metabolic process	245/2535	4869/75580	4	2.04
GO:0050789	regulation of biological process	111/2535	2041/75580	2	2.04
GO:0044249	cellular biosynthetic process	151/2535	3012/75580	3	2.01
GO:0006139	nucleobase-containing compound metabolic process	192/2535	3938/75580	4	2.0
GO:0016070	RNA metabolic process	122/2535	2310/75580	6	1.97
GO:0044260	cellular macromolecule metabolic process	302/2535	5672/75580	4	1.95
GO:1901566	organonitrogen compound biosynthetic process	91/2535	1648/75580	4	1.94
GO:0034641	cellular nitrogen compound metabolic process	245/2535	4998/75580	3	1.93
GO:1901564	organonitrogen compound metabolic process	340/2535	6708/75580	3	1.92
GO:0043170	macromolecule metabolic process	432/2535	8280/75580	3	1.89
GO:0006807	nitrogen compound metabolic process	500/2535	9762/75580	2	1.87
GO:0044237	cellular metabolic process	541/2535	10414/75580	2	1.86
GO:0071704	organic substance metabolic process	610/2535	11484/75580	2	1.84
GO:0044238	primary metabolic process	554/2535	10641/75580	2	1.84
GO:0008152	metabolic process	643/2535	12234/75580	1	1.83
GO:0046394	carboxylic acid biosynthetic process	31/2535	376/75580	6	1.83
GO:0031323	regulation of cellular metabolic process	58/2535	932/75580	4	1.77
GO:0045892	negative regulation of transcription, DNA-templated	9/2535	43/75580	10	1.69
GO:2000112	regulation of cellular macromolecule biosynthetic process	52/2535	816/75580	6	1.59
GO:0010468	regulation of gene expression	58/2535	936/75580	5	1.58
GO:0065007	biological regulation	113/2535	2204/75580	1	1.57
GO:1901576	organic substance biosynthetic process	150/2535	3124/75580	3	1.54
GO:0005975	carbohydrate metabolic process	49/2535	755/75580	3	1.54
GO:0050794	regulation of cellular process	99/2535	1887/75580	3	1.48
GO:0009086	methionine biosynthetic process	7/2535	26/75580	10	1.47
GO:0006520	cellular amino acid metabolic process	51/2535	803/75580	6	1.47
GO:0080090	regulation of primary metabolic process	57/2535	927/75580	4	1.44
GO:0044283	small molecule biosynthetic process	45/2535	674/75580	3	1.42
GO:0008652	cellular amino acid biosynthetic process	25/2535	291/75580	7	1.41
GO:1901605	alpha-amino acid metabolic process	29/2535	364/75580	7	1.35

Table 7 Significantly enriched terms relating to biological processes in the positively selected hemibiotroph proteins.

GO number	Name	Ratio in study	Ratio in population	Depth	-log10 of p value
GO:0009086	methionine biosynthetic process	13/6255	64/222540	10	3.38
GO:0051274	beta-glucan biosynthetic process	18/6255	104/222540	8	3.31
GO:0051253	negative regulation of RNA metabolic process	16/6255	91/222540	7	3.26
GO:0043648	dicarboxylic acid metabolic process	20/6255	150/222540	6	3.25
GO:0009082	branched-chain amino acid biosynthetic process	18/6255	81/222540	5	3.06
GO:0010605	negative regulation of macromolecule metabolic process	34/6255	300/222540	5	2.96
GO:0034637	cellular carbohydrate biosynthetic process	29/6255	226/222540	5	2.94
GO:0031324	negative regulation of cellular metabolic process	20/6255	129/222540	5	2.79
GO:0009312	oligosaccharide biosynthetic process	19/6255	146/222540	5	2.74
GO:2000113	negative regulation of cellular macromolecule biosynthetic process	15/6255	101/222540	7	2.7
GO:0044262	cellular carbohydrate metabolic process	35/6255	455/222540	4	2.67
GO:0016051	carbohydrate biosynthetic process	37/6255	333/222540	4	2.61
GO:0000097	sulfur amino acid biosynthetic process	15/6255	121/222540	5	2.46
GO:0007017	microtubule-based process	72/6255	1355/222540	2	2.39
GO:0006355	regulation of transcription, DNA-templated	75/6255	1176/222540	9	2.39
GO:2000112	regulation of cellular macromolecule biosynthetic process	84/6255	1307/222540	6	2.39
GO:0051273	beta-glucan metabolic process	19/6255	190/222540	7	2.37
GO:0006396	RNA processing	95/6255	1887/222540	7	2.33
GO:0016071	mRNA metabolic process	52/6255	831/222540	7	2.32
GO:0051171	regulation of nitrogen compound metabolic process	95/6255	1522/222540	4	2.31
GO:0006974	cellular response to DNA damage stimulus	68/6255	1193/222540	4	2.29
GO:0010468	regulation of gene expression	104/6255	1526/222540	5	2.28
GO:0006950	response to stress	88/6255	1528/222540	2	2.27
GO:0006397	mRNA processing	42/6255	664/222540	8	2.27
GO:0019222	regulation of metabolic process	118/6255	1794/222540	3	2.27
GO:0051252	regulation of RNA metabolic process	81/6255	1206/222540	6	2.26
GO:0060255	regulation of macromolecule metabolic process	117/6255	1749/222540	4	2.25
GO:0033554	cellular response to stress	72/6255	1253/222540	3	2.24
GO:0051716	cellular response to stimulus	73/6255	1254/222540	2	2.24
GO:0050896	response to stimulus	96/6255	1605/222540	1	2.22
GO:0031323	regulation of cellular metabolic process	100/6255	1555/222540	4	2.21
GO:0005975	carbohydrate metabolic process	156/6255	2565/222540	3	2.19
GO:0050789	regulation of biological process	172/6255	3715/222540	2	2.12
GO:0048519	negative regulation of biological process	35/6255	527/222540	3	2.11
GO:0050794	regulation of cellular process	154/6255	3447/222540	3	2.05
GO:0044249	cellular biosynthetic process	208/6255	5071/222540	3	2.01
GO:0065007	biological regulation	178/6255	4153/222540	1	2.01
GO:0016070	RNA metabolic process	148/6255	3483/222540	6	2.01
GO:0006468	protein phosphorylation	215/6255	4156/222540	7	2.0
GO:0016310	phosphorylation	235/6255	4699/222540	5	2.0
GO:0009058	biosynthetic process	230/6255	5732/222540	2	1.96
GO:0006793	phosphorus metabolic process	292/6255	6981/222540	3	1.94
GO:0043412	macromolecule modification	366/6255	7831/222540	4	1.94
GO:0006464	cellular protein modification process	327/6255	7218/222540	6	1.94
GO:0044267	cellular protein metabolic process	372/6255	8250/222540	5	1.92
GO:0006796	phosphate-containing compound metabolic process	291/6255	6936/222540	4	1.9
GO:0019538	protein metabolic process	459/6255	11568/222540	4	1.84
GO:0034645	cellular macromolecule biosynthetic process	92/6255	1977/222540	5	1.82
GO:1901576	organic substance biosynthetic process	209/6255	5300/222540	3	1.82
GO:0044260	cellular macromolecule metabolic process	524/6255	13650/222540	4	1.81
GO:1901564	organonitrogen compound metabolic process	558/6255	15181/222540	3	1.79
GO:0043170	macromolecule metabolic process	713/6255	19213/222540	3	1.74
GO:0044237	cellular metabolic process	841/6255	22391/222540	2	1.71
GO:0006807	nitrogen compound metabolic process	788/6255	22412/222540	2	1.7
GO:0044238	primary metabolic process	941/6255	25043/222540	2	1.69
GO:0071704	organic substance metabolic process	1007/6255	26763/222540	2	1.67
GO:0008152	metabolic process	1069/6255	28228/222540	1	1.67
GO:0055085	transmembrane transport	202/6255	5180/222540	4	1.66
GO:0006813	potassium ion transport	25/6255	336/222540	6	1.6
GO:0045892	negative regulation of transcription, DNA-templated	11/6255	82/222540	10	1.43
GO:0009059	macromolecule biosynthetic process	99/6255	2252/222540	4	1.39

Table 8 Enriched terms relating to biological processes in the positively selected plant necrotrophs.

GO number	Name	Ratio in study	Ratio in population	Depth	-log10 of p value
GO:0006190	inosine salvage	8/9880	8/129511	11	3.63
GO:0009088	threonine biosynthetic process	8/9880	8/129511	10	3.63
GO:0006425	glutaminyl-tRNA aminoacylation	8/9880	8/129511	11	3.63
GO:1901031	regulation of response to reactive oxygen species	9/9880	11/129511	6	3.5
GO:0046168	glycerol-3-phosphate catabolic process	13/9880	18/129511	7	3.47
GO:0072350	tricarboxylic acid metabolic process	17/9880	32/129511	6	3.28
GO:0006537	glutamate biosynthetic process	14/9880	15/129511	10	3.23
GO:0006166	purine ribonucleoside salvage	14/9880	23/129511	10	3.22
GO:0009084	glutamine family amino acid biosynthetic process	27/9880	86/129511	9	3.14
GO:0005992	trehalose biosynthetic process	26/9880	73/129511	7	3.13
GO:0009082	branched-chain amino acid biosynthetic process	31/9880	67/129511	5	3.1
GO:0060271	cilium assembly	30/9880	122/129511	7	3.06
GO:0009064	glutamine family amino acid metabolic process	47/9880	153/129511	8	2.99
GO:0000096	sulfur amino acid metabolic process	28/9880	123/129511	4	2.98
GO:0006144	purine nucleobase metabolic process	18/9880	50/129511	7	2.97
GO:0006555	methionine metabolic process	19/9880	56/129511	9	2.96
GO:0006536	glutamate metabolic process	22/9880	39/129511	9	2.96
GO:0033014	tetrapyrrole biosynthetic process	29/9880	109/129511	5	2.95
GO:0051274	beta-glucan biosynthetic process	29/9880	75/129511	8	2.93
GO:0016573	histone acetylation	22/9880	75/129511	11	2.93
GO:0003341	cilium movement	25/9880	75/129511	4	2.93
GO:0001522	pseudouridine synthesis	45/9880	154/129511	8	2.93
GO:0071897	DNA biosynthetic process	18/9880	34/129511	7	2.92
GO:0006816	calcium ion transport	27/9880	124/129511	7	2.91
GO:0006102	isocitrate metabolic process	9/9880	16/129511	7	2.91
GO:0000097	sulfur amino acid biosynthetic process	21/9880	82/129511	5	2.89
GO:0032012	regulation of ARF protein signal transduction	16/9880	45/129511	9	2.88
GO:0030488	tRNA methylation	17/9880	51/129511	11	2.85
GO:0007154	cell communication	28/9880	76/129511	2	2.83
GO:0018205	peptidyl-lysine modification	32/9880	148/129511	8	2.82
GO:0017038	protein import	33/9880	133/129511	9	2.8
GO:0006606	protein import into nucleus	25/9880	83/129511	10	2.8
GO:0051056	regulation of small GTPase mediated signal transduction	28/9880	70/129511	7	2.78
GO:0006414	translational elongation	28/9880	90/129511	6	2.78
GO:0006075	(1->3)-beta-D-glucan biosynthetic process	22/9880	64/129511	9	2.75
GO:0009086	methionine biosynthetic process	18/9880	46/129511	10	2.74
GO:0034637	cellular carbohydrate biosynthetic process	56/9880	165/129511	5	2.73
GO:0006525	arginine metabolic process	19/9880	58/129511	9	2.72
GO:0016051	carbohydrate biosynthetic process	61/9880	238/129511	4	2.72
GO:0001510	RNA methylation	43/9880	205/129511	8	2.72
GO:0006400	tRNA modification	55/9880	228/129511	10	2.71
GO:0009250	glucan biosynthetic process	30/9880	84/129511	7	2.71
GO:0051273	beta-glucan metabolic process	30/9880	150/129511	7	2.69
GO:0006096	glycolytic process	41/9880	150/129511	12	2.69
GO:2000113	negative regulation of cellular macromolecule biosynthetic process	20/9880	71/129511	7	2.69
GO:0070925	organelle assembly	41/9880	197/129511	5	2.68
GO:0043648	dicarboxylic acid metabolic process	34/9880	120/129511	6	2.67
GO:0043547	positive regulation of GTPase activity	15/9880	21/129511	6	2.67
GO:0010605	negative regulation of macromolecule metabolic process	52/9880	240/129511	5	2.64
GO:0009312	oligosaccharide biosynthetic process	31/9880	106/129511	5	2.63
GO:0006073	cellular glucan metabolic process	31/9880	159/129511	6	2.62
GO:0016570	histone modification	39/9880	199/129511	7	2.6
GO:1901615	organic hydroxy compound metabolic process	35/9880	190/129511	3	2.59
GO:0006091	generation of precursor metabolites and energy	57/9880	294/129511	3	2.55
GO:0006401	RNA catabolic process	38/9880	181/129511	7	2.53
GO:0072330	monocarboxylic acid biosynthetic process	28/9880	138/129511	7	2.48
GO:1901607	alpha-amino acid biosynthetic process	57/9880	356/129511	8	2.47
GO:0006418	tRNA aminoacylation for protein translation	56/9880	328/129511	10	2.46
GO:0048583	regulation of response to stimulus	51/9880	300/129511	3	2.46
GO:0016052	carbohydrate catabolic process	51/9880	261/129511	4	2.45
GO:0032259	methylation	72/9880	386/129511	2	2.45
GO:0008033	tRNA processing	96/9880	403/129511	9	2.45
GO:0034472	snRNA 3'-end processing	10/9880	22/129511	10	2.45
GO:0072594	establishment of protein localization to organelle	43/9880	251/129511	6	2.44
GO:0006457	protein folding	60/9880	404/129511	2	2.43
GO:0044262	cellular carbohydrate metabolic process	66/9880	330/129511	4	2.43
GO:0046034	ATP metabolic process	43/9880	227/129511	2	2.43
GO:0072521	purine-containing compound metabolic process	74/9880	488/129511	4	2.42
GO:0031324	negative regulation of cellular metabolic process	22/9880	95/129511	5	2.37
GO:0043414	macromolecule methylation	52/9880	320/129511	5	2.34
GO:0008652	cellular amino acid biosynthetic process	79/9880	427/129511	7	2.33
GO:0032787	monocarboxylic acid metabolic process	81/9880	463/129511	6	2.32
GO:0034470	ncRNA processing	140/9880	735/129511	8	2.3
GO:0006813	potassium ion transport	49/9880	309/129511	6	2.29
GO:0018193	peptidyl-amino acid modification	85/9880	518/129511	7	2.28
GO:0046394	carboxylic acid biosynthetic process	107/9880	573/129511	6	2.28
GO:1901605	alpha-amino acid metabolic process	87/9880	557/129511	7	2.26
GO:0009451	RNA modification	120/9880	540/129511	7	2.26
GO:0016053	organic acid biosynthetic process	144/9880	682/129511	4	2.23
GO:0006310	DNA recombination	40/9880	233/129511	7	2.22
GO:0034660	ncRNA metabolic process	201/9880	1090/129511	7	2.22
GO:0034613	cellular protein localization	44/9880	274/129511	4	2.22
GO:0006399	tRNA metabolic process	157/9880	748/129511	8	2.21
GO:0007018	microtubule-based movement	116/9880	839/129511	3	2.21
GO:0006814	sodium ion transport	17/9880	63/129511	6	2.19
GO:0006520	cellular amino acid metabolic process	167/9880	1097/129511	6	2.19
GO:0006355	regulation of transcription, DNA-templated	145/9880	1045/129511	9	2.18
GO:0044283	small molecule biosynthetic process	179/9880	1074/129511	3	2.18
GO:0006812	cation transport	162/9880	1214/129511	5	2.17
GO:0007017	microtubule-based process	141/9880	1103/129511	2	2.16
GO:0080090	regulation of primary metabolic process	174/9880	1307/129511	4	2.15
GO:0010468	regulation of gene expression	183/9880	1307/129511	5	2.15
GO:1901575	organic substance catabolic process	189/9880	1657/129511	3	2.15
GO:0051252	regulation of RNA metabolic process	147/9880	1056/129511	6	2.14
GO:0006629	lipid metabolic process	197/9880	1627/129511	3	2.14
GO:0000398	mRNA splicing, via spliceosome	50/9880	321/129511	11	2.12
GO:0005975	carbohydrate metabolic process	200/9880	1738/129511	3	2.11
GO:0006066	alcohol metabolic process	25/9880	121/129511	4	2.1
GO:0006082	organic acid metabolic process	321/9880	1814/129511	3	2.1
GO:0006396	RNA processing	263/9880	1641/129511	7	2.1
GO:0031323	regulation of cellular metabolic process	177/9880	1325/129511	4	2.09

Table 9 Enriched terms relating to biological processes in the positively selected plant necrotrophs (continued).

GO number	Name	Ratio in study	Ratio in population	Depth	-log10 of p value
GO:2000112	regulation of cellular macromolecule biosynthetic process	157/9880	1151/129511	6	2.08
GO:0005976	polysaccharide metabolic process	32/9880	175/129511	4	2.08
GO:0019222	regulation of metabolic process	215/9880	1513/129511	3	2.08
GO:0060255	regulation of macromolecule metabolic process	210/9880	1484/129511	4	2.07
GO:0031047	gene silencing by RNA	8/9880	15/129511	8	2.07
GO:0034645	cellular macromolecule biosynthetic process	217/9880	1621/129511	5	2.06
GO:0043604	amide biosynthetic process	114/9880	946/129511	5	2.04
GO:0016043	cellular component organization	226/9880	2030/129511	3	2.02
GO:0016070	RNA metabolic process	453/9880	2964/129511	6	2.01
GO:0009056	catabolic process	208/9880	1779/129511	2	2.0
GO:0050896	response to stimulus	141/9880	1228/129511	1	2.0
GO:0043087	regulation of GTPase activity	17/9880	65/129511	5	2.0
GO:0071840	cellular component organization or biogenesis	233/9880	2151/129511	2	2.0
GO:0009059	macromolecule biosynthetic process	231/9880	1856/129511	4	1.99
GO:0022607	cellular component assembly	110/9880	906/129511	4	1.99
GO:0019752	carboxylic acid metabolic process	270/9880	1645/129511	5	1.99
GO:1901566	organonitrogen compound biosynthetic process	319/9880	2320/129511	4	1.99
GO:0050789	regulation of biological process	385/9880	3249/129511	2	1.94
GO:0009112	nucleobase metabolic process	24/9880	116/129511	6	1.93
GO:0044281	small molecule metabolic process	405/9880	3119/129511	2	1.92
GO:0006885	regulation of pH	15/9880	53/129511	8	1.91
GO:0065007	biological regulation	409/9880	3527/129511	1	1.91
GO:0050794	regulation of cellular process	337/9880	3032/129511	3	1.91
GO:0046148	pigment biosynthetic process	17/9880	66/129511	3	1.91
GO:0050790	regulation of catalytic activity	38/9880	227/129511	3	1.88
GO:0006811	ion transport	328/9880	3267/129511	4	1.85
GO:0044249	cellular biosynthetic process	543/9880	4181/129511	3	1.85
GO:0090304	nucleic acid metabolic process	569/9880	5033/129511	5	1.84
GO:1901576	organic substance biosynthetic process	564/9880	4428/129511	3	1.83
GO:0006464	cellular protein modification process	551/9880	5511/129511	6	1.82
GO:0009058	biosynthetic process	613/9880	4761/129511	2	1.81
GO:0006793	phosphorus metabolic process	512/9880	5401/129511	3	1.8
GO:0006419	alanyl-tRNA aminoacylation	8/9880	16/129511	11	1.8
GO:0006101	citrate metabolic process	8/9880	16/129511	7	1.8
GO:0008612	peptidyl-lysine modification to peptidyl-hypusine	8/9880	16/129511	9	1.8
GO:0006423	cysteinyl-tRNA aminoacylation	8/9880	16/129511	11	1.8
GO:0044267	cellular protein metabolic process	655/9880	6393/129511	5	1.79
GO:0044271	cellular nitrogen compound biosynthetic process	265/9880	2604/129511	4	1.79
GO:0006796	phosphate-containing compound metabolic process	506/9880	5361/129511	4	1.79
GO:0043412	macromolecule modification	672/9880	6051/129511	4	1.78
GO:0006139	nucleobase-containing compound metabolic process	688/9880	6267/129511	4	1.78
GO:0006777	Mo-molybdopterin cofactor biosynthetic process	13/9880	42/129511	7	1.77
GO:1901360	organic cyclic compound metabolic process	786/9880	6882/129511	3	1.77
GO:0006725	cellular aromatic compound metabolic process	753/9880	6732/129511	3	1.77
GO:0034641	cellular nitrogen compound metabolic process	877/9880	7987/129511	3	1.74
GO:0046483	heterocycle metabolic process	759/9880	6710/129511	3	1.74
GO:0010629	negative regulation of gene expression	30/9880	166/129511	6	1.71
GO:0016071	mRNA metabolic process	84/9880	674/129511	7	1.69
GO:0044260	cellular macromolecule metabolic process	972/9880	9893/129511	4	1.69
GO:1901564	organonitrogen compound metabolic process	1213/9880	12929/129511	3	1.63
GO:0043170	macromolecule metabolic process	1472/9880	15069/129511	3	1.62
GO:0016310	phosphorylation	341/9880	3524/129511	5	1.6
GO:0006807	nitrogen compound metabolic process	1772/9880	17828/129511	2	1.59
GO:0044237	cellular metabolic process	1906/9880	17064/129511	2	1.57
GO:0009098	leucine biosynthetic process	8/9880	17/129511	8	1.57
GO:0006435	threonyl-tRNA aminoacylation	8/9880	17/129511	11	1.57
GO:0044238	primary metabolic process	1962/9880	19530/129511	2	1.57
GO:0008152	metabolic process	2337/9880	22313/129511	1	1.56
GO:0071704	organic substance metabolic process	2192/9880	21160/129511	2	1.54
GO:0048519	negative regulation of biological process	57/9880	414/129511	3	1.53
GO:0033554	cellular response to stress	108/9880	929/129511	3	1.51
GO:0006325	chromatin organization	47/9880	323/129511	4	1.42
GO:1901136	carbohydrate derivative catabolic process	18/9880	79/129511	4	1.39
GO:0006950	response to stress	129/9880	1162/129511	2	1.38
GO:0042364	water-soluble vitamin biosynthetic process	28/9880	156/129511	5	1.37
GO:0019243	methylglyoxal catabolic process to D-lactate via S-lactoyl-glutathione	8/9880	18/129511	9	1.35
GO:0019310	inositol catabolic process	8/9880	18/129511	7	1.35
GO:0018344	protein geranylgeranylation	8/9880	18/129511	8	1.35
GO:0006566	threonine metabolic process	11/9880	34/129511	9	1.3

Table 10 enriched terms relating to biological processes in the positively selected animal necrotrophs.

go number	name	ratio in study	ratio in population	depth	-log10 of p value
GO:0006190	inosine salvage	8/7214	8/114793	11	4.22
GO:0032955	regulation of division septum assembly	6/7214	6/114793	8	3.9
GO:0036159	inner dynein arm assembly	6/7214	6/114793	8	3.9
GO:0046168	glycerol-3-phosphate catabolic process	12/7214	20/114793	7	3.7
GO:0015940	pantothenate biosynthetic process	15/7214	20/114793	8	3.7
GO:0072350	tricarboxylic acid metabolic process	13/7214	26/114793	6	3.41
GO:2000113	negative regulation of cellular macromolecule biosynthetic process	17/7214	50/114793	7	3.29
GO:0002098	tRNA wobble uridine modification	18/7214	50/114793	12	3.29
GO:0030488	tRNA methylation	15/7214	44/114793	11	3.24
GO:0051253	negative regulation of RNA metabolic process	17/7214	44/114793	7	3.24
GO:0003341	cilium movement	24/7214	57/114793	4	3.24
GO:0006536	glutamate metabolic process	15/7214	38/114793	9	3.23
GO:0009086	methionine biosynthetic process	14/7214	33/114793	10	3.11
GO:0006414	translational elongation	21/7214	80/114793	6	3.09
GO:0071897	DNA biosynthetic process	14/7214	39/114793	7	3.09
GO:1901031	regulation of response to reactive oxygen species	6/7214	7/114793	6	3.08
GO:0006425	glutaminyl-tRNA aminoacylation	6/7214	7/114793	11	3.08
GO:0051103	DNA ligation involved in DNA repair	6/7214	7/114793	8	3.08
GO:0032958	inositol phosphate biosynthetic process	13/7214	18/114793	7	3.05
GO:0001522	pseudouridine synthesis	29/7214	138/114793	8	3.03
GO:0006606	protein import into nucleus	24/7214	81/114793	10	3.0
GO:1901617	organic hydroxy compound biosynthetic process	19/7214	81/114793	4	3.0
GO:0017186	peptidyl-pyroglutamic acid biosynthetic process, using glutaminyl-peptide cyclotransferase	11/7214	14/114793	9	2.98
GO:0043648	dicarboxylic acid metabolic process	27/7214	105/114793	6	2.98
GO:0045892	negative regulation of transcription, DNA-templated	16/7214	40/114793	10	2.94
GO:0006525	arginine metabolic process	15/7214	53/114793	9	2.93
GO:0051056	regulation of small GTPase mediated signal transduction	27/7214	82/114793	7	2.92
GO:0009082	branched-chain amino acid biosynthetic process	24/7214	60/114793	5	2.91
GO:0016573	histone acetylation	21/7214	83/114793	11	2.83
GO:0019751	polyol metabolic process	19/7214	83/114793	5	2.83
GO:0009084	glutamine family amino acid biosynthetic process	22/7214	83/114793	9	2.83
GO:0007154	cell communication	18/7214	54/114793	2	2.82
GO:0006537	glutamate biosynthetic process	11/7214	19/114793	10	2.81
GO:0006166	purine ribonucleoside salvage	13/7214	19/114793	10	2.81
GO:0034637	cellular carbohydrate biosynthetic process	25/7214	133/114793	5	2.8
GO:0042398	cellular modified amino acid biosynthetic process	23/7214	116/114793	5	2.79
GO:0006066	alcohol metabolic process	26/7214	116/114793	4	2.79
GO:0017038	protein import	28/7214	116/114793	9	2.79
GO:0009064	glutamine family amino acid metabolic process	40/7214	142/114793	8	2.79
GO:0033014	tetrapyrrole biosynthetic process	21/7214	76/114793	5	2.78
GO:0001510	RNA methylation	40/7214	188/114793	8	2.78
GO:0009966	regulation of signal transduction	43/7214	245/114793	5	2.77
GO:0046034	ATP metabolic process	32/7214	198/114793	2	2.75
GO:0006096	glycolytic process	26/7214	134/114793	12	2.74
GO:0000398	mRNA splicing, via spliceosome	50/7214	286/114793	11	2.73
GO:0051274	beta-glucan biosynthetic process	17/7214	62/114793	8	2.71
GO:0016052	carbohydrate catabolic process	37/7214	236/114793	4	2.69
GO:0072330	monocarboxylic acid biosynthetic process	31/7214	101/114793	7	2.68
GO:0031324	negative regulation of cellular metabolic process	18/7214	70/114793	5	2.65
GO:0000097	sulfur amino acid biosynthetic process	20/7214	70/114793	5	2.65
GO:0032012	regulation of ARF protein signal transduction	16/7214	56/114793	9	2.6
GO:0060271	cilium assembly	27/7214	119/114793	7	2.59
GO:0008380	RNA splicing	53/7214	321/114793	8	2.59
GO:1901615	organic hydroxy compound metabolic process	37/7214	192/114793	3	2.58
GO:0046394	carboxylic acid biosynthetic process	96/7214	495/114793	6	2.53
GO:0006480	N-terminal protein amino acid methylation	6/7214	8/114793	9	2.52
GO:0070925	organelle assembly	35/7214	215/114793	5	2.52
GO:0018205	peptidyl-lysine modification	29/7214	165/114793	8	2.52
GO:0008033	tRNA processing	92/7214	407/114793	9	2.51
GO:0046129	purine ribonucleoside biosynthetic process	14/7214	50/114793	9	2.51
GO:0008652	cellular amino acid biosynthetic process	65/7214	388/114793	7	2.5
GO:0018193	peptidyl-amino acid modification	65/7214	497/114793	7	2.5
GO:0032787	monocarboxylic acid metabolic process	64/7214	357/114793	6	2.47
GO:0006400	tRNA modification	52/7214	220/114793	10	2.46
GO:0010605	negative regulation of macromolecule metabolic process	42/7214	220/114793	5	2.46
GO:0032259	methylation	64/7214	377/114793	2	2.46
GO:0016053	organic acid biosynthetic process	126/7214	602/114793	4	2.45
GO:0016071	mRNA metabolic process	80/7214	646/114793	7	2.45
GO:1901607	alpha-amino acid biosynthetic process	48/7214	329/114793	8	2.42
GO:0048583	regulation of response to stimulus	49/7214	258/114793	3	2.42
GO:1901605	alpha-amino acid metabolic process	71/7214	546/114793	7	2.4
GO:0042364	water-soluble vitamin biosynthetic process	24/7214	131/114793	5	2.39
GO:0043414	macromolecule methylation	48/7214	304/114793	5	2.38
GO:0009451	RNA modification	107/7214	512/114793	7	2.35
GO:0006165	nucleoside diphosphate phosphorylation	27/7214	159/114793	8	2.35
GO:0006397	mRNA processing	67/7214	528/114793	8	2.35
GO:0006399	tRNA metabolic process	132/7214	725/114793	8	2.35
GO:0034660	ncRNA metabolic process	167/7214	1014/114793	7	2.33
GO:0043604	amide biosynthetic process	110/7214	802/114793	5	2.32
GO:0007017	microtubule-based process	116/7214	1016/114793	2	2.32
GO:0006518	peptide metabolic process	111/7214	990/114793	5	2.31
GO:0034470	ncRNA processing	127/7214	690/114793	8	2.28
GO:0006412	translation	91/7214	714/114793	7	2.28
GO:0044283	small molecule biosynthetic process	158/7214	943/114793	3	2.27
GO:0072594	establishment of protein localization to organelle	35/7214	227/114793	6	2.27
GO:0006355	regulation of transcription, DNA-templated	112/7214	1004/114793	9	2.24
GO:0031323	regulation of cellular metabolic process	130/7214	1282/114793	4	2.21
GO:0010468	regulation of gene expression	143/7214	1284/114793	5	2.2
GO:0007018	microtubule-based movement	93/7214	774/114793	3	2.2
GO:1901575	organic substance catabolic process	147/7214	1452/114793	3	2.19
GO:0043603	cellular amide metabolic process	135/7214	1100/114793	4	2.19
GO:0034645	cellular macromolecule biosynthetic process	188/7214	1454/114793	5	2.19
GO:0006520	cellular amino acid metabolic process	132/7214	1045/114793	6	2.18
GO:0044262	cellular carbohydrate metabolic process	37/7214	258/114793	4	2.17
GO:0051171	regulation of nitrogen compound metabolic process	134/7214	1260/114793	4	2.17
GO:0006082	organic acid metabolic process	258/7214	1638/114793	3	2.17
GO:2000112	regulation of cellular macromolecule biosynthetic process	118/7214	1107/114793	6	2.16
GO:0051252	regulation of RNA metabolic process	113/7214	1021/114793	6	2.16
GO:0006396	RNA processing	240/7214	1505/114793	7	2.13
GO:0006091	generation of precursor metabolites and energy	38/7214	272/114793	3	2.12
GO:0009056	catabolic process	154/7214	1580/114793	2	2.12

Table 11 Enriched terms relating to biological processes in the positively selected animal necrotrophs (continued).

GO number	Name	Ratio in study	Ratio in population	Depth	-log10 of p value
GO:0006629	lipid metabolic process	143/7214	1475/114793	3	2.12
GO:0060255	regulation of macromolecule metabolic process	160/7214	1442/114793	4	2.11
GO:0019752	carboxylic acid metabolic process	216/7214	1483/114793	5	2.09
GO:0034613	cellular protein localization	36/7214	248/114793	4	2.08
GO:1901566	organonitrogen compound biosynthetic process	277/7214	2065/114793	4	2.08
GO:0009059	macromolecule biosynthetic process	189/7214	1710/114793	4	2.07
GO:0006102	isocitrate metabolic process	7/7214	13/114793	7	2.05
GO:0050789	regulation of biological process	306/7214	3078/114793	2	2.03
GO:0016043	cellular component organization	184/7214	2008/114793	3	2.02
GO:0044281	small molecule metabolic process	317/7214	2782/114793	2	2.01
GO:0006401	RNA catabolic process	29/7214	185/114793	7	2.01
GO:0044271	cellular nitrogen compound biosynthetic process	215/7214	2275/114793	4	2.0
GO:0016070	RNA metabolic process	372/7214	2742/114793	6	2.0
GO:0050794	regulation of cellular process	270/7214	2861/114793	3	1.97
GO:0071840	cellular component organization or biogenesis	189/7214	2119/114793	2	1.96
GO:0090304	nucleic acid metabolic process	466/7214	3855/114793	5	1.96
GO:0044249	cellular biosynthetic process	430/7214	3741/114793	3	1.96
GO:1901576	organic substance biosynthetic process	464/7214	4002/114793	3	1.93
GO:0065007	biological regulation	330/7214	3366/114793	1	1.93
GO:0006139	nucleobase-containing compound metabolic process	548/7214	4921/114793	4	1.92
GO:0009058	biosynthetic process	506/7214	4268/114793	2	1.91
GO:0051273	beta-glucan metabolic process	22/7214	122/114793	7	1.88
GO:0006725	cellular aromatic compound metabolic process	604/7214	5329/114793	3	1.87
GO:1901360	organic cyclic compound metabolic process	632/7214	5489/114793	3	1.86
GO:0043412	macromolecule modification	580/7214	6499/114793	4	1.86
GO:0006555	methionine metabolic process	15/7214	64/114793	9	1.85
GO:0034641	cellular nitrogen compound metabolic process	717/7214	6287/114793	3	1.84
GO:0006464	cellular protein modification process	469/7214	5979/114793	6	1.83
GO:0046483	heterocycle metabolic process	600/7214	5319/114793	3	1.82
GO:0022607	cellular component assembly	90/7214	869/114793	4	1.81
GO:0044267	cellular protein metabolic process	566/7214	6741/114793	5	1.81
GO:0006793	phosphorus metabolic process	450/7214	5743/114793	3	1.78
GO:0019538	protein metabolic process	725/7214	9516/114793	4	1.77
GO:0044260	cellular macromolecule metabolic process	824/7214	9171/114793	4	1.74
GO:0006310	DNA recombination	31/7214	211/114793	7	1.7
GO:0044237	cellular metabolic process	1576/7214	15465/114793	2	1.67
GO:1901564	organonitrogen compound metabolic process	1025/7214	12079/114793	3	1.67
GO:0043170	macromolecule metabolic process	1243/7214	13653/114793	3	1.66
GO:0006541	glutamine metabolic process	10/7214	31/114793	9	1.65
GO:0006418	tRNA aminoacylation for protein translation	40/7214	302/114793	10	1.64
GO:0005975	carbohydrate metabolic process	121/7214	1279/114793	3	1.63
GO:0006796	phosphate-containing compound metabolic process	444/7214	5726/114793	4	1.63
GO:0006807	nitrogen compound metabolic process	1481/7214	15957/114793	2	1.63
GO:0044238	primary metabolic process	1583/7214	17428/114793	2	1.62
GO:0008152	metabolic process	1908/7214	19763/114793	1	1.61
GO:0071704	organic substance metabolic process	1793/7214	18791/114793	2	1.59
GO:0009987	cellular process	2251/7214	23958/114793	1	1.56
GO:0008150	biological_process	2760/7214	30756/114793	0	1.53
GO:0000096	sulfur amino acid metabolic process	22/7214	130/114793	4	1.43
GO:0002943	tRNA dihydrouridine synthesis	6/7214	11/114793	11	1.41
GO:0009072	aromatic amino acid family metabolic process	26/7214	169/114793	4	1.4

Table 12 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:0055085	transmembrane transport	16/62	557/11080	4	3.26	Achlya hypogyna
GO:0019637	organophosphate metabolic process	8/62	213/11080	4	1.36	Achlya hypogyna
GO:0043412	macromolecule modification	107/760	755/8600	4	2.26	Albugo candida
GO:0071704	organic substance metabolic process	271/760	2317/8600	2	2.12	Albugo candida
GO:0044237	cellular metabolic process	251/760	2142/8600	2	2.09	Albugo candida
GO:0044238	primary metabolic process	254/760	2154/8600	2	2.07	Albugo candida
GO:0006464	cellular protein modification process	95/760	671/8600	6	2.06	Albugo candida
GO:0008152	metabolic process	283/760	2451/8600	1	2.04	Albugo candida
GO:0044260	cellular macromolecule metabolic process	145/760	1177/8600	4	1.58	Albugo candida
GO:0043170	macromolecule metabolic process	196/760	1696/8600	3	1.45	Albugo candida
GO:0006807	nitrogen compound metabolic process	223/760	1970/8600	2	1.44	Albugo candida
GO:0006796	phosphate-containing compound metabolic process	90/760	660/8600	4	1.41	Albugo candida
GO:0044262	cellular carbohydrate metabolic process	12/504	37/8647	4	2.94	Albugo laibachii
GO:0034645	cellular macromolecule biosynthetic process	35/504	228/8647	5	2.47	Albugo laibachii
GO:0051560	mitochondrial calcium ion homeostasis	4/504	4/8647	10	1.7	Albugo laibachii
GO:0042592	homeostatic process	8/504	21/8647	3	1.64	Albugo laibachii
GO:0051274	beta-glucan biosynthetic process	6/504	11/8647	8	1.62	Albugo laibachii
GO:0098771	inorganic ion homeostasis	6/504	12/8647	6	1.34	Albugo laibachii
GO:0034637	cellular carbohydrate biosynthetic process	8/504	23/8647	5	1.31	Albugo laibachii
GO:0006821	chloride transport	6/477	22/17944	7	1.48	Aphanomyces astaci
GO:0045048	protein insertion into ER membrane	4/140	5/14252	8	4.08	Aphanomyces euteiches
GO:0046434	organophosphate catabolic process	4/140	16/14252	5	1.56	Aphanomyces euteiches
GO:0009084	glutamine family amino acid biosynthetic process	8/272	19/6501	9	3.14	Bremia lactucae
GO:0006561	proline biosynthetic process	6/272	10/6501	10	2.81	Bremia lactucae
GO:1901264	carbohydrate derivative transport	6/272	13/6501	7	1.94	Bremia lactucae
GO:0044271	cellular nitrogen compound biosynthetic process	0/272	274/6501	4	1.59	Bremia lactucae
GO:0006259	DNA metabolic process	18/843	803/12755	6	2.4	Globisporangium splendens
GO:0015074	DNA integration	2/843	666/12755	7	2.31	Globisporangium splendens
GO:0044260	cellular macromolecule metabolic process	81/843	1879/12755	4	1.68	Globisporangium splendens
GO:0034220	ion transmembrane transport	12/197	68/7213	5	3.39	Hyaloperonospora arabidopsidis
GO:0098656	anion transmembrane transport	8/197	33/7213	6	2.45	Hyaloperonospora arabidopsidis
GO:0055085	transmembrane transport	20/197	247/7213	4	1.69	Hyaloperonospora arabidopsidis
GO:0043933	protein-containing complex subunit organization	44/3329	134/20260	4	2.26	Nothophytophthora sp
GO:1901564	organonitrogen compound metabolic process	379/3329	2777/20260	3	1.4	Nothophytophthora sp
GO:0006665	sphingolipid metabolic process	4/123	20/13965	6	1.32	Phytophthora cinnamomi
GO:0007186	G protein-coupled receptor signaling pathway	4/201	12/19214	5	1.98	Phytophthora fragariae
GO:0098771	inorganic ion homeostasis	4/201	13/19214	6	1.83	Phytophthora fragariae
GO:0090304	nucleic acid metabolic process	2/201	1600/19214	5	1.81	Phytophthora fragariae
GO:1901360	organic cyclic compound metabolic process	4/201	1865/19214	3	1.4	Phytophthora fragariae
GO:0034219	carbohydrate transmembrane transport	2/41	2/8291	8	1.37	Phytophthora kernoviae
GO:0006643	membrane lipid metabolic process	6/103	37/18043	5	3.97	Phytophthora megakarya
GO:0009247	glycolipid biosynthetic process	4/103	18/18043	7	2.29	Phytophthora megakarya
GO:0006470	protein dephosphorylation	18/1465	55/12653	7	1.3	Phytophthora nicotianae
GO:0070536	protein K63-linked deubiquitination	4/748	5/17216	10	1.48	Phytophthora parasitica
GO:0006069	ethanol oxidation	2/43	2/9298	7	1.41	Pythium aphanidermatum

Figure 9 Differences in annotated cellular pathways from the Stramenopile dataset. Shown are pathways which have up to 36 repeated values per taxa. The clusters from Table 1 are encapsulated in a labeled square.

Figure 10 Differences in annotated cellular pathways for the members of the Saprolegniaceae family in the Stramenopile dataset. Shown are pathways which are different in at least one taxa and have at least one complete loss in any of the taxa.

Figure 11 Differences in annotated cellular pathways for the members of the Pythiaceae family and *Globisporangium* genus in the Stramenopile dataset. Shown are pathways which are different in at least one taxa and have at least one complete loss in any of the taxa.

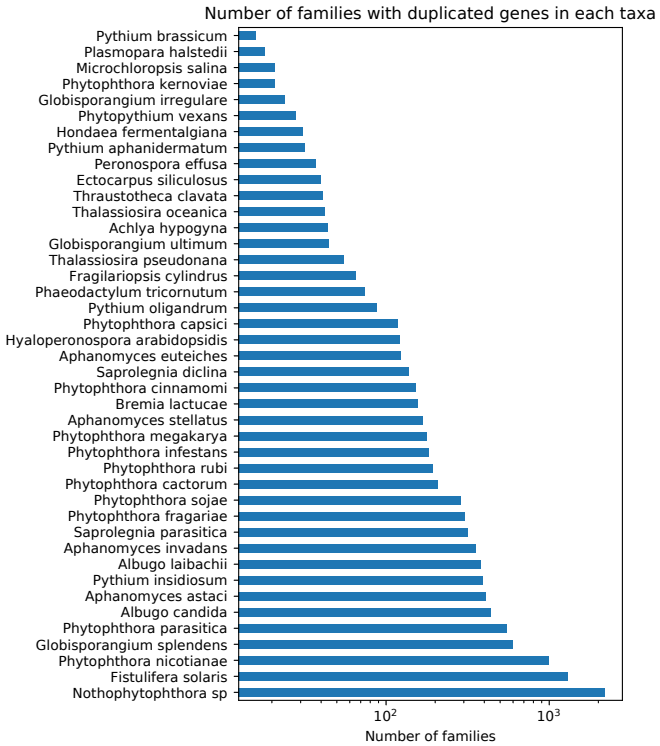

Figure 12 Duplicates in protein families in the Stramenopile dataset. Number of families with five or more members from different taxa that contain paralogs in the dataset.

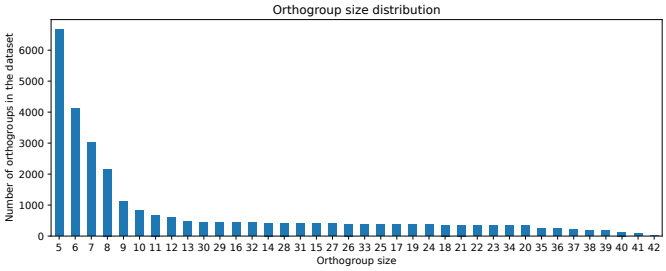

Figure 13 Distribution of protein family size in the dataset. Number of families with the same member size.

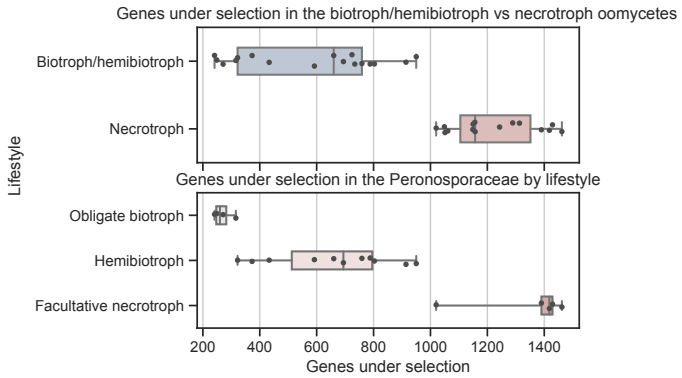

Figure 14 Comparison of the distribution of positive genes under selection for different lifestyles. Significance between the different categories is $p < 0.01$ in both the upper graph (Mann-Whitney test) and the lower graph (ANOVA one-tailed test).

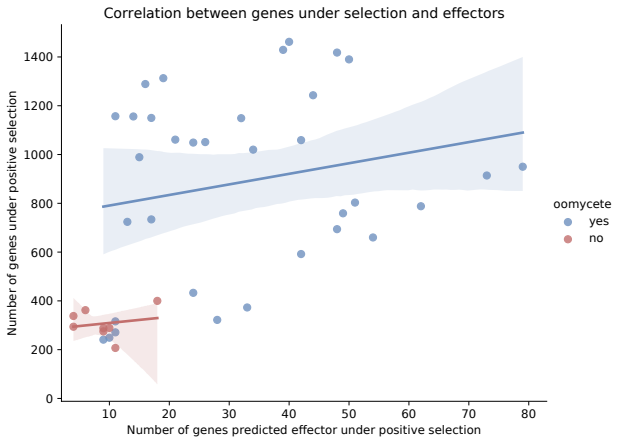

Figure 15 Correlation between genes under selection and effectors. Oomycetes represented in blue (Pearson's correlation, $r = 0.22$, $p = 0.22$) and non-oomycetes in red (Pearson's correlation, $r = 0.19$, $p = 0.65$). Pearson correlation represented as a straight line and the confidence interval represented as a lighter shade.

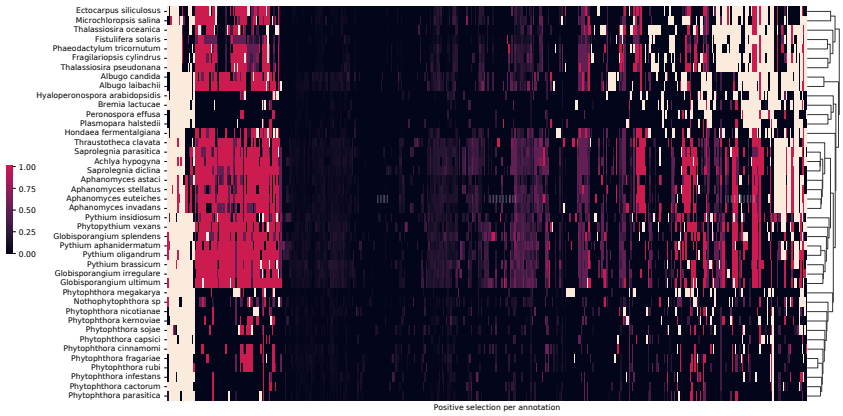

Figure 16 Heatmap of positive selection ratio of functional annotations in the Stramenopile dataset. The color gradient from black to red represents the ratio of genes with a particular functional annotation that are under selection. Uncolored cells represent the absence of the annotation in a species. Weighted-based clustering of the distance between the taxa is represented.

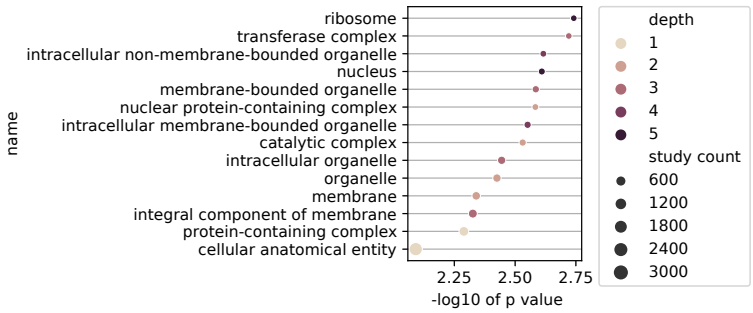

Figure 17 Cellular compartment enrichment under selection in all oomycetes in the Stramenopile dataset. The color represents the GO depth. GO depth is a measure of the number of parent nodes in the GO tree. The size of the dots corresponds to the total number of proteins under selection in the Stramenopile dataset that belong to said term.

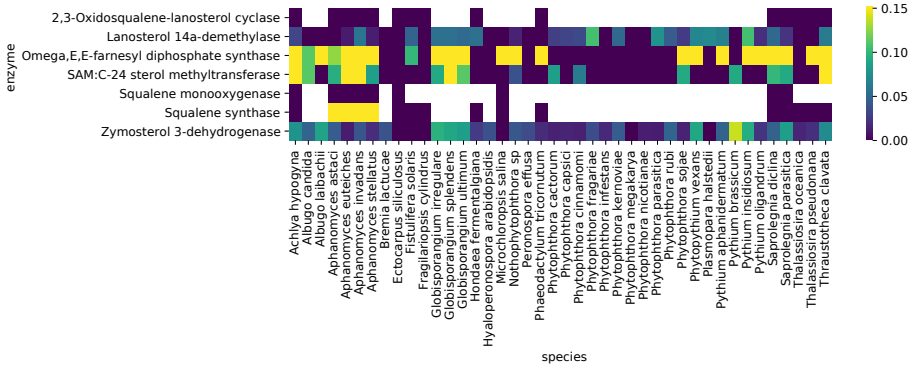

Figure 18 Sterol biosynthesis-related enzymes in Stramenopiles. Heatmap of the presence and absence of the enzymes relating to sterol biosynthesis pathway in the Stramenopiles. The yellow gradient represents the normalized ratio of predicted positive selection in genes with this annotation.

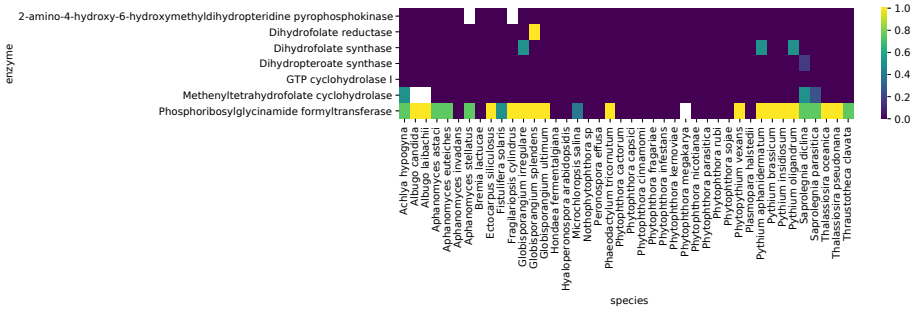

Figure 19 Tetrahydrofolate salvage and biosynthesis-related enzymes in Stramenopiles. Heatmap of the presence and absence of the enzymes relating to tetrahydropholate metabolism in the Stramenopiles. The yellow gradient represents the ratio of predicted positive selection in genes with this annotation.

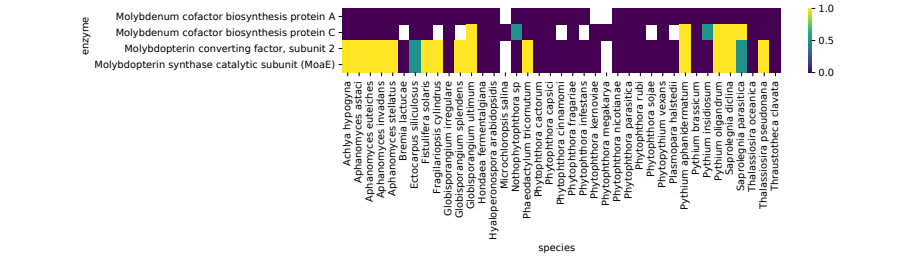

Figure 20 Molybdopterin biosynthesis-related enzymes in Stramenopiles. Heatmap of the presence and absence of the enzymes relating to molybdopterin biosynthesis in the Stramenopiles. The yellow gradient represents the ratio of predicted positive selection in genes with this annotation.

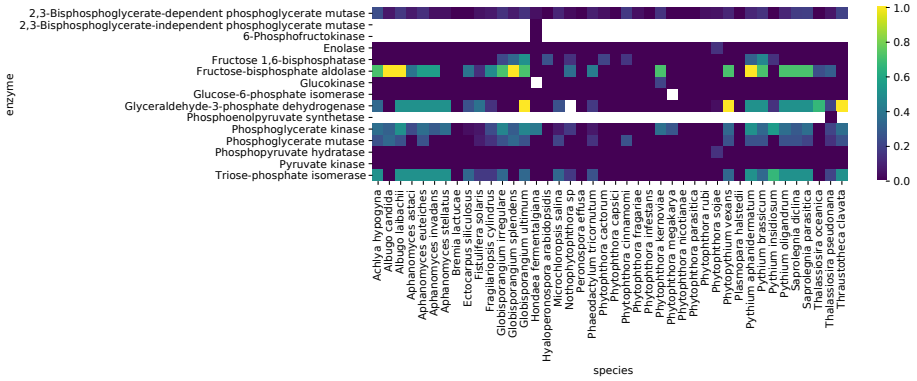

Figure 21 Glycolysis I, II and III-related enzymes in Stramenopiles. Heatmap of the presence and absence of the enzymes relating to glycolysis pathway in the Stramenopiles. The yellow gradient represents the normalized ratio of predicted positive selection in genes with this annotation.