Comparative Analysis of Oomycete Genome Evolution Using the Oomycete Gene Order Browser (OGOB)

Jamie McGowan[1,2], Kevin P. Byrne[3], and David A. Fitzpatrick[1,2,*]

[1]Genome Evolution Laboratory, Department of Biology, Maynooth University, Co. Kildare, Ireland

[2]Human Health Research Institute, Maynooth University, Co. Kildare, Ireland

[3]School of Medicine, UCD Conway Institute, University College Dublin, Ireland

*Corresponding author: E-mail: david.fitzpatrick@mu.ie.

Accepted: December 10, 2018

Abstract

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

Key words: oomycetes, synteny, gene order, effectors, comparative genomics, database, *Phytophthora*, phylostratigraphy.

Introduction

The oomycetes are a class of filamentous, eukaryotic microorganisms that include some of the most devastating plant and animal pathogens (Beakes et al. 2012). They represent one of the biggest threats to worldwide food security and natural ecosystems (Fisher et al. 2012). Oomycetes resemble fungi in terms of their morphology, filamentous growth, ecological niches, and modes of nutrition (Richards et al. 2006). Despite their extensive similarities, the evolutionary relationship between oomycetes and fungi represent one of the most distantly related evolutionary groupings within the eukaryotes (Burki 2014). Oomycetes are members of the stramenopiles lineage of the stramenopiles–alveolata–rhizaria eukaryotic supergroup, with close relationships to the diatoms and brown algae (Burki 2014). Within the oomycete class, there are a number of highly diverse orders, including the Peronosporales, Pythiales, Albuginales, and Saprolegniales orders (fig. 1). There is significant diversity both between and within these orders in terms of lifestyle, pathogenicity, and host range. The Peronosporales order is the most extensively studied order, consisting largely of phytopathogens, including the hemibiotrophic genus *Phytophthora* (the "plant destroyers") (fig. 1). The most notorious of which is *Phytophthora infestans*, the causative agent of late potato blight and causative agent of the Irish potato famine which resulted in the death of 1 million people in Ireland and the emigration of another million (Haas et al. 2009). *Phytophthora infestans* is reported to cause billions of euros' worth of worldwide potato crop loss annually (Haverkort et al. 2008). Other highly destructive *Phytophthora* species include *Ph. sojae* and *Ph. ramorum*. *Phytophthora sojae* has a narrow host range, infecting soybean, and costs between 1 and 2 billion dollars in crop loss per year (Tyler et al. 2006; Tyler 2007). *Phytophthora*

© The Author(s) 2018. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

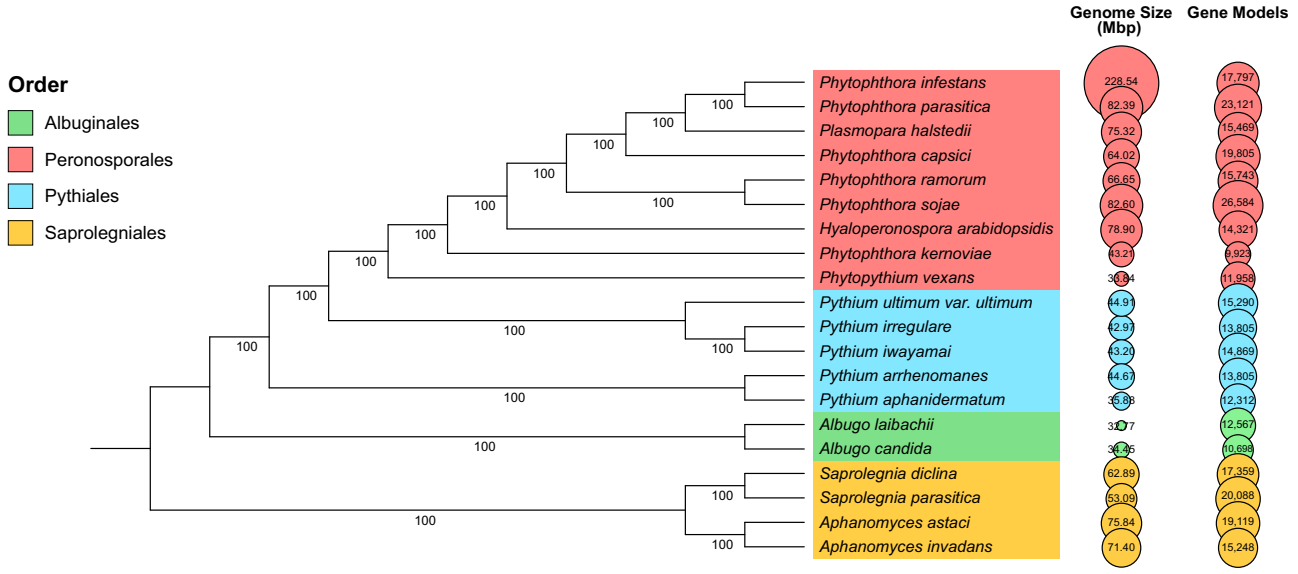

Fig. 1.—Matrix Representation with Parsimony (MRP) supertree of the 17,738 OGOB pillars that contain at least four genes. The supertree was generated in CLANN and is rooted at the Saprolegniales order. All nodes have 100% bootstrap support. Species are colored by order as follows: green, Albuginales;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

ramorum, in contrast, has a very wide host range with more than 100 host species reported and is destroying forest ecosystems worldwide (Rizzo et al. 2005; Tyler et al. 2006). Other members of the Peronosporales order include the obligate biotrophic genera *Hyaloperonospora* and *Plasmopara* (fig. 1), which cause downy mildew in a number of economically important crops (Coates and Beynon 2010; Gascuel et al. 2015; Sharma et al. 2015). The Pythiales order includes the cosmopolitan genus *Pythium* (fig. 1), which are generalist necrotrophs with broad host ranges that cause root rot in many important crops and ornamental plants (Lévesque et al. 2010; Adhikari et al. 2013). *Pythium ultimum* var. *ultimum* (hereafter referred to as *Py. ultimum*) is one of the most pathogenic *Pythium* species, with a broad host range including corn, soybean, wheat, and ornamental plants (Cheung et al. 2008). The Albuginales order is a more basal order (fig. 1) that includes the obligate biotrophic genus *Albugo* which causes "white blister rust" disease in various Brassicaceae species, including mustard and cabbage family plants (Kemen et al. 2011; Links et al. 2011). The Saprolegniales order (fig. 1) include animal and plant pathogens from the *Aphanomyces* genus (Diéguez-Uribeondo et al. 2009; Makkonen et al. 2016), and the *Saprolegnia* genus which causes severe infection of animals, in particular they cause "cotton mould" disease in many fish that are important in the global aquaculture industry (Jiang et al. 2013; van den Berg et al. 2013).

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

Despite their economic impact and the threat that they pose to worldwide food security, there is a lack of tools and databases available to study oomycete genes and genomes. This is particularly striking when compared with other taxonomic groups such as fungi. Databases, such as the *Pythium* Genome Database and the Comprehensive Phytopathogen

Table 1.
Genome Statistics and References for Oomycete Species Hosted on OGOB

Species	Assembly Size (Bp)	Scaffolds	Scaffold N50	Scaffold L50	GC%	% Gaps	Genes	Orphans[a]	G50	G90	BUSCO[b]	Source	Reference
Ph. infestans	228,543,505	4,921	1,588,622	38	50.9	16.81	17,797	793	18	103	95.80%	NCBI	Haas et al. (2009)
Ph. parasitica	82,389,172	708	888,348	25	49.5	34.61	23,121	3,130	20	87	94.10%	NCBI	Broad Institute (INRA-310)
Pl. halstedii	75,321,785	3,162	1,546,205	16	45.3	11.32	15,469	4,830	16	52	93.20%	ENSEMBL	Sharma et al. (2015)
Ph. capsici	64,023,748	917	705,730	29	50.4	12.47	19,805	390	27	89	91.00%	JGI	Lamour et al. (2012)
Hy. arabidopsidis	78,897,814	3,044	332,402	70	47.2	10.22	14,321	4,132	59	242	91.80%	ENSEMBL	Baxter et al. (2010)
Ph. sojae	82,601,618	83	7,609,242	4	54.6	3.96	26,584	1,725	4	13	97.90%	JGI	Tyler et al. (2006)
Ph. ramorum	66,652,401	2,576	308,042	63	53.8	18.35	15,743	150	47	548	97.40%	JGI	Tyler et al. (2006)
Ph. kernoviae	43,208,681	1,805	72,999	162	49.6	0.39	9,923	492	163	603	96.20%	ENSEMBL	PhyKer238/432
Pp. vexans	33,844,883	3,685	29,235	340	58.9	0.77	11,958	251	338	1,365	93.20%	ENSEMBL	Adhikari et al. (2013)
Py. iwayamai	43,199,174	11,541	11,008	1107	55.0	3.55	14,869	336	1,086	4,351	80.30%	ENSEMBL	Adhikari et al. (2013)
Py. irregulare	42,968,084	5,887	23,217	494	53.7	0.68	13,805	167	473	2,343	94.00%	ENSEMBL	Adhikari et al. (2013)
Py. ultimum	44,913,582	975	837,833	19	52.3	4.72	15,290	557	19	54	97.00%	ENSEMBL	Lévesque et al. (2010)
Py. arrhenomanes	44,672,625	10,972	9,784	1195	56.9	4.18	13,805	299	1,010	4,252	81.20%	ENSEMBL	Adhikari et al. (2013)
Py. aphanidermatum	35,876,849	1,774	37,384	270	53.8	4.49	12,312	189	244	873	90.60%	ENSEMBL	Adhikari et al. (2013)
Al. laibachii	32,766,811	3,827	69,384	130	44.3	0.00	12,567	2,238	106	391	80.30%	ENSEMBL	Kemen et al. (2011)
Al. candida	34,452,595	5,216	51,405	171	43.1	0.00	10,698	2,334	108	443	79.90%	ENSEMBL	Links et al. (2011)
Sa. diclina	62,885,792	390	602,571	34	58.6	35.66	17,359	361	26	95	93.20%	ENSEMBL	Broad Institute (VS20)
Sa. parasitica	53,093,248	1,442	280,942	46	58.4	9.33	20,088	603	41	371	87.20%	ENSEMBL	Jiang et al. (2013)
Ap. invadans	71,402,472	481	1,130,244	19	54.1	41.95	15,248	986	14	56	90.60%	ENSEMBL	Broad Institute (9901)
Ap. astaci	75,844,385	835	657,536	31	49.7	22.77	19,119	2,127	24	109	87.20%	ENSEMBL	Broad Institute (APO3)

NOTE.—Orphans were identified as species-specific in our phylostratigraphy analysis. G50 and G90 value correspond to the minimum number of scaffolds that contain at least 50% and 90% of total genes, respectively.
[a]Identified as species specific in our phylostratigraphy analysis.
[b]Percentage completeness as determined by BUSCO v3 against the Alveolata/Stramenopile BUSCO data set.

Genomics Resource (Hamilton et al. 2011), have been retired. FungiDB contains genome data for 16 oomycete species including information pertaining to orthology and synteny (Basenko et al. 2018), however the genome browser in FungiDB displays a to-scale representation of chromosomal regions making it unsuitable for the analysis of gene order and evolution. EumicrobeDB (Panda et al. 2018) was recently published and contains the genomes of several oomycete species and a large number of bioinformatics tools. EumicrobeDB has a tool for comparing syntenic regions between species, however, it is limited to comparing two species and displays a to-scale representation of chromosomal regions. Furthermore, it is not immediately obvious if an ortholog is absent in a species or if it is present in another area of the genome. These issues make EumicrobeDB unsuitable for detailed analysis of gene order and evolution across multiple species. To overcome this, we have developed the Oomycete Gene Order Browser (OGOB).

OGOB is a curated database that currently hosts genomic data for 20 oomycete species. Species included in OGOB were selected to include a broad range of representatives from the oomycete class and also based on the availability of gene sets. A recent review carried out a survey to rank the "top 10" oomycetes in terms of their economic and scientific importance (Kamoun et al. 2015). OGOB hosts eight of these species. OGOB also hosts a number of useful bioinformatics tools that allow users to carry out bioinformatic analyses in the web browser without installing local command line tools. This makes OGOB useful for comparative genomic, syntenic, and evolutionary analyses of oomycete genomes as well as for the analysis of individual genes and gene families. OGOB is based on the original synteny engine developed for the Yeast Gene Order Browser (YGOB) (Byrne and Wolfe 2005, 2006), to which we have made a number of functional and visual upgrades.

Here we describe the structure and functionality of OGOB. We have also undertaken a number of comparative genomic analyses using the genome data housed in database. These analyses yield insights into the evolution of oomycete genomes and the effect that gene duplication has had in shaping the gene repertoire of individual species. Using OGOB, we have investigated the overall conserved core of Oomycete genes as well as individual orders. Furthermore, we have also completed a comprehensive analysis of the 190 possible pairwise synteny comparisons between the 20 species hosted in OGOB. OGOB is available at https://ogob.ie, last accessed December 31, 2018.

Materials and Methods

OGOB Database Construction

Genomic data for the 20 oomycete species were retrieved from the sources listed in table 1, including genome assemblies and gene sets. Gene sets were manually inspected and dubious gene calls were removed. For genes with alternative transcripts, the longest transcript was retained. The final data set contains 319,881 protein coding genes. BUSCO v3 (Waterhouse et al. 2018) was used to assess the gene space completeness of each assembly with the alveolata/stramenopiles data set of BUSCOs. InterProScan 5 (Jones et al. 2014) was run on all 319,881 oomycete proteins in the OGOB database. Proteins were annotated for functional domains using the InterPro (Finn et al. 2017), Pfam (Finn et al. 2016), and PANTHER (Mi et al. 2017) databases, as well as for Gene Ontology terms (Ashburner et al. 2000). Signal peptides were predicted using SignalP (Bendtsen et al. 2004) and transmembrane domains were predicted with TMHMM (Krogh et al. 2001). Functional annotations are displayed on OGOB gene information pages and link back to the original annotation databases. Metabolic pathways were also annotated using the KEGG (Ogata et al. 1999), MetaCyc (Caspi et al. 2018), and Reactome (Fabregat et al. 2018) databases. All annotations can be downloaded from the OGOB data page (https://ogob.ie/gob/data.html, last accessed December 31, 2018).

Phylogenetic Analysis

A maximum-parsimony supertree approach was carried out to generate the oomycete species phylogeny (fig. 1). All pillars containing at least four genes (17,738 pillars) were retrieved and individually aligned using MUSCLE (Edgar 2004). Individual phylogenies for each of the 17,738 pillars were generated using FastTree v2.1.9 (Price et al. 2010). A supertree was constructed using the Matrix Representation with Parsimony (MRP) method implemented in Clann (Creevey et al. 2004; Creevey and McInerney 2005) with 100 bootstrap replicates. The phylogeny was visualized and annotated using the Interactive Tree of Life (Letunic and Bork 2007).

Phylostratigraphy Analysis

Individual phylostratigraphic maps for each of the 20 oomycetes were constructed following previously published methods (Quint et al. 2012; Drost et al. 2015). The data set used by Drost *et al* (2015) was retrieved. This data set contains amino acid sequences for 4,557 species including 1,787 eukaryotes (883 animals, 364 plants, 344 fungi, and 196 other eukaryotes) and 2,770 prokaryotes (2,511 bacteria and 259 archaea). We added all sequences hosted by OGOB to this database, resulting in a final database of 17,826,795 amino acid sequences. Each oomycete protein was searched against this database using BlastP (Altschul et al. 1997). Each protein is assigned to the oldest phylostrata that contains at least one BLAST hit with an *E* value cut-off of $1e^{-5}$. A gene is

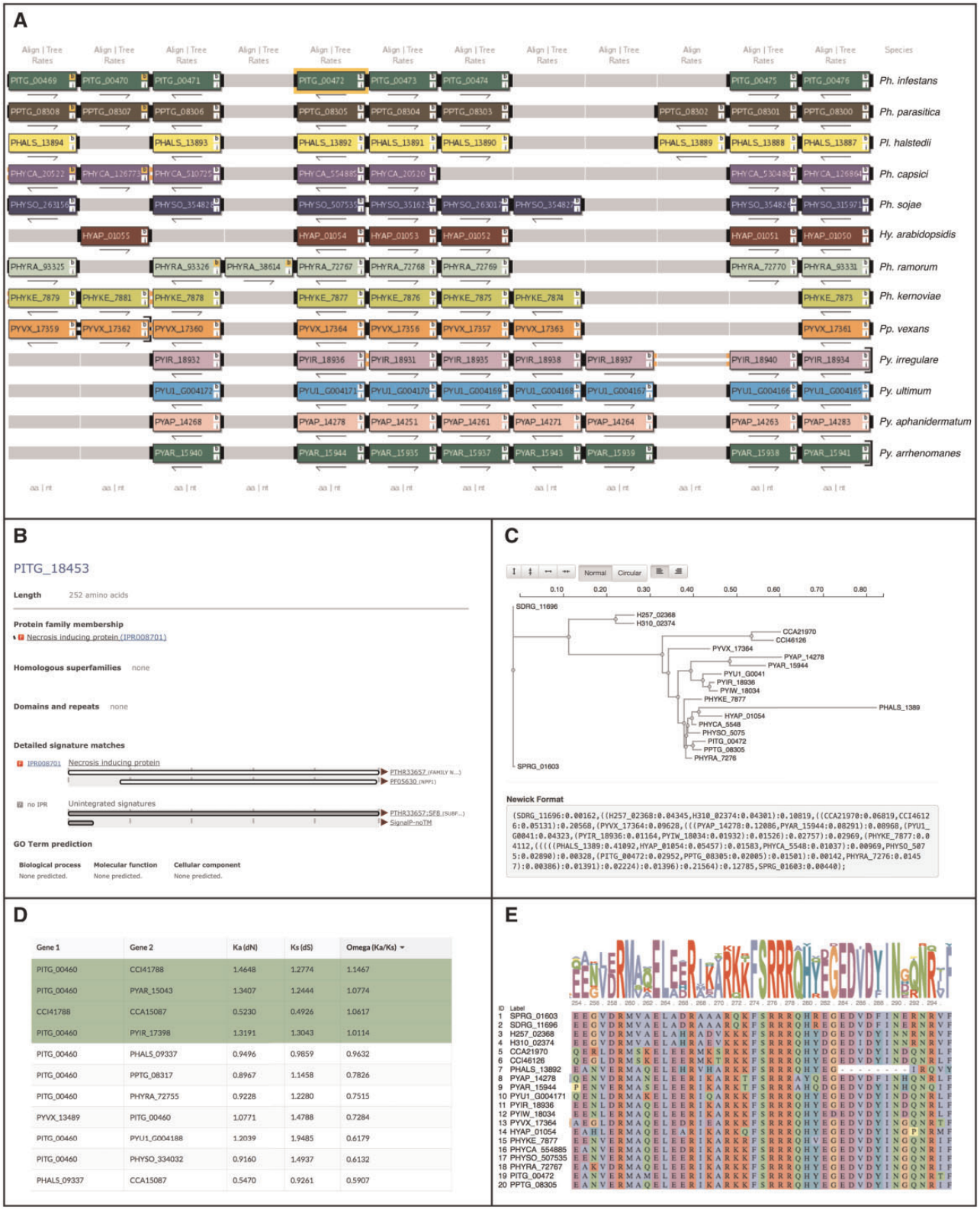

Fig. 2.—The Oomycete Gene Order Browser. (*A*) OGOB Screenshot. Each horizontal track represents a chromosomal segment from one species, with species labels on the right. Each box represents a protein coding gene, with the gene ID shown. Genes that are in the same pillar are orthologous. Each color represents a chromosome/scaffold. A change in color represents a breakdown in synteny. Genes colored gray indicate a nonsyntenic ortholog. Arrows under gene boxes represent the relative transcriptional orientation. Adjacent genes are connected by a solid black connector. Connector are colored gray if there is

assigned to the youngest phylostratum (i.e., species-specific orphan) if it does not have any such BLAST hit.

Gene Duplications

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

Multigene families were identified for each species by performing all-versus-all BlastP searches (Altschul et al. 1997) of each gene against every other gene in its genome with an *E* value cut-off of $1e^{-30}$, followed by Markov clustering using MCL (Enright et al. 2002) with an inflation value of 1.5.

Gene Enrichment Analyses

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

Results and Discussion

OGOB Structure and Functions

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

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

Links below each pillar (fig. 2*A*) allow users to retrieve the corresponding amino acid and nucleotide sequences for that pillar. Clicking the "i" button on any gene launches an information page for that gene showing any functional annotations (fig. 2*B*). Clicking on any annotation will link the user to the relevant annotation database (e.g., Pfam and InterPro). The information page for each gene has also a BLAST facility that permits users to search the corresponding nucleotide/protein sequence against OGOB's gene/protein data sets. Users can also search against the full genomes sequences or intergenic regions only. This facility allows users to confirm that a gene is missing from an assembly for example and in part helps overcome any of the shortfalls associated with genes that may have been missed during gene calling. Links at the top of each pillar (fig. 2*A*) allow users to construct phylogenetic trees ("Tree"), perform multiple sequence alignments ("Align"), and calculate evolutionary rates ("Rates"). Maximum likelihood phylogenies are generated using PhyML (Guindon et al. 2010) and displayed in an interactive interface that allows users to manipulate and root trees, implemented using phylotree.js (fig. 2*C*). Furthermore, users can also download phylogenies in Newick format for further processing. The rate of nonsynonymous (d*N*) and synonymous substitutions (d*S*) are calculated using yn00 (Yang and Nielsen 2000). A d*N*/d*S* ratio > 1 is indicative of positive selection and is highlighted in OGOB (fig. 2*D*). Multiple sequence alignments are performed using MUSCLE (Edgar 2004) and displayed in an

a gap in that genome. Double small bars connect nonadjacent genes that are <5 genes apart and single small bars connect genes that are <20 genes apart. Connectors are colored orange if there is an inversion. In this screenshot, the browser is focused on PITG_00472, an mRNA splicing factor SYF2 gene. This gene is in a genomic segment that is syntenically conserved in the Peronosporales and Pythiales, as shown by the colored blocks. (*B*) Gene info page showing functional annotations carried out by InterProScan, accessed by clicking the "i" button on gene boxes in OGOB. (*C*) Interactive maximum likelihood phylogeny of the genes in the same pillar as PITG_00472, accessed by clicking the "Tree" button at the top of pillars. (*D*) Pairwise yn00 evolutionary rates of genes in OGOB accessed by clicking the "Rates" button at the top of pillars. Genes highlighted green show evidence of positive selection, that is a d*N*/d*S* ratio $\geq$1. (*E*) MUSCLE multiple sequence alignment of all genes in the same pillar as PITG_00472, accessed by clicking the "Align" button at the top of pillars. The consensus sequence of the pillar is also shown.

interactive interface implemented using MSAViewer (Yachdav et al. 2016), the consensus sequence for the pillar is also shown (fig. 2*E*). Clicking the "b" button on any gene in OGOB launches a BLAST search of that gene against the entire OGOB database. BLAST results are colored to highlight orthologs, paralogs, tandem duplicates, singletons, and syntenically conserved hits. Users can select hits from the BLAST search and perform the above functions. This allows users to quickly analyze BLAST hits without having to manually obtain their sequences. We have also integrated a search interface into OGOB that makes it easier to study particular genes or gene families without the need to know individual gene identifiers. For example, users can search for genes that contain specific Gene Ontology terms or Pfam domains and easily compare the presence or absence of gene families across species and investigate their syntenic context. In addition, we have incorporated BLAST search support allowing users to search their own protein or nucleotide sequences against the OGOB database. The BLAST results page provides links to view hits in OGOB.

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

N50 scores are the most common score used to assess the quality of genome assemblies. However, it is well known that the metric suffers many problems. N50 scores can be artificially inflated in assemblies with large proportions of gaps or misassemblies. Furthermore, it can be difficult to determine how fragmented an assembly is based on N50 score alone without knowing the number of contigs/scaffolds or predicted genome size and it does not take gene models into account. Here we report an alternative metric which we call a "G50 score." An assembly's G50 score is the minimum number of scaffolds that contain at least 50% of genes. More generally, G*x* is the minimum number of contigs/scaffolds that contain at least *x* % of genes. G50 scores make it more immediately obvious how fragmented an assembly is and is better suited for synteny analyses than N50 scores. G50 scores are roughly proportional to L50 scores (table 1), except they do not take into account scaffolds that do not contain genes. For example, *Ph. sojae* has a G50 score of 4 (table 1) indicating a very high quality, contiguous assembly whereby four scaffolds contain at least 50% of the total genes. In comparison, *Py. iwayami* has a G50 score of 1,086 (table 1) indicating that this is a very fragmented assembly where 50% of the genes are distributed across 1,086 scaffolds. Without knowing any other metrics of the assembly we can tell that many scaffolds have only one gene therefore it is not possible to determine the syntenic context of these genes. Due to the contiguous nature, low N50 scores and poor G50 scores of some assemblies in OGOB, the levels of global synteny reported herein may be underestimates.

Orthology Curation and Syntenolog-Search

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

Genes were initially added to orthology pillars in OGOB using a reciprocal best BLAST hits strategy. Genes that are each other's best hits in a reciprocal BlastP search (*E* value cut-off $1e^{-10}$) are considered orthologs and added to the same pillar. This strategy initially placed the 319,881 oomycete genes into 146,768 pillars. A large number of pillars were singleton pillars (pillars with only one gene) that had significant BLAST hits to genes in other pillars but not reciprocal best hits. Using a similar approach to Synteno-BLAST which was used in CGOB (Maguire et al. 2013) and SearchDOGS (OhEigeartaigh et al. 2014), we have developed an automated strategy called "Syntenolog-Search" that combines results from BLAST searches with synteny information to identify microsyntenically conserved orthologs that cannot be identified using reciprocal best BLAST hits searches alone. We use the term "syntenolog" to describe syntenically

Table 2
Oomycete Tandem Duplication Analysis

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

conserved orthologs. Syntenolog-Search systematically examines all singleton pillars to check if they can be merged with another pillar based on homology and microsyntenic context. Compared with Synteno-BLAST, we use a stricter *E* value cut-off and a more permissive definition of microsynteny. Briefly, each singleton is searched against the OGOB database using BlastP (*E* value cut-off $1e^{-10}$). Hits are then examined for microsyntenic conservation to determine if there exists a pair of neighboring orthologs, within a distance of 20 genes that serve as anchor points, that is the query gene and hit gene are in a conserved genomic neighborhood. If such a hit exists, the two genes are considered syntenologs and the pillars are merged. For example, consider the two genes PITG_00248 and PPTG_10928 from *Ph. infestans* and *Ph. parasitica*, respectively. Both of these genes are annotated as Papain family cysteine proteases (PF00112). These genes do not have reciprocal best hits in a BlastP search, likely because they are both members of large paralogous families. However, they were identified as syntenologs by Syntenolog-Search and as a result both genes were moved to the same pillar (supplementary fig. S2*A*, Supplementary Material online). Upon manual inspection in OGOB, they are obvious orthologs. They share significant sequence similarity ($1e^{-125}$) (supplementary fig. S2*B* and *C*, Supplementary Material online) to each other and are syntenically conserved, co-occurring at the same loci (i.e., distance = 0) (supplementary fig. S2*A*, Supplementary Material online). This highlights the power of Syntenolog-Search in identifying reliable orthologous relationships that cannot be identified using BLAST alone. Syntenolog-Search inferred orthologous relationships for a further 22,708 oomycete genes, resulting in a final pillar count of 124,060. Thus, on average each pillar in OGOB has 2.58 genes.

Tandem Gene Duplications

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

In total 12,541 tandem clusters, corresponding to 29,717 genes, were identified across the 20 oomycete species (table 2, supplementary table S1, Supplementary Material online). The overall average number of genes per tandem cluster is 2.3 (table 2). *Phytophthora sojae* has the highest number of tandem clusters with 1,389 tandemly duplicated clusters, which corresponds to 3,411 genes or 12.83% of its total gene count (table 2). The obligate biotrophic species *Pl. halstedii*, *Al. laibachii*, and *Al. candida* have the smallest number of tandem clusters (149, 135 and 195 clusters respectively)

and have also the smallest proportions of their proteome belonging to tandem clusters (2.02%, 2.19%, and 3.91%, respectively) (table 2). Tandemly duplicated genes in *Sa. diclina* represent the highest proportion of the proteome (15.35% corresponding to 2,664 genes) (table 2) compared with the other species.

We set out to identify biological functions that are enriched or under-represented in tandemly duplicated clusters for each species. This was achieved by comparing the frequency of Gene Ontology (GO) Slim terms in tandem clusters relative to the nontandemly duplicated proportion of the proteome using the Fisher exact test, corrected for false discovery rate (FDR) using the Benjamini–Hochberg procedure. Here we report corrected P values < 0.05 as significant. Our results show enrichment in tandem clusters for a number of GO Slim terms in each species (supplementary table S2, Supplementary Material online), except for the two *Albugo* species, where no GO term was detected as being enriched or purified. We detected enrichment for terms related to transport, including establishment of localization (GO:0051234; 12 species), transmembrane transport (GO:0051234; 16 species), and transmembrane transporter activity (GO:0022857; 14 species) (supplementary table S2, Supplementary Material online). As with previous analyses (Martens and Van de Peer 2010), we also detected enrichment for terms that are potentially involved in pathogenicity such as extracellular region (GO:0005576; 15 species), hydrolase activity, acting on glycosyl bonds (GO:0016798; 13 species), carbohydrate metabolic process (GO:0005975; nine species), catalytic activity, acting on a protein (GO:0140096; eight species), hydrolase activity (GO:0016787; eight species), and peptidase activity (GO:0008233; eight species) (supplementary table S2, Supplementary Material online). Significantly, we also detected enrichment of proteins that contain signal peptides in tandem clusters for all species (supplementary table S2, Supplementary Material online), suggesting that tandem duplication events may be a major driving force for the evolution and expansion of secreted oomycete effectors.

Our analysis detected a number of terms related to housekeeping functions that are significantly under-represented in tandem clusters (supplementary table S2, Supplementary Material online), including intracellular part (GO:0044424; 17 species), nucleic acid metabolic process (GO:0090304; 16 species), intracellular organelle (GO:0043229; 15 species), nucleic acid binding (GO:0003676; 15 species), biosynthetic process (GO:0009058; 15 species), translation (GO:0006412; 14 species), ribosome (GO:0005840; 14 species), DNA metabolic process (GO:0006259; 13 species), RNA metabolic process (GO:0016070; 12 species), tRNA metabolic process (GO:0006399; 10 species), ncRNA metabolic process (GO:0034660; 10 species), and RNA binding (GO:0003723; 10 species). The majority of these terms describe cellular "housekeeping" genes that are usually members of large protein interaction networks. In yeast, these categories of genes have been shown to be recalcitrant to gene duplication as they interfere with highly constrained cellular systems and the dosage-balance hypothesis predicts that selection will remove these from populations (Papp et al. 2003; He and Zhang 2006; Li et al. 2006). Similarly in angiosperms single-copy genes are often involved in essential housekeeping functions that are highly conserved across all eukaryotes and are also resistant to duplication (De Smet et al. 2013).

Sequence similarity searches are not sufficient to identify highly divergent tandem duplicates. Using synteny information hosted by OGOB, it is possible to use slowly evolving tandem duplicates in one species to identify rapidly evolving tandems in other species that are so divergent that they cannot be identified by BLAST homology searches. For example, consider the tandem cluster of four *Ph. infestans* genes (PITG_01020, PITG_01022, PITG_01023, and PITG_01024) which have the elicitin Pfam domain (PF00964). This tandem cluster is conserved in *Ph. parasitica* and *Ph. sojae* but in *Ph. capsici* only two members were identified as tandemly duplicated (supplementary fig. S3*A*, Supplementary Material online). Upon manual inspection in OGOB we see orthologs in *Ph. capsici* to the two remaining tandem duplicates. These were not defined as tandemly duplicated as they did not meet our initial BLAST criteria, both genes are considerable longer than the orthologs in the other three species so violated the HSP cutoff. Furthermore, by comparing tandem duplicates between species, we can use OGOB to identify genes that arose via tandem duplication and later dispersed elsewhere in the genome. For example, *Ph. infestans* contains a cluster of five tandemly duplicated sugar efflux transporters (PITG_04998–PITG_05002). This cluster is syntenically conserved in *Ph. parasitica*, *Ph. capsici*, *Ph. ramorum*, and *Ph. sojae* (supplementary fig. S3*B*, Supplementary Material online). However, two members of the *Ph. sojae* tandem cluster have relocated to another loci on the same scaffold (supplementary fig. S3*B*, Supplementary Material online).

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

The Oomycete Paranome

We also identified the paranome for each oomycete species, that is the set of all paralogous multigene families. *Saprolegnia parasitica* has the highest number of multigene families (3,010), whereas *Ph. kernoviae* (757) has the lowest

Table 3
The Oomycete Paranome

Species	Genes	Unique Genes	% Unique Genes	Genes in Multigene Families	% Genes in Multigene Families	Multigene Families	Avg # Genes per Family	2 Members	3 Members	4 Members	≥5 Members
Ph. infestans	17,797	8,097	45.50%	9,700	54.50%	2,167	4.48	12.16%	6.39%	4.81%	31.15%
Ph. parasitica	23,121	11,922	51.56%	11,199	48.44%	2,162	5.18	9.42%	4.90%	2.99%	31.12%
Pl. halstedii	15,469	11,139	72.01%	4,330	27.99%	1,054	4.11	7.01%	4.09%	2.33%	14.56%
Ph. capsici	19,805	8,310	41.96%	11,495	58.04%	2,007	5.73	8.95%	5.12%	4.20%	39.77%
Hy. arabidopsidis	14,321	8,743	61.05%	5,578	38.95%	1,645	3.39	15.57%	5.36%	2.32%	15.70%
Ph. sojae	26,584	9,776	36.77%	16,808	63.23%	2,723	6.17	9.25%	4.83%	3.64%	45.50%
Ph. ramorum	15,743	5,827	37.01%	9,916	62.99%	1,755	5.65	11.13%	5.68%	3.96%	42.22%
Ph. kernoviae	9,923	6,730	67.82%	3,193	32.18%	757	4.22	8.00%	3.60%	2.82%	17.76%
Pp. vexans	11,958	7,025	58.75%	4,933	41.25%	1,163	4.24	9.82%	5.75%	3.38%	22.31%
Py. iwayamai	14,869	8,876	59.69%	5,993	40.31%	1,318	4.55	9.24%	4.50%	2.74%	23.82%
Py. irregulare	13,805	7,966	57.70%	5,839	42.30%	1,240	4.71	8.68%	4.89%	3.13%	25.60%
Py. ultimum	15,290	8,761	57.30%	6,529	42.70%	1,313	4.97	7.89%	4.53%	3.22%	27.06%
Py. arrhenomanes	13,805	8,231	59.62%	5,574	40.38%	1,260	4.42	9.03%	4.67%	3.53%	23.14%
Py. aphanidermatum	12,312	7,360	59.78%	4,952	40.22%	1,173	4.22	9.36%	5.48%	3.02%	22.36%
Al. laibachii	12,567	6,842	54.44%	5,725	45.56%	1,536	3.73	16.95%	4.92%	2.96%	20.73%
Al. candida	10,698	7,203	67.33%	3,495	32.67%	1,049	3.33	12.68%	4.43%	2.99%	12.57%
Sa. diclina	17,359	8,711	50.18%	8,648	49.82%	1,700	5.09	9.76%	4.96%	3.39%	31.71%
Sa. parasitica	20,088	7,966	39.66%	12,122	60.34%	3,010	4.03	18.66%	6.54%	3.94%	31.20%
Ap. invadans	15,248	8,955	58.73%	6,293	41.27%	1,389	4.53	9.80%	4.19%	3.46%	23.82%
Ap. astaci	19,119	9,875	51.65%	9,244	48.35%	1,700	5.44	8.65%	4.53%	3.54%	31.63%

NOTE—Multigene families are identified by a BlastP search with an E value cut-off of 1e-30 followed by MCL clustering with an inflation value of 1.5.

number (table 3). *Phytophthora ramorum* has the lowest number (5,827) of genes that do not belong to multigene families, whereas *Ph. parasitica* has the highest (11,922) (table 3). The proportion of genes that belong to multigene families varies greatly between oomycete species. In *Pl. halstedii*, only 28% of genes belong to a multigene family, whereas 63% of *Ph. sojae* genes belong to multigene families (table 3). On average, ~45.5% of oomycete genes housed in OGOB belong to a multigene family. The average number of genes in each family ranges from 3.3 to 6.2, with an overall average of 4.6 genes per family (table 3).

We also carried out a GO enrichment analysis to determine if any GO terms are over or under-represented in the paranome of each species. This identified enrichment of terms including ion binding (GO:0043167; 20 species), ATPase activity (GO:0016887; 20 species), cellular proteins modification process (GO:0006464; 20 species), and regulation of cellular process (GO:0050794; nine species) (supplementary table S2, Supplementary Material online). Similar to our tandem duplication analysis, we see enrichment of terms related to transport (GO:0006810; 17 species) and establishment of localization (GO:0051234; 17 species) (supplementary table S2, Supplementary Material online). We also see enrichment of terms potentially involved in pathogenicity including carbohydrate metabolic process (GO:0005975; 20 species) and extracellular region (GO:0005576; six species) (supplementary table S2, Supplementary Material online). In terms of under-represented GO terms in the oomycete paranome, our results largely match that of tandem clusters. We see GO terms related to housekeeping functions under-represented in the paranome of most species, including cytoplasmic part (GO:0044444; 20 species), RNA processing (GO:0006396; 20 species), cellular component organization (GO:0016043; 20 species), nucleus (GO:0016043; 20 species), RNA binding (GO:0003723; 20 species), translation (GO:0006412; 19 species), ribosome (GO:0005840; 19 species), ribosome biogenesis (GO:0042254; 18 species), nuclease activity (GO:0004518; 17 species), and cell cycle (GO:0007049; 14 species) (supplementary table S2, Supplementary Material online). Our results above are in line with previous analyses of *Phytophthora* and *Pythium* species that have shown that pathogenicity related genes are typically expanded relative to genes not directly linked to pathogenicity (Tyler et al. 2006; Haas et al. 2009; Lévesque et al. 2010).

Phylostratigraphy Analysis

To further elucidate oomycete genome evolution we carried out a phylostratigraphic analysis of each species housed in OGOB. Phylostratigraphy is a statistical approach for reconstructing macroevolutionary transitions by identifying the evolutionary emergence of founder genes across the tree of life (Domazet-Lošo et al. 2007; Tautz and Domazet-Lošo 2011; Sestak and Domazet-Loso 2015). We estimated the age and

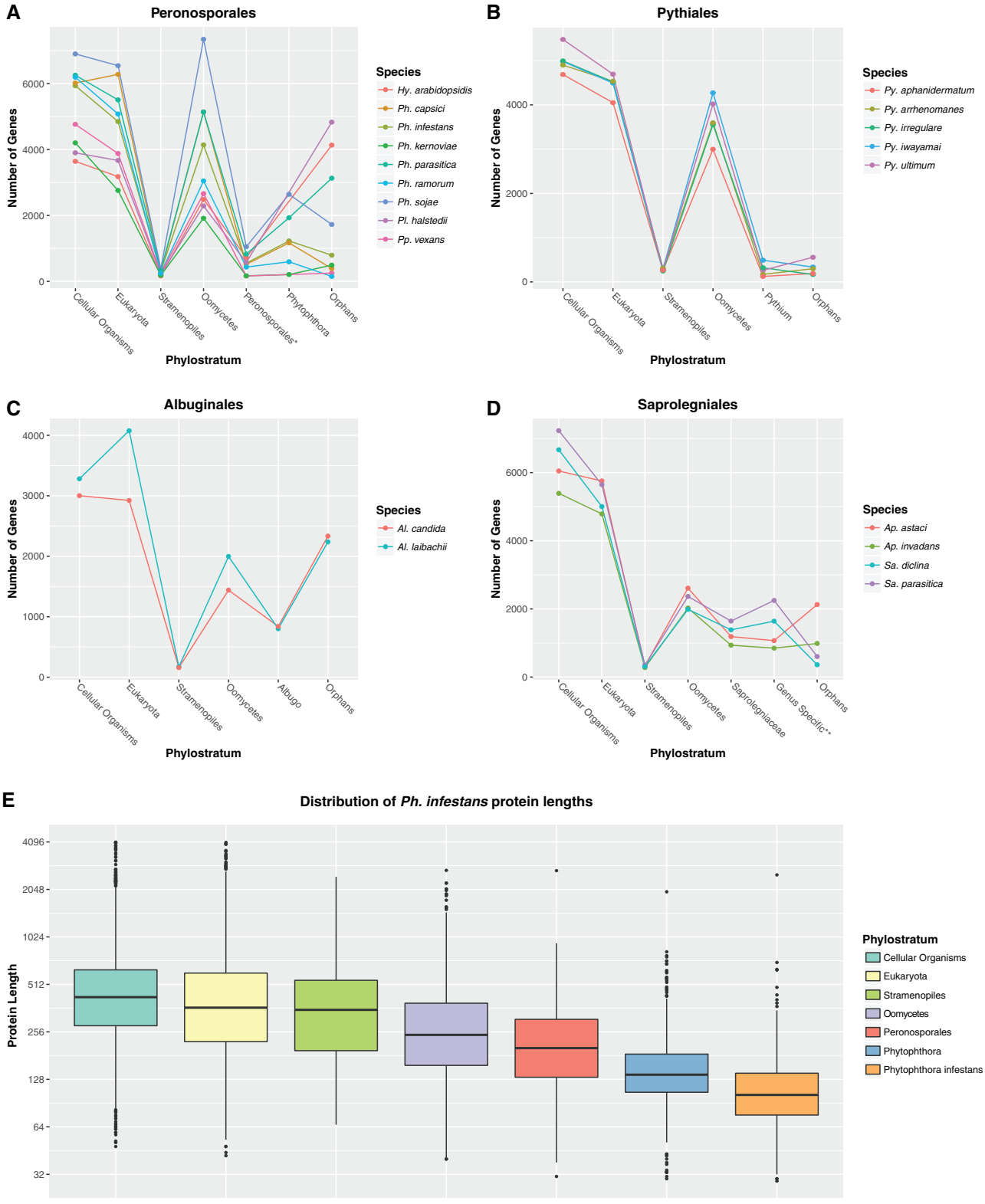

Fig. 3.—Phylostratigraphic analysis of 20 oomycete species to determine the number of founder genes that have arisen in each phylostratum in Peronosporales (*A*), Pythiales (*B*), Albuginales (*C*), and Saprolegniales (*D*). (*E*) Distribution of protein lengths in *Ph. infestans* across phylostrata shows a continuous increase in length with evolutionary age.

emergence of all 319,881 oomycete genes by identifying their founder genes in a database containing species across the tree of life. We generated our database by merging all sequences hosted on OGOB with those from a previous analysis with broad phyletic distribution (Drost et al. 2015), resulting in a final database of 17,826,795 amino acid sequences (see Materials and Methods section). Each oomycete gene was searched against the database using BlastP (*E* value cutoff $1e^{-5}$) and genes were assigned to a phylostratum based on their most ancient hit.

In total for the 20 species, 104,662 genes (32.7%) were placed at the origin of cellular organisms (i.e., homologs were identified in bacteria or archaea), 92,218 genes (28.8%) were eukaryotic in origin, 5,355 genes (1.7%) arose in the stramenopiles, and 65,015 genes (20.3%) arose in the oomycetes (fig. 3*A*–*D*). The remaining genes were assigned as unique to particular oomycete lineages, 26,090 (8.2%) of which were determined to be unique to individual species (orphan genes). By comparing phylostratigraphic maps for each species we can identify macroevolutionary trends. The overall trends identified for each oomycete order are largely similar. With few exceptions, genes of ancient Prokaryotic origin represent the largest proportion of each oomycete genome (fig. 3*A*–*D*). A slightly smaller proportion arose in the evolution of eukaryotes (fig. 3*A*–*D*). *Phytophthora capsici* and *Al. laibachii* are exceptions, whereby more genes were identified as eukaryotic in origin than Prokaryotic (fig. 3*A* and *C*). Very few genes (between 159 and 382 genes in each species) were detected to have arisen during the evolution of stramenopiles (fig. 3*A*–*D*). Our results suggest that either very few genes were gained during the evolution of stramenopiles or else a large number of genes of stramenopiles origin were later lost. Following this, in each species we see a large burst of oomycete genes being formed (fig. 3*A*–*D*). On average, genes of oomycete origin correspond to 21.34% of Peronosporales genomes (fig. 3*A*), 26.24% of *Pythium* genomes (fig. 3*B*), 14.68% of *Albugo* genomes (fig. 3*C*), and only 12.53% of Saprolegniales genomes (fig. 3*D*). *Phytophthora sojae* has more genes of Oomycete origin than any other species, making up the largest proportion of its total gene count (27.62%), suggesting large scale duplication of genes of oomycete origin (fig. 3*A*). For species in the Peronosporales order (fig. 3*A*) we see that genes of Peronosporales origin correspond to very few genes (between 163 genes in *Pp. vexans* to 1,052 genes in *Ph. sojae*). In addition, genes of *Phytophthora* origin represent only a small proportion of *Phytophthora* genomes, corresponding to an average of only 6.15% of genes (fig. 3*A*). We see bursts of emerging orphans in the downy-mildews (*Hy. arabidopsidis* and *Pl. halstedii*) and also in some *Phytophthora* species (*Ph. parasitica* and *Ph. sojae*) (fig. 3*A*). Seven hundred and ninety-three genes (4.46%) in *Ph. infestans* were identified as unique to *Ph. infestans*. The same is true for *Pythium* species as very few genes were assigned to the *Pythium* phylostratum (between 124 and 490 genes) (fig. 3*B*). Furthermore, we see very few orphan genes in each *Pythium* species (between 167 and 557 genes) (fig. 3*B*), suggesting there is not a great deal of gene content diversity. This is in contrast to the Albuginales and Saprolegniales species where larger numbers of genus specific and species specific genes were detected (fig. 3*C* and *D*).

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

Our phylostratigraphy approach allows us to account for differences in the number of genes each species has. For example, *Al. laibachii* has more genes than its closest relative *Al. candida* (12,567 vs 10,698 genes). *Albugo laibachii* has 4,077 genes that were assigned to the eukaryotic node, these genes are distributed across 2,549 gene families. Similarly, *Al. candida* has 2,923 eukaryotic genes located across 2,378 families. Furthermore, at the oomycete node, *Al. laibachii* has 2,000 genes that are grouped into 1,400 families while *Al. candida* has 1,440 grouped into 1,250 families. The number of orphan genes found in both species is very similar (2,334 vs 2,238). Therefore, *Al. laibachii* has more genes than *Al. candida* not because of the de novo formation of orphan genes but, rather, it has more copies of genes that can be mapped back to the eukaryotic and oomycete nodes (fig. 3*C*). These differences are due to retention and expansion of gene families from these nodes in *Al. laibachii*. A similar trend can be seen when we compare the gene content of *Ph. sojae* and *Ph. kernoviae* (26,584 vs 9,923 genes). *Phytophthora sojae* has 6,543 genes of eukaryotic origin distributed across 3,752 families, while *Ph. kernoviae* only has 2,757 genes of eukaryotic origin belong to 2,331 families. Furthermore, *Ph. sojae* has 7,342 genes of oomycete origin distributed across 3,416 families whereas *Ph. kernoviae* only has 1,915 genes of oomycete origin distributed across 1,664 families. It should be noted however that *Ph. sojae* has an additional 1,233 orphan genes relative to *Ph. kernoviae* (1,725 vs 492). Interestingly *Pl. halstedii*, *Hy. arabidopsidis*, and *Ph. parasitica* have experienced large bursts of orphan gene formation (4,830, 4,132 and 3,130 genes) that correspond to large proportions of their genomes (fig. 3*A*).

Comparisons of average protein length across phylostrata in *Ph. infestans* reveal that the length of proteins increases with evolutionary age (fig. 3*E*). Proteins found at the youngest phylostratum (i.e., orphan genes) are the shortest, and protein length increases across each older phylostrata, with the longest proteins being found at the oldest phylostratum (Cellular Organisms) (fig. 3*E*). A similar trend was identified in all species in our data set (supplementary fig. S4, Supplementary Material online). This has also been observed in other species such as yeast (Carvunis et al. 2012), *Arabidopsis thaliana* (Guo 2013) and metazoa (Neme and Tautz 2013), suggesting that similar evolutionary pressures are influencing genome and molecular evolution across distantly related eukaryotic species.

The Core Oomycete Ortholog Gene Set

We used OGOB's orthology pillars to identify core oomycete genes. We define core orthologs as the set of orthologs that are present in all species (i.e., pillars with 20 genes). We also define syntenologs as core orthologs that are microsyntenically conserved in all species (i.e., pillars with 20 core orthologs whereby each gene is microsyntenically conserved with every other gene). Overall, our analysis revealed 1,835 core oomycete orthology pillars. Thus, on average 12% of all oomycete genes have an ortholog in every other species (supplementary table S3, Supplementary Material online). Only 37 syntenolog pillars (2% of core pillars) were identified (supplementary table S3, Supplementary Material online). Oomycete syntenologs correspond to an average of only 0.25% of total genes in a genome. However, this is a very strict approach as each ortholog must be microsyntenically conserved with every other gene. Furthermore, the contiguous nature of some of the assemblies in OGOB may contribute to this low number. Therefore, we repeated this analysis to identify core and syntenolog orthologs individually in the Peronosporales order, the Saprolegniales order, the *Pythium* genus and the *Albugo* genus.

Overall we identified 4,063 core orthology pillars in the eight Peronosporales species, of which 2,279 (56%) belong to the syntenolog category, corresponding to between 8.57% and 22.97% of total genes in Peronosporales species (supplementary table S3, Supplementary Material online). Core pillars (6,483) were identified for the five species in the *Pythium* genus, of which 2,863 (44.16%) belong to the syntenolog category (supplementary table S3, Supplementary Material online). This corresponds to between 18.72% and 23.25% of total genes in *Pythium* species. Analysis of the four species in the Saprolegniales order revealed an even more extensive degree of syntenic conservation, 8,910 core pillars were identified, of which 7,718 (86.62%) belong to the syntenolog category (supplementary table S3, Supplementary Material online). This corresponds to between 38.42% and 50.62% of total genes in Saprolegniales genomes being both ubiquitous and microsyntenically conserved. The highest degree of synteny was detected in the *Albugo* genus where 6,719 core pillars and 6,313 syntenolog pillars were detected (supplementary table S3, Supplementary Material online). This means that 93.96% of orthologs within *Albugo* are microsyntenically conserved. This corresponds to between 50.23% and 59% of total *Albugo* genes. This result may be biased, however, as there are only two closely related *Albugo* genomes in OGOB.

For each group of species (Peronosporales, Saprolegniales, *Pythium*, and *Albugo*) we carried out an enrichment analysis of syntenically conserved core orthologs by comparing the frequency of GO terms associated with genes found in syntenolog pillars relative to nonsyntenolog pillars. As expected, our results show that syntenically conserved orthologs are enriched for housekeeping functions including GO terms such as ribosome (GO:0005840), translation (GO:0006412), cellular macromolecule biosynthetic process (GO:0034645), amide biosynthetic process (GO:0043604), RNA binding (GO:0003723), nucleus (GO:0005634), nucleolus (GO:0005730) (supplementary table S3, Supplementary Material online). These findings are consistent with the hypothesis that oomycetes contain "gated communities" where conserved and housekeeping genes reside (Bhowmick and Tripathy 2014). Terms under-represented in syntenic orthologs include establishment of localization (GO:0051234), carbohydrate metabolic process (GO:0005975), transmembrane transport (GO:0055085), extracellular region (GO:0005576), and ATPase activity (GO:0016887) (supplementary table S3, Supplementary Material online).

To fine tune our analysis even further, we also investigated the degree of microsynteny in each possible pair of the 20 oomycete species. For each pair of species, we identify orthologs and quantify the proportion that are microsyntenically (microsyntenologs) conserved. We consider a pair of orthologs to be microsyntenically conserved if there exists another pair of orthologs within a distance of 20 genes. Unsurprisingly, our results reveal very high levels of microsynteny between closely related species within oomycete orders, and a breakdown in synteny between more distantly related species across orders (supplementary table S4, Supplementary Material online). We use the proportion of microsyntenically conserved genes to generate a distance matrix and use this to cluster species based on microsyntenic conservation (fig. 4). As expected, more closely related organisms share a higher degree of microsynteny and are clustered together into their orders and genera (fig. 4). When comparing any two oomycete species the proportion of orthologs that are microsyntenically conserved is between 27.57% and 96.39% (supplementary table S4, Supplementary Material online). *Saprolegnia diclina* and *Sa. parasitica* share the highest degree of microsynteny (82.41% of genes or 96.35% of orthologs), followed by *Ap. astaci* and *Ap. invadans* (75.55% of genes or

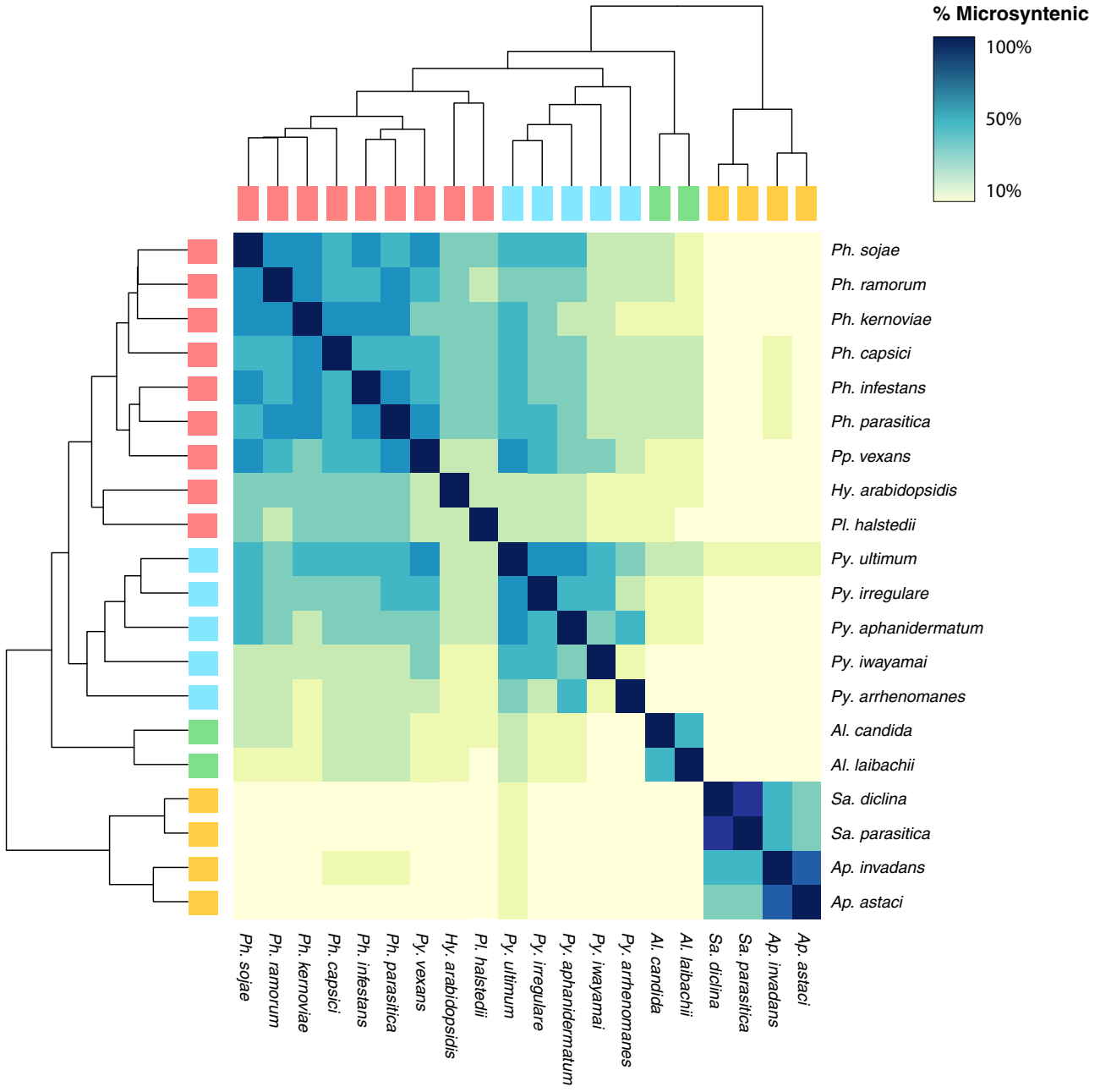

Fig. 4.—Pairwise microsyntenic analysis of oomycete species. Heatmap values represent the total proportion of genes in the smallest genome that were identified as syntenologs. These values are used to cluster the species based on microsynteny.

96.39% of orthologs) (fig. 4 and supplementary table S4, Supplementary Material online). The two species showing the lowest level of microsynteny are *Ap. astaci* and *Py. arrhenomanes* (10.34% of genes or 28.01% of orthologs) (fig. 4 and supplementary table S4, Supplementary Material online). On average 33.90% of total genes or 64.91% of orthologs are microsyntenic when comparing any two oomycete species in OGOB (supplementary table S4, Supplementary Material online). It is worth noting that we also performed the above analysis with more restrictive window sizes (i.e., a window size of 5 instead of 20) and results were largely congruent (not shown).

Our results are largely in agreement with previous analyses. For example, a previous study determined that over 75% of exons in *Ph. ramorum* and *Ph. sojae* aligned in a whole genome alignment (Tyler et al. 2006). Here, we find that 11,070 orthologs are shared between *Ph. ramorum* and *Ph. sojae*, of which 10,158 (91.76%) were detected to be microsyntenic. This corresponds to approximately 65% of genes in *Ph. ramorum*. The authors of the *Ph. infestans* genome reported that

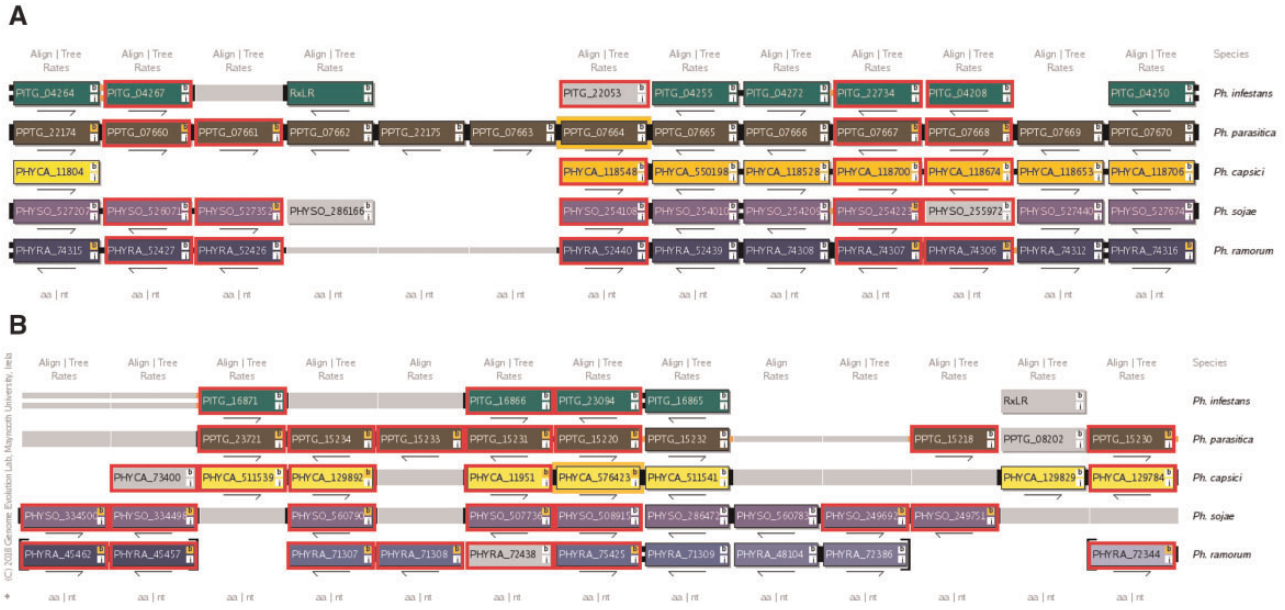

Fig. 5.—Two loci with multiple NLPs. Orthologs that contain a PFAM NLP domain (PF05630) are indicated with a red box. For display purposes only *Ph. infestans*, *Ph. parasitica*, *Ph. capsici*, *Ph. sojae*, and *Ph. ramorum* are shown. *Albugo*, *Aphanomyces*, and *Saprolegnia* species lack proteins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

90% of orthologs shared by *Ph. infestans*, *Ph. ramorum*, and *Ph. sojae* are found in blocks of conserved gene order (Haas et al. 2009). Our results are in agreement with this. We identified 9,219 orthologs that are present in all three species, of which 8,322 are microsyntenically conserved (90%). Another analysis reported extensive synteny between *Py. ultimum* and *Phytophthora* species (Lévesque et al. 2010). We detect that up to 94% of orthologs between *Py. ultimum* and any *Phytophthora* species are syntenologs (supplementary table S4, Supplementary Material online). Surprisingly, our results suggest that there is a greater degree of microsyntenic conservation between *Phytophthora* species and *Py. ultimum* than between *Phytophthora* species and *Hy. arabidopsidis* or *Pl. halstedii* (fig. 4). This may be due to gene loss events or extensive genome rearrangements in the evolution of obligate biotrophy in these downy mildew species (Baxter et al. 2010).

When comparing microsynteny between orders, the Saprolegniales species are most divergent from other species. On average only 40.70% of orthologs or 16.05% of total genes are microsyntenically conserved when compared with species outside of the order (fig. 4, supplementary table S4, Supplementary Material online). This is not surprising as the Saprolegniales are thought to have diverged from other oomycetes approximately 200 Ma (Matari and Blair 2014). However, species within the Saprolegniales order have the highest degree of microsyntenic conservation, on average 91.68% of their orthologs are microsyntenically conserved or 61% of total genes. When comparing any two Peronosporales species, between 73.79% and 94.9% of orthologs are syntenically conserved. This corresponds to between 34.8% and 68.03% of total genes (supplementary table S4, Supplementary Material online). In *Pythium* species this range is 54.72–91.23% of orthologs or 28.82–68.63% of total genes (supplementary table S4, Supplementary Material online).

Surprisingly *Pl. halstedii* and *Hy. arabidopsidis* are clustered together based on microsynteny (fig. 4), despite having closer related species in the OGOB data set (fig. 1). This suggests that they have convergently evolved similar genome organizations.

Using OGOB to Visualize Expansions in Proteins with Necrosis-Inducing Domains

Necrosis-inducing proteins (NLPs) are apoplastic effectors found in bacteria, fungi, and oomycetes (Feng et al. 2014). The mechanisms by which NLPs act are not fully understood but they are known to induce necrosis, elicit immune responses, and trigger ethylene accumulation in dicotyledons (Oome and Van den Ackerveken 2014). Previously we

reported that putative proteins containing the NLP PFAM domain (PF05630) are significantly overrepresented in numerous *Phytophthora* and Pythiales species but are completely absent from *Albugo*, *Aphanomyces*, and *Saprolegnia* species (McGowan and Fitzpatrick 2017). NLPs are highly expanded in *Phytophthora* species. In particular, *Ph. capsici*, *Ph. ramorum*, *Ph. parasitica*, and *Ph. sojae* have 65, 69, 74, and 80 putative proteins with NLP domains (McGowan and Fitzpatrick 2017).

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

Conclusion

We report here the development of OGOB, a database and tool for performing comparative genomic and synteny analyses of oomycete species. We highlight the usefulness of synteny information in identifying orthologs and use synteny to identify orthologous relationships for 22,708 genes that could not be identified using BLAST searches alone. Phylostratigraphy was used to determine the composition of 20 oomycete genomes and estimate the evolutionary age and emergence of 319,881 oomycete genes. The extent of gene duplication was determined and tandem duplication events were identified as a driving force for the expansion of secreted effector arsenals. Core conserved genes for each oomycete order were identified. Synteny analysis of the 20 oomycete species hosted by OGOB revealed a high degree of syntenic conservation. Our results suggest that conserved genes with housekeeping functions are more likely to be syntenically conserved. Going forward, it is our goal to include additional gold standard genomes from diverse clades to OGOB. For example, currently of the ten recognized *Phytophthora* clades, only data for five clades (clades 1, 2, 7, 8, and 10) are represented. Furthermore, we will also investigate the possibility of implementing robust automated pipelines to locate putative genes that may have been missed at the gene calling stage of annotating genomes. OGOB is a valuable, central resource that will be of interest to plant pathologists and the oomycete community. OGOB is available at https://ogob.ie, last accessed December 31, 2018.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

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

Literature Cited

Adhikari BN, et al. 2013. Comparative genomics reveals insight into virulence strategies of plant pathogenic oomycetes. PLoS One 8(10):e75072.

Altschul SF, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25(17):3389–3402.

Ashburner M, et al. 2000. Gene Ontology: tool for the unification of biology. Nat Genet. 25(1):25–29.

Basenko EY, et al. 2018. FungiDB: an integrated bioinformatic resource for fungi and oomycetes. J Fungi (Basel, Switzerland) 4:39.

Baxter L, et al. 2010. Signatures of adaptation to obligate biotrophy in the *Hyaloperonospora arabidopsidis* genome. Science 330(6010):1549–1551.

Beakes GW, Glockling SL, Sekimoto S. 2012. The evolutionary phylogeny of the oomycete 'fungi'. Protoplasma 249(1):3–19.

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

Bhowmick S, Tripathy S. 2014. A tale of effectors; their secretory mechanisms and computational discovery in pathogenic, non-pathogenic and commensal microbes. Mol Biol. 3:118.

Burki F. 2014. The eukaryotic tree of life from a global phylogenomic perspective. Cold Spring Harb Perspect Biol. 6(5):a016147.

Byrne KP, Wolfe KH. 2005. The Yeast Gene Order Browser: combining curated homology and syntenic context reveals gene fate in polyploid species. Genome Res. 15(10):1456–1461.

Byrne KP, Wolfe KH. 2006. Visualizing syntenic relationships among the hemiascomycetes with the Yeast Gene Order Browser. Nucleic Acids Res. 34(90001):D452–D455.

Carvunis A-R, et al. 2012. Proto-genes and de novo gene birth. Nature 487(7407):370–374.

Caspi R, et al. 2018. The MetaCyc database of metabolic pathways and enzymes. Nucleic Acids Res. 46(D1):D633–D639.

Cheung F, et al. 2008. Analysis of the *Pythium ultimum* transcriptome using Sanger and Pyrosequencing approaches. BMC Genomics 9(1):542.

Coates ME, Beynon JL. 2010. *Hyaloperonospora arabidopsidis* as a pathogen model. Annu Rev Phytopathol. 48(1):329–345.
Creevey CJ, et al. 2004. Does a tree-like phylogeny only exist at the tips in the prokaryotes? Proc R Soc B Biol Sci. 271(1557):2551–2558.
Creevey CJ, McInerney JO. 2005. Clann: investigating phylogenetic information through supertree analyses. Bioinformatics 21(3):390–392.
De Smet R, et al. 2013. Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. Proc Natl Acad Sci U S A. 110(8):2898–2903.
Diéguez-Uribeondo J, et al. 2009. Phylogenetic relationships among plant and animal parasites, and saprotrophs in Aphanomyces (Oomycetes). Fungal Genet Biol. 46(5):365–376.
Domazet-Lošo T, Brajković J, Tautz D. 2007. A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. Trends Genet. 23(11):533–539.
Drost H-G, Gabel A, Grosse I, Quint M. 2015. Evidence for active maintenance of phylotranscriptomic hourglass patterns in animal and plant embryogenesis. Mol Biol Evol. 32(5):1221–1231.
Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32(5):1792–1797.
Enright AJ, Van Dongen S, Ouzounis CA. 2002. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. 30(7):1575–1584.
Fabregat A, et al. 2018. The reactome pathway knowledgebase. Nucleic Acids Res. 46(D1):D649–D655.
Feng BZ, et al. 2014. Characterization of necrosis-inducing NLP proteins in *Phytophthora capsici*. BMC Plant Biol. 14(1):126.
Finn RD, et al. 2017. InterPro in 2017—beyond protein family and domain annotations. Nucleic Acids Res. 45(D1):D190–D199.
Finn RD, et al. 2016. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 44(D1):D279–D285.
Fisher MC, et al. 2012. Emerging fungal threats to animal, plant and ecosystem health. Nature 484(7393):186–194.
Gascuel Q, et al. 2015. The sunflower downy mildew pathogen *Plasmopara halstedii*. Mol Plant Pathol. 16(2):109–122.
Guindon S, et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 2.0. Syst Biol. 59(3):307–321.
Guo Y-L. 2013. Gene family evolution in green plants with emphasis on the origination and evolution of *Arabidopsis thaliana* genes. Plant J. 73(6):941–951.
Haas BJ, et al. 2009. Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. Nature 461(7262):393–398.
Hamilton JP, et al. 2011. The Comprehensive Phytopathogen Genomics Resource: a web-based resource for data-mining plant pathogen genomes. Database (Oxford) 2011:bar053.
Haverkort AJ, et al. 2008. Societal costs of late blight in potato and prospects of durable resistance through cisgenic modification. Potato Res. 51(1):47–57.
He X, Zhang J. 2006. Higher duplicability of less important genes in yeast genomes. Mol Biol Evol. 23(1):144–151.
Jiang RHY, et al. 2013. Distinctive expansion of potential virulence genes in the genome of the oomycete fish pathogen *Saprolegnia parasitica*. PLoS Genet. 9(6):e1003272.
Jiang RHY, Tyler BM. 2012. Mechanisms and evolution of virulence in oomycetes. Annu Rev Phytopathol. 50(1):295–318.
Jiang RHY, Tyler BM, Govers F. 2006. Comparative analysis of *Phytophthora* genes encoding a secreted proteins reveals conserved synteny and lineage-specific gene duplications and deletions. Mol Plant Microbe Interact. 19(12):1311–1321.
Jones P, et al. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30(9):1236–1240.
Kaessmann H. 2010. Origins, evolution, and phenotypic impact of new genes. Genome Res. 20(10):1313–1326.
Kamoun S. 2006. A catalogue of the effector secretome of plant pathogenic oomycetes. Annu Rev Phytopathol. 44(1):41–60.
Kamoun S, et al. 2015. The Top 10 oomycete pathogens in molecular plant pathology. Mol Plant Pathol.16(4):413–434.
Kemen E, et al. 2011. Gene gain and loss during evolution of obligate parasitism in the white rust pathogen of *Arabidopsis thaliana*. PLoS Biol. 9(7):e1001094.
Klopfenstein DV, et al. 2018. GOATOOLS: a python library for gene ontology analyses. Sci Rep. 8(1):10872.
Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 305(3):567–580.
Lamour KH, et al. 2012. Genome sequencing and mapping reveal loss of heterozygosity as a mechanism for rapid adaptation in the vegetable pathogen *Phytophthora capsici*. Mol Plant Microbe Interact. 25(10):1350–1360.
Letunic I, Bork P. 2007. Interactive Tree of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics 23(1):127–128.
Lévesque CA, et al. 2010. Genome sequence of the necrotrophic plant pathogen *Pythium ultimum* reveals original pathogenicity mechanisms and effector repertoire. Genome Biol. 11(7):R73.
Li L, Huang Y, Xia X, Sun Z. 2006. Preferential duplication in the sparse part of yeast protein interaction network. Mol Biol Evol. 23(12):2467–2473.
Links MG, et al. 2011. De novo sequence assembly of *Albugo candida* reveals a small genome relative to other biotrophic oomycetes. BMC Genomics 12:1–12.
Liu D, Hunt M, Tsai IJ. 2018. Inferring synteny between genome assemblies: a systematic evaluation. BMC Bioinformatics 19:26.
Maguire SL, et al. 2013. Comparative genome analysis and gene finding in Candida species using CGOB. Mol Biol Evol. 30(6):1281–1291.
Makkonen J, et al. 2016. Mitochondrial genomes and comparative genomics of *Aphanomyces astaci* and *Aphanomyces invadans*. Sci Rep. 6:36089.
Martens C, Van de Peer Y. 2010. The hidden duplication past of the plant pathogen *Phytophthora* and its consequences for infection. BMC Genomics 11(1):353.
Matari NH, Blair JE. 2014. A multilocus timescale for oomycete evolution estimated under three distinct molecular clock models. BMC Evol Biol. 14(1):101.
McGowan J, Fitzpatrick DA. 2017. Genomic, network, and phylogenetic analysis of the oomycete effector arsenal. mSphere 2(6):e00408–17.
Mi H, et al. 2017. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Res. 45(D1):D183–D189.
Neme R, Tautz D. 2013. Phylogenetic patterns of emergence of new genes support a model of frequent de novo evolution. BMC Genomics 14(1):117.
Ogata H, et al. 1999. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 27(1):29–34.
OhEigeartaigh SS, Armisen D, Byrne KP, Wolfe KH. 2014. SearchDOGS bacteria, software that provides automated identification of potentially missed genes in annotated bacterial genomes. J Bacteriol. 196(11):2030–2042.
Oome S, Van den Ackerveken G. 2014. Comparative and functional analysis of the widely occurring family of Nep1-like proteins. Mol Plant Microbe Interact. 27:1–51.
Ospina-Giraldo MD, Griffith JG, Laird EW, Mingora C. 2010. The CAZyome of *Phytophthora* spp.: a comprehensive analysis of the gene complement coding for carbohydrate-active enzymes in species of the genus *Phytophthora*. BMC Genomics 11(1):525.

Panda A, et al. 2018. EumicrobeDBLite: a lightweight genomic resource and analytic platform for draft oomycete genomes. Mol Plant Pathol. 19(1):227–237.
Papp B, Pál C, Hurst LD. 2003. Dosage sensitivity and the evolution of gene families in yeast. Nature 424(6945):194–197.
Phillippy AM, Schatz MC, Pop M. 2008. Genome assembly forensics: finding the elusive mis-assembly. Genome Biol. 9(3):R55.
Price MN, Dehal PS, Arkin AP. 2010. FastTree 2—approximately maximum-likelihood trees for large alignments. PLoS One 5(3):e9490.
Quint M, et al. 2012. A transcriptomic hourglass in plant embryogenesis. Nature 490(7418):98–101.
Richards TA, Dacks JB, Jenkinson JM, Thornton CR, Talbot NJ. 2006. Evolution of filamentous plant pathogens: gene exchange across eukaryotic kingdoms. Curr Biol. 16(18):1857–1864.
Rizzo DM, Garbelotto M, Hansen EM. 2005. *Phytophthora ramorum*: integrative research and management of an emerging pathogen in California and Oregon forests. Annu Rev Phytopathol. 43(1):309–335.
Sestak MS, Domazet-Loso T. 2015. Phylostratigraphic profiles in zebrafish uncover chordate origins of the vertebrate brain. Mol Biol Evol. 32:299–312.
Sharma R, et al. 2015. Genome analyses of the sunflower pathogen *Plasmopara halstedii* provide insights into effector evolution in downy mildews and *Phytophthora*. BMC Genomics 16:741.
Sperschneider J, Williams AH, Hane JK, Singh KB, Taylor JM. 2015. Evaluation of secretion prediction highlights differing approaches needed for oomycete and fungal effectors. Front Plant Sci. 6:1168.
Tautz D, Domazet-Lošo T. 2011. The evolutionary origin of orphan genes. Nat Rev Genet. 12(10):692–702.
Tyler BM. 2007. *Phytophthora sojae*: root rot pathogen of soybean and model oomycete. Mol Plant Pathol. 8(1):1–8.
Tyler BM, et al. 2006. *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. Science 313(5791):1261–1266.
van den Berg AH, McLaggan D, Diéguez-Uribeondo J, van West P. 2013. The impact of the water moulds *Saprolegnia diclina* and *Saprolegnia parasitica* on natural ecosystems and the aquaculture industry. Fungal Biol Rev. 27(2):33–42.
Waterhouse RM, et al. 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol Biol Evol. 35(3):543–548.
Yachdav G, et al. 2016. MSAViewer: interactive JavaScript visualization of multiple sequence alignments. Bioinformatics 32:btw474.
Yang Z, Nielsen R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol Biol Evol. 17(1):32–43.

Associate editor: Sandra Baldauf