MPMI Vol. 36, No. 1, 2023, pp. 26–46, https://doi.org/10.1094/MPMI-06-22-0133-R

Comparative Genomic Analysis of 31 *Phytophthora* Genomes Reveals Genome Plasticity and Horizontal Gene Transfer

Brent A. Kronmiller,[1,2,†] Nicolas Feau,[3] Danyu Shen,[4] Javier F. Tabima,[5] Shahin S. Ali,[6] Andrew D. Armitage,[7] Felipe Arredondo,[1,2] Bryan A. Bailey,[6] Stephanie R. Bollmann,[8] Angela Dale,[3,9] Richard J. Harrison,[10] Kelly Hrywkiw,[3] Takao Kasuga,[11] Rebecca McDougal,[12] Charlotte F. Nellist,[10] Preeti Panda,[13] Sucheta Tripathy,[14] Nari M. Williams,[12,15] Wenwu Ye,[4] Yuanchao Wang,[4] Richard C. Hamelin,[3,16,17] and Niklaus J. Grünwald[18]

[1] Center for Quantitative Life Sciences, Oregon State University, Corvallis, OR, U.S.A.
[2] Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR, U.S.A.
[3] Department of Forest and Conservation Sciences, The University of British Columbia, Vancouver, Canada
[4] Department of Plant Pathology, Nanjing Agricultural University, Nanjing, China
[5] Department of Biology, Clark University, Worcester, MA, U.S.A.
[6] Sustainable Perennial Crops Laboratory, Northeast Area, USDA/ARS, Beltsville Agricultural Research Center-West, Beltsville, MD, U.S.A.
[7] Natural Resources Institute, University of Greenwich, Chatham Maritime, U.K.
[8] Department of Integrative Biology, Oregon State University, Corvallis, OR, U.S.A.
[9] SC-New Construction Materials, FPInnovations, Vancouver, V6T 1Z4, Canada
[10] NIAB EMR, East Malling, U.K.
[11] Crops Pathology and Genetics Research Unit, Agricultural Research Service, United States Department of Agriculture, Davis, CA, U.S.A.
[12] Scion (Zealand Forest Research Institute), 49 Sala Street, Te Papa Tipu Innovation Park, Private Bag 3020, Rotorua, New Zealand
[13] The New Zealand Institute for Plant and Food Research Ltd, 74 Gerald Street, Lincoln, 7608, New Zealand
[14] CSIR Indian Institute of Chemical Biology, Kolkata, India
[15] Department of Pathogen Ecology and Control, Plant and Food Research, Private Bag 1401, Havelock North, New Zealand
[16] Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec, Canada
[17] Département des sciences du bois et de la forêt, Faculté de Foresterie et Géographie, Université Laval, Québec, Canada
[18] Horticultural Crop Research Unit, United States Department of Agriculture, Agricultural Research Service, Corvallis, OR, U.S.A.

Accepted for publication 24 October 2022.

[†]Corresponding author: B. A. Kronmiller; brent.kronmiller@oregonstate.edu

Nicolas Feau, Danyu Shen, and Javier F. Tabima contributed equally to this work. Author order is alphabetical.

Funding: This work was supported in part by grants to Y. Wang from China National Funds for Distinguished Young Scientists (31225022). Genome sequencing of *Phytophthora pini* was supported by the California Walnut Board and the United States Department of Agriculture Agriculture Research Service (USDA ARS), CRIS Project grant 5306-22000-014-00D to T. Kasuga. N. J. Grunwald was supported by USDA ARS project 2072-22000-041-000-D, National Institute of Food and Agriculture grant 2018-67013-27823, and the J Frank Schmidt Foundation. R. J. Harrison, A. D. Armitage, and C. F. Nellist were supported by grants from the U.K. Biotechnology and Biological Sciences Research Council (BB/K017071/1, BB/K017071/2, and BB/N006682/1). R. McDougal, P. Panda, and N. Williams were funded by the New Zealand Ministry of Business, Innovation and Employment (grant number CO4X1305), the Forest Growers Levy Trust (administered by the New Zealand Forest Owners Association) and the Radiata Pine Breeding Company under the "Healthy trees, Healthy future" research program at Scion (NZFRI, Ltd).

Phytophthora **species are oomycete plant pathogens that cause great economic and ecological impacts. The *Phytophthora* genus includes over 180 known species, infecting a wide range of plant hosts, including crops, trees, and ornamentals. We sequenced the genomes of 31 individual *Phytophthora* species and 24 individual transcriptomes to study genetic relationships across the genus. *De novo* genome assemblies revealed variation in genome sizes, numbers of predicted genes, and in repetitive element content across the *Phytophthora* genus. A genus-wide comparison evaluated orthologous groups of genes. Predicted effector gene counts varied across *Phytophthora***

e-Xtra: Supplementary material is available online.

The author(s) declare no conflict of interest.

species by effector family, genome size, and plant host range. Predicted numbers of apoplastic effectors increased as the host range of *Phytophthora* species increased. Predicted numbers of cytoplasmic effectors also increased with host range but leveled off or decreased in *Phytophthora* species that have enormous host ranges. With extensive sequencing across the *Phytophthora* genus, we now have the genomic resources to evaluate horizontal gene transfer events across the oomycetes. Using a machine-learning approach to identify horizontally transferred genes with bacterial or fungal origin, we identified 44 candidates over 36 *Phytophthora* species genomes. Phylogenetic reconstruction indicates that the transfers of most of these 44 candidates happened in parallel to major advances in the evolution of the oomycetes and *Phytophthora* spp. We conclude that the 31 genomes presented here are essential for investigating genus-wide genomic associations in genus *Phytophthora*.

Keywords: effectors, horizontal gene transfer, oomycete plant pathogens, *Phytophthora*

Members of the *Phytophthora* genus are oomycete plant pathogens that collectively infect a wide range of plants (Erwin and Ribeiro 1996; Judelson and Blanco 2005) and cause great economic, environmental, and societal impact (Drake and Jones 2017). Oomycetes are morphologically similar to filamentous fungi (Gunderson et al. 1987; Jiang and Tyler 2012; Thines 2014; Thines and Kamoun 2010) but are classified as stramenopiles, a group that also includes diatoms and brown algae (Dick 2001; Gunderson et al. 1987). Oomycetes include many plant-pathogenic species besides *Phytophthora*, including numerous *Pythium* species that cause seed, seedling, root, and fruit rots and a broad diversity of obligate biotrophs that cause downy mildew.

Phytophthora species infect numerous plants, including crops, trees, and ornamentals, in managed and natural ecosystems. The agent responsible for potato blight, *Phytophthora infestans* triggered the 1840s potato famine (Haas et al. 2009; Turner 2005), and *Phytophthora ramorum* is responsible for sudden oak death in North America and sudden larch death in the United Kingdom, which have destroyed millions of trees in addition to infecting hundreds of additional tree and ornamental species (Goheen et al. 2002; Grunwald et al. 2008; Rizzo et al. 2002, 2005). Some *Phytophthora* species are relatively host-specific, such as the soybean pathogen *Phytophthora sojae* (Tyler 2007), the strawberry pathogen *Phytophthora fragariae* (Kennedy et al. 1986), and the lychee pathogen *Phytophthora litchi* (Kao and Leu 1980). In contrast, others including *Phytophthora cinnamomi* (Hardham 2005), *Phytophthora palmivora* (Erwin and Ribeiro 1996), and *Phytophthora parasitica* (Erwin and Ribeiro 1996), can infect a vast assortment of plant hosts. The mechanistic basis of this large variation in apparent host specificity is currently unknown (Haas et al. 2009; Raffaele and Kamoun 2012).

There are over 180 known *Phytophthora* species (Kroon et al. 2012; Yang et al. 2017) phylogenetically falling into 12 or more clades (Brasier et al. 2022; Jung et al. 2017; Kroon et al. 2012) and further divided into numerous subclades (Yang et al. 2017). Previous genome sequencing studies have examined species within the *Phytophthora* genus (Ali et al. 2017; Armitage et al. 2018; Dong et al. 2015; Jung et al. 2017; Raffaele and Kamoun 2012; Tabima et al. 2017; Turner 2005; Yang et al. 2017; Ye et al. 2016a), with *P. sojae* and *P. ramorum* (Tyler 2006), *P. infestans* (Haas et al. 2009; Knaus et al. 2020), and *Phythophthora capsici* (Lamour et al. 2012) serving as key models for the genus. Overall, these studies revealed highly dynamic genomes containing both rapidly evolving and conserved regions.

Phytophthora species infect plant hosts through the use of two broad classes of secreted effector proteins (Asai and Shirasu 2015; Jiang and Tyler 2012; McGowan and Fitzpatrick 2017; Wang and Jiao 2019; Wang et al. 2017). Apoplastic effectors act outside the plant cells and include glycoside hydrolases, necrosis-inducing proteins (NLPs) (Dong et al. 2012), proteases, lipases, lipid-binding proteins, and protease inhibitors (Tian et al. 2007). Roles of apoplastic effectors include weakening of plant physical and chemical defenses and are a source of nutrition early in infection. In contrast, cytoplasmic effectors enter plant cells, often through the differentiation of specialized structures called haustoria, and include RxLR effectors (Jiang et al. 2008; Morgan and Kamoun 2007; Tyler 2006), crinkler effectors (CRN) (Schornack et al. 2010; Stam et al. 2013; Torto 2003; Win et al. 2006; Zhang et al. 2014) and non-conventionally secreted effectors (Liu et al. 2014). In oomycetes, as well as other pathogens, cytoplasmic effectors manipulate numerous aspects of host physiology and morphology to promote susceptibility, including suppression of host immunity and programmed cell death (Oh et al. 2009; Wang et al. 2011), stimulating and inhibiting the release of nutrients (Bozkurt et al. 2011; Caillaud et al. 2012; Jiang and Tyler 2012; Torto-Alalibo et al. 2010; Wang et al. 2019). The genomes of oomycete pathogens sequenced to date include large rapidly evolving gene families encoding these effectors (Haas et al. 2009; Lamour et al. 2012; Tyler 2006). Many of these effector genes, especially those encoding RxLR effectors, display evidence of accelerated evolution due to host-pathogen coevolutionary conflict (Jiang et al. 2008; Shen et al. 2013; Ye et al. 2016b).

Examination of previously sequenced genomes has identified two distinct partitions, namely gene-dense, repeat-poor regions and gene-sparse, repeat-rich regions (Haas et al. 2009; Knaus et al. 2020; Lamour et al. 2012; Tyler 2006). Highly conserved housekeeping genes are typically found in gene-dense regions, while rapidly evolving gene families associated with infection are typically found in gene-sparse regions that are transposon-rich (Haas et al. 2009; Gijzen 2009; Knaus et al. 2020; Lamour et al. 2012; Tyler 2006). This arrangement has been labeled "the two-speed genome" (Dong et al. 2015; Raffaele and Kamoun 2012). It has been hypothesized that transposons in the gene-sparse, transposon-rich regions may contribute to the genomic diversity and possibly to epigenetic variability of expression of genes in those regions including infection-associated genes (Wang et al. 2020).

To investigate phylogenetic relationships, horizontal gene transfer (HGT), effector genomics, and possible mechanisms underlying host ranges across members of the genus *Phytophthora*, we sequenced 31 genomes, using Illumina short read technology. Our newly sequenced genomes include most species in clade 7 as well as representative species from nine of the phylogenetic clades (Blair et al. 2008). Several genomes have already been published individually as a result of this project, i.e., *Phytophthora litchii* (Ye et al. 2016a), *Phytophthora megakarya* and *Phytophthora palmivora* (Ali et al. 2017), *Phytophthora fragariae* and *Phytophthora rubi* (Tabima et al. 2017), and *Phytophthora cactorum* (Armitage et al. 2018; Nellist et al. 2021). Here, we present a combined analysis of the genome sequences of 37 *Phytophthora* species, including the 31 species newly sequenced by our sequencing consortium, resulting in the first large-scale comparative genomic study including species from nine *Phytophthora* clades. This work provides insights into genome architecture and evolution in the genus *Phytophthora* as well as novel genomic resources of broad interest.

Results

Sequencing and assembly of the 31 *Phytophthora* genomes.

The 31 *Phytophthora* genomes sequenced by our consortium produced between 16.6 and 44.7 million raw reads per genome (Table 1; Supplementary Table S1). The sizes of the *Phytophthora* genome assemblies varied greatly, ranging from 37.3 to 107.8 Mb (mean 61.3 ± 17.6 Mb). Due to variations in the number of reads per genome and differences in genome sizes, the quality of assembled genomes varied as well. The number of contigs per assembly ranged from 2,131 to 28,263 (mean 11,380 ± 7,919.2). The more completely assembled genomes benefited from deeper coverage. For example, the *Phytophthora boehmeriae* assembly was composed of 2,866 contigs with an N_{50} of 41,917 bp; this assembly benefitted from 87× coverage as a result of receiving 34.7 million reads to cover a 39.7 Mb assembly. In contrast, the *Phytophthora lateralis* assembly had only 45× coverage (23.2 million reads across a 50.5 Mb assembly size), resulting in 28,263 contigs and an N_{50} of 2,396 bp (Table 1). Sequence read length did not seem to make a difference in assembly quality. Two genomes (*P. cactorum* and *Phytophthora idaei*) with 250-bp reads (compared with most assemblies with 50-bp reads) had an N_{50} and number of contigs that were not better than the other genomes when compared with genome size (*P. cactorum*: contigs 7,888, N_{50} 15,053, size 56 Mb; *P. idaei*: contigs 7,163, N_{50} 14,461, size 53.5 Mb) (Table 1).

To aid with gene calling and identify active genes, RNA sequencing was conducted for 25 *Phytophthora* species on V8-grown mycelia and either germinated cysts or Plich-grown mycelia for those species that did not readily yield zoospores (van der Lee et al. 1997). RNA sequencing produced 12.5 to 36.0 million paired-end sequence reads per sample, with resulting transcriptome assemblies averaging 36,189 contigs per assembly, with an average N_{50} of 1,822 bp. Transcriptome assemblies for most species had between 20,000 and 40,000 contigs, with the exception of *Phytophthora parvispora*, which produced more than 109,000 contigs. TransDecoder, which reduces duplication in *de novo* transcriptome assemblies by identifying candidate coding regions and removing duplicates, was run on the Trinity assemblies and lowered average transcriptome assembly to 33,200 contigs and raised average N_{50} to 2,753 bp. Transcriptome size differences between species are not correlated with genome size differences with the two largest genome assemblies (*P. megakarya* and *P. palmivora*) both resulting in midrange transcriptome assembly sizes (Supplementary Table S1).

De novo repeat identification and analysis.

Repeats were identified, classified, and masked to prepare genomes for gene prediction. *De novo* repeat prediction identified between 27 and 295 different repeat subfamilies per species. The percent repeat content of the assemblies varied greatly across the *Phytophthora* genomes sequenced. The genome assemblies of *Phytophthora kernoviae*, *P. litchii*, and *Phytophthora agathidicida* contained very low repeat content of 4.15, 5.76, and 5.98%, respectively (Fig. 1). On the other end, several genome assemblies had high repeat content, namely, *P. megakarya*, *Phytophthora hibernalis*, and *Phytophthora pinifolia* contained 32.94, 32.46, and 33.00%, respectively.

Repeat annotations were classified into types (class I retrotransposons, class II DNA transposons, and other), families, and subfamilies (Fig. 1). Fifteen different DNA transposons were identified across all 31 species. Long terminal repeat (LTR) retrotransposons were more diverse across *Phytophthora* species. As many as 57 Copia LTR retrotransposons and 144 Gypsy LTR retrotransposon types were found in each genome assembly.

Table 1. Sequencing and assembly statistics for 31 *Phytophthora* species, grouped by clade

		Genome			Annotations	
Species	**Clade**	**Contigs**	**N_{50}**	**Assembly length**	**Repeat percent**	**Predicted genes**
P. cactorum	1	7,888	15,053	56,443,298	19.96	18,027
P. idaei	1	7,163	14,461	53,468,943	16.19	18,038
P. pini	2	2,131	42,987	38,730,000	7.31	14,019
P. multivora	2	2,844	46,133	40,059,192	10.86	13,682
P. pluvialis	3	4,340	30,816	53,616,150	16.04	16,285
P. litchii	4	2,543	34,546	38,200,938	5.98	12,391
P. palmivora	4	24,815	6,694	107,798,747	29.62	37,283
P. megakarya	4	24,073	7,093	101,609,312	31.94	33,614
P. agathidicida	5	3,754	19,544	37,337,699	5.76	12,923
P. taxon totara	5	4,425	30,809	55,576,372	16.56	17,619
P. parvispora	7	9,906	6,820	46,825,958	8.75	15,642
P. pisi	7	7,667	15,253	58,856,683	16.80	18,953
P. robiniae	7	14,865	8,754	69,938,814	25.58	23,128
P. niederhauseri	7	26,463	4,805	90,270,009	20.96	29,587
P. cajani	7	18,255	5,113	64,854,085	20.65	19,840
P. vignae	7	10,330	8,363	56,137,732	17.45	18,535
P. melonis	7	11,353	15,342	73,416,743	25.93	21,276
P. pistaciae	7	10,414	10,302	63,209,321	18.56	19,423
P. uliginosa	7	8,955	10,095	57,072,031	24.16	17,226
P. europaea	7	8,301	11,551	58,787,065	23.33	17,117
P. fragariae	7	8,544	20,362	76,969,737	30.81	20,448
P. rubi	7	9,434	17,808	74,863,594	29.48	23,476
P. pinifolia	6	22,610	6,021	74,478,861	33.00	23,717
P. lateralis	8	28,263	2,396	50,496,828	23.39	19,503
P. hibernalis	8	6,587	21,408	71,256,216	32.46	23,578
P. foliorum	8	5,320	15,800	48,973,082	19.26	16,083
P. brassicae	8	12,447	12,337	72,849,437	28.39	26,010
P. syringae	8	6,572	15,987	57,045,526	21.71	18,234
P. cryptogea	8	25,944	4,730	69,446,343	17.65	24,936
P. boehmeriae	10	2,866	41,917	39,747,814	7.83	13,325
P. kernoviae	10	13,710	5,225	42,698,878	4.15	14,322

Phytophthora gene prediction and annotation.

Gene predictions ranged from 12,391 to 37,283 (mean 19,943 ± 5,864) across the newly sequenced genomes. These gene counts per species are within the general range of previously published *Phytophthora* genomes (*P. sojae*, *P. ramorum*, *P. infestans*, *P. capsici*, *P. cinnamomi*, and *P. parasitica*) ranging from 16,066 to 26,584 genes per species, considering that we observed higher gene counts in genomes with genome duplications (*P. palmivora* and *P. megakarya* [Ali et al. 2017]).

Some of the previously published *Phytophthora* genome sequences were annotated with the MAKER gene prediction process outlined here to validate the methods. The *P. capsici* genome was reported to contain 19,805 predicted genes (Lamour et al. 2012), while we obtained 18,917 predicted genes. The *P. sojae* genome v3.0 was reported to contain 26,584 predicted genes (Tyler 2006), whereas we identified 21,447 MAKER-predicted genes. Therefore, our MAKER pipeline may slightly undercount the gene content compared with other methods.

The completeness of the gene sets predicted from the genomes and transcriptomes was assessed by identifying single-copy core orthologs using the benchmarking universal single-copy orthologs (BUSCO) pipeline with the Alveolata_Stramenopiles database (234 genes) as the reference. BUSCO analysis of the 31 genome assemblies identified 147 to 231 complete genes (mean 204 ± 27) of the 234 single-copy genes in the database (Fig. 2A). Analysis of the 24 transcriptome assemblies identified 18 to 228 complete genes (mean 187 ± 54) (Fig. 2B). The *P. kernoviae* and *P. lateralis* transcriptomes were outliers, with only 20 and 18 complete genes identified, respectively. When *P. kernoviae* and *P. lateralis* were removed, transcriptome assemblies ranged from 142 to 228 complete (single and duplicated) genes (mean 203 ± 18). Results of BUSCO analyses run on predicted proteins ranged from 139 to 222 complete (single and duplicated) genes (mean 192 ± 25) of the 234 conserved orthologous proteins in the database (Fig. 2C).

Predicted proteins from the MAKER gene prediction were functionally annotated by matching to published *Phytophthora*, stramenopile, and fungal proteins. Across the 614,862 proteins predicted in the 31 *Phytophthora* species, when aligned to the National Center for Biotechnology Information (NCBI) database and UniProt TrEMBL, 294,146 *Phytophthora* proteins produced alignments that passed the cutoff filter. Removing 'Uncharacterized Protein' or similarly uninformative functional annotations yielded 196,652 proteins with functional classifications (mean 6,343.6 ± 3,191.7). When aligned to the *Phytophthora*, stramenopile, and fungal sequence databases, 445,458 proteins passed the alignment cutoffs, with 304,951 proteins (mean 9,837.1 ± 2,523.3) that had informative functional annotations.

InterProScan was used to identify domains and motifs in all predicted proteins. Of the 614,862 proteins, 173,727 had domains identified. Gene ontology (GO) terms were assigned from BLASTX alignments between the UniProt BLASTX alignments and the InterProScan predictions, identifying 321,953 proteins with GO terms assigned.

Effector protein identification in *Phytophthora* spp.

The cytoplasmic effector identification process predicted a total of 10,354 RxLR effector proteins and 4,415 CRN effector proteins from the genomes and transcriptomes of the species

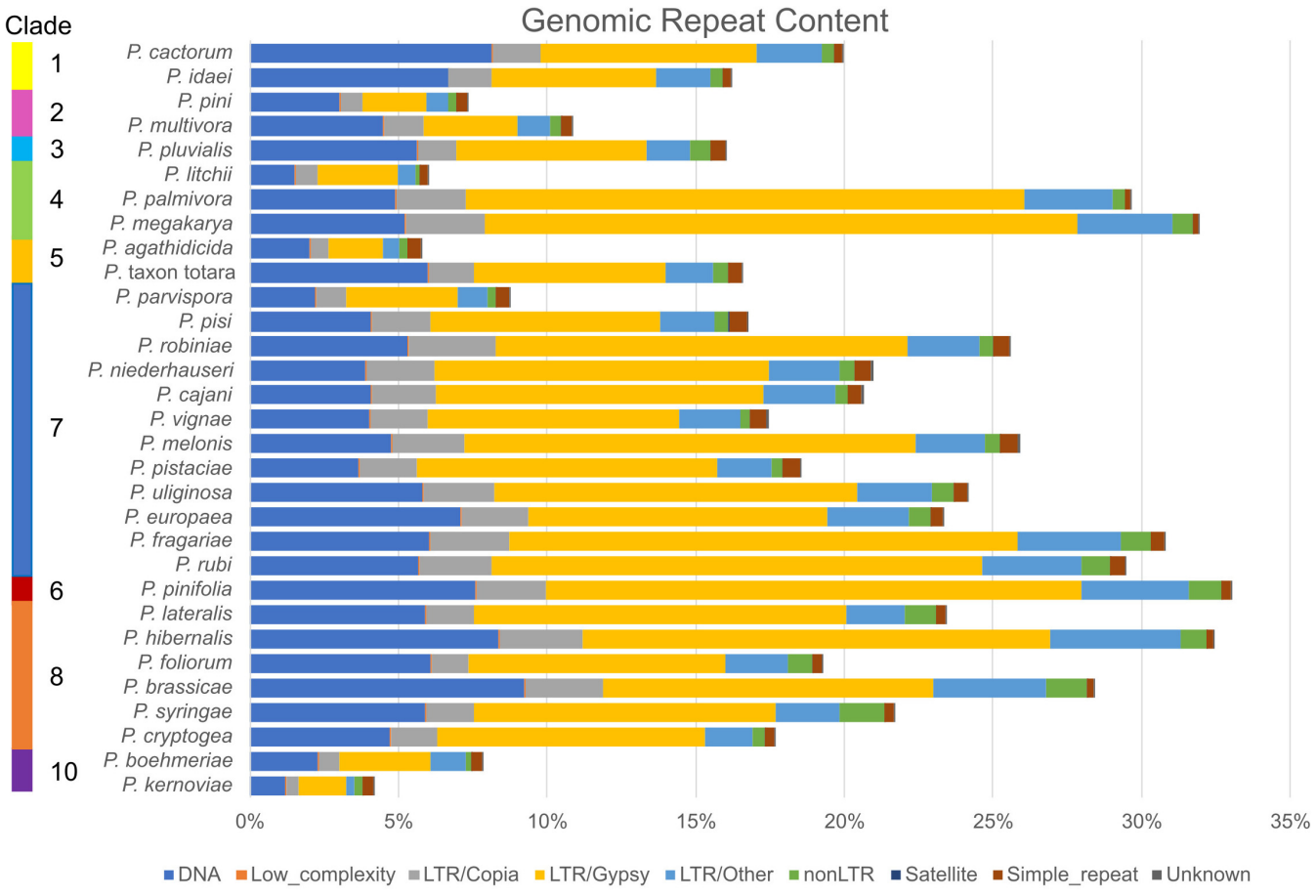

Fig. 1. Analysis of 31 *Phytophthora* species shows an abundance of repetitive elements. Species are shown in phylogenetic clade order; clade designations are shown on the left. Repeat content is displayed as percentage of the total genome content. Repeat classifications are shown as colored bar segments.

sequenced in this study (Fig. 3). The numbers of predicted RxLR effector genes differed greatly across genomes. *P. pinifolia* exhibited the lowest number, with 46 predicted RxLR effectors, while *P. megakarya* exhibited the highest number, with 1,183 predicted proteins. We also observed great variation in the number of predicted CRN effectors, with *P. litchii* showing the lowest number of CRN effectors, 27, while *Phytophthora cajani* showed the highest, 274.

Our search for apoplastic effectors identified 6,671 glycoside hydrolases, 1,191 NLPs, and 1,046 protease inhibitors across the 31 *Phytophthora* genomes (Fig. 3). The predicted glycoside hydrolase genes range from 139 (*Phytophthora brassicae*) to 386 (*P. palmivora*). The numbers of NLP genes range from five (*P. pinifolia*) to 78 (*Phytophthora niederhauserii*). The counts of protease inhibitor genes range from 23 (*P. brassicae*) to 67 (*P. palmivora*).

Orthology clustering of *Phytophthora* proteins.

The 715,980 predicted genes from the 31 genome assemblies along with those of *P. sojae*, *P. ramorum*, *P. infestans*, *P. capsici*, *P. cinnamomi*, and *P. parasitica* were subjected to orthology analysis. In the first step, the *Phytophthora* genes were matched against the pre-computed publicly available orthologous groups. From this analysis, 560,201 genes were assigned to 7,829 unique clusters, leaving 155,779 genes unassigned. In step two, these remaining genes were clustered using OrthoMCL, yielding 13,474 additional unique clusters. In total, the 715,980 genes were assigned to 21,303 orthologous clusters.

Orthologous groups were assigned functional annotations based on the proteins that composed the group. Of the 21,303 orthologous clusters, 9,806 could be assigned informative functional annotations, as defined in the gene annotation section.

The numbers of genes present in orthologous groups encompassing all 37 *Phytophthora* species shows a bimodal frequency distribution (Fig. 4), with peaks observed at one to six genes per group and 34 to 41 genes per group. We hypothesize this first peak represents rapidly evolving genes that are conserved in only a few of the 37 *Phytophthora* species in this study. This first peak was much smaller when only genes with meaningful annotations were considered, suggesting an enrichment for genes that have previously uncharacterized functions. This peak included both small orthologous groups in which all genes were from the same *Phytophthora* clade and groups consisting of genes from multiple clades. The second large peak was centered at 37 genes per ortholog group, thus representing ortholog groups that have one gene per species. Additional groups include one or a few species missing the orthologous gene or one or a few species with a second copy of an orthologous gene.

Phylogenetic relationship across *Phytophthora* species.

We reconstructed the phylogenetic relationships of the sequenced *Phytophthora* species (Fig. 5) based on 61 single-copy core orthologous genes shared across 37 species. Predicted genes and amino acid protein sequences used to build the phylogeny are found in Supplementary Table S2. The RAxML phylogenetic tree clustered these species into phylogenetic clades consistent with previous studies (Blair et al. 2008; Martin et al. 2014; McCarthy and Fitzpatrick 2017; Kroon et al. 2012; Yang et al. 2017), with the exception of *Phytophthora taxon totara* placed into clade 5. The separation of *P. taxon totara* from clade 3 containing *Phytophthora pluvialis* has been reported previously (McCarthy and Fitzpatrick 2017).

Relationship of *Phytophthora* effector gene numbers to plant host range.

To examine the assemblies for clues as to the genomic basis for the diverse host ranges of the sequenced species, *Phytophthora* species were categorized into plant host ranges as follows. Thirteen species were defined as having narrow, 12 having multiple, six having wide, and six having huge host ranges (Fig. 5). The numbers of genes predicted to encode various families of effectors were plotted for each host range class. Paralogous effectors with greater than 95% nucleotide identity over the full sequence length were counted only once. Counts were plotted

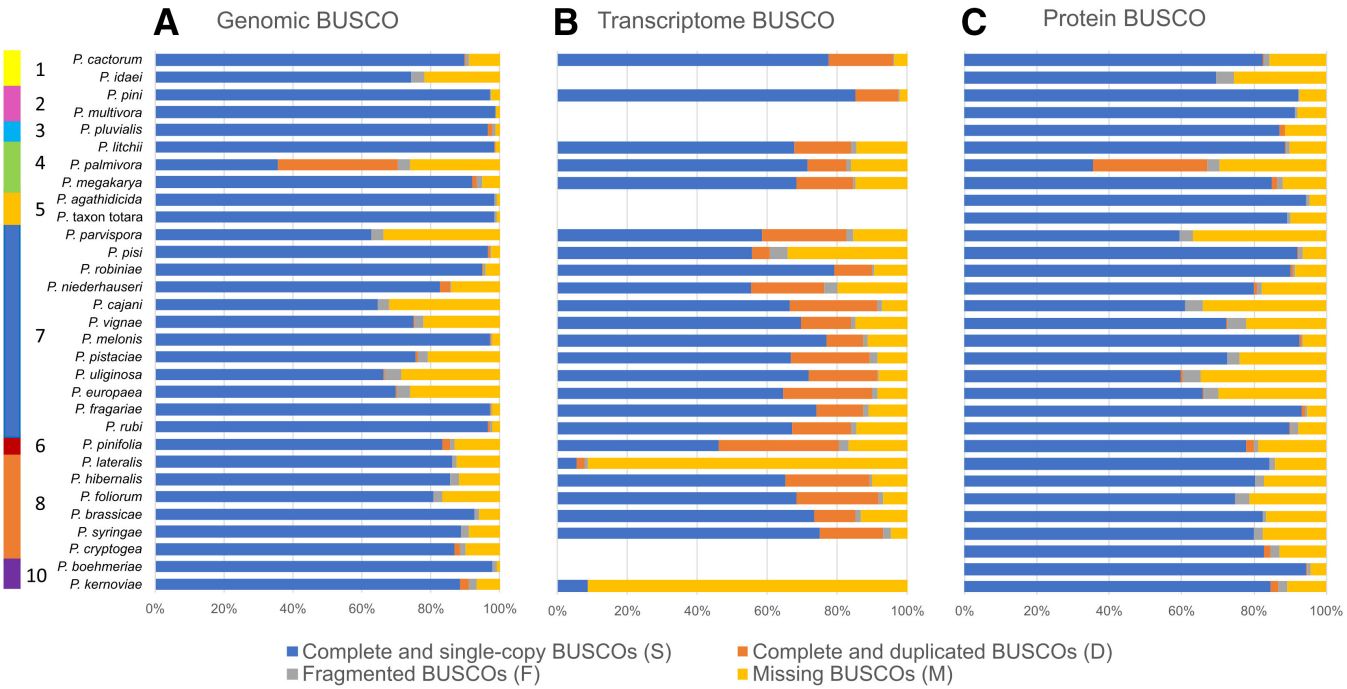

Fig. 2. Benchmarking universal single-copy orthologs (BUSCO) analysis demonstrates completeness of the 31 *Phytophthora* species in this study. **A,** Genomic assembly, **B,** transcriptome assembly, and **C,** predicted proteins. Species are shown in phylogenetic clade order and clade designations are shown on the left. For each BUSCO analysis, results from searching 234 single-copy orthologs in the Alveolata_Stramenopiles dataset are shown.

for each of two cytoplasmic effector categories (RxLR and CRN) and three apoplastic effector categories (glycoside hydrolases, protease inhibitors, and NLPs) (Fig. 6). Subcategories of the apoplastic effectors glycoside hydrolases and protease inhibitors were individually plotted (Supplementary Fig. S1A and B). In each case, numbers of genes predicted to encode each effector subcategory for each species were plotted against the host range.

Apoplastic effectors show a distinct overall pattern; *Phytophthora* species with smaller host ranges had fewer predicted effector genes, with numbers of predicted effector genes increasing with increased host range (Fig. 6A, B, and C). Predicted cytoplasmic effector genes show a similar pattern, starting with low numbers of predicted genes in *Phytophthora* species with narrow host ranges and increasing in those with multiple and wide host ranges. However, for both the RxLR and CRN effector categories, the numbers of predicted genes per species decrease from the wide to the huge host range (Fig. 6D and E). Two species were outliers with respect to the number of apoplastic NLP effector genes, *Phytophthora pistaciae* in the narrow host range category and *P. sojae* in the multiple host range category had many more predicted apoplastic effector genes than the other species in each of their host range categories, respectively. Four species were outliers with respect to the numbers of predicted RxLR effector genes, namely, *P. megakarya* and *P. pistaciae* in the narrow host range category, *P. parvispora* in the multiple host range category, and *P. palmivora* in the huge host range category all had many more predicted RxLR genes than the other species in those categories. One outlier was observed in the CRN effector genes; *P. infestans* in the wide host range category had many more genes than the other species in that category.

HGT.

We evaluated all 31 genomes for evidence of HGT. We used machine learning to identify HGT candidates and phylogenetic approaches to validate candidate HGT genes.

Support vector machine (SVM) classifier predicted HGT candidates. Analysis of the 722,232 transcripts with our SVM classifier over the 31 genome assemblies and the six previously published *Phytophthora* genomes identified 35,246 HGT candidates. A total of 28,791 of these transcripts that could be regrouped in orthology groups encoding putative transposable elements (TEs) were discarded, resulting in 6,455 non-TE HGT candidates. The number of candidates predicted ranged from 91 in *P. agathidicida* to 233 in *P. megakarya* and 262 in *P. palmivora* (mean 160.13 ± 36.68). *P. agathidicida* has one of the lowest numbers of transcripts annotated (12,923 transcripts), while *P. palmivora* and *P. megakarya* have the highest gene content with 37,283 and 33,614 transcripts, respectively. Overall, we identified a significant linear correlation between gene space in each genome and the number of HGT candidates predicted with the SVM classifier ($r^2 = 0.45$; $P < 0.0001$).

Phylogenetic filtering of HGT candidates. The 6,455 non-TE HGT candidates predicted with the SVM classifier were subject to a two-step filtering process to discard false positives. In the first step, we searched for homologs among a database of sequences built from seven clades (including putative fungal and bacterial donors), followed by phylogenetic tree reconstruction with bootstrap analysis. The phylogenetic filter retained 2,214 candidates, among which 1,113 (50.3%) showed a strong phylogenetic discordance and were seen nested within a distantly related clade (Fungi, Bacteria, or Amoebozoa) in direct contra-

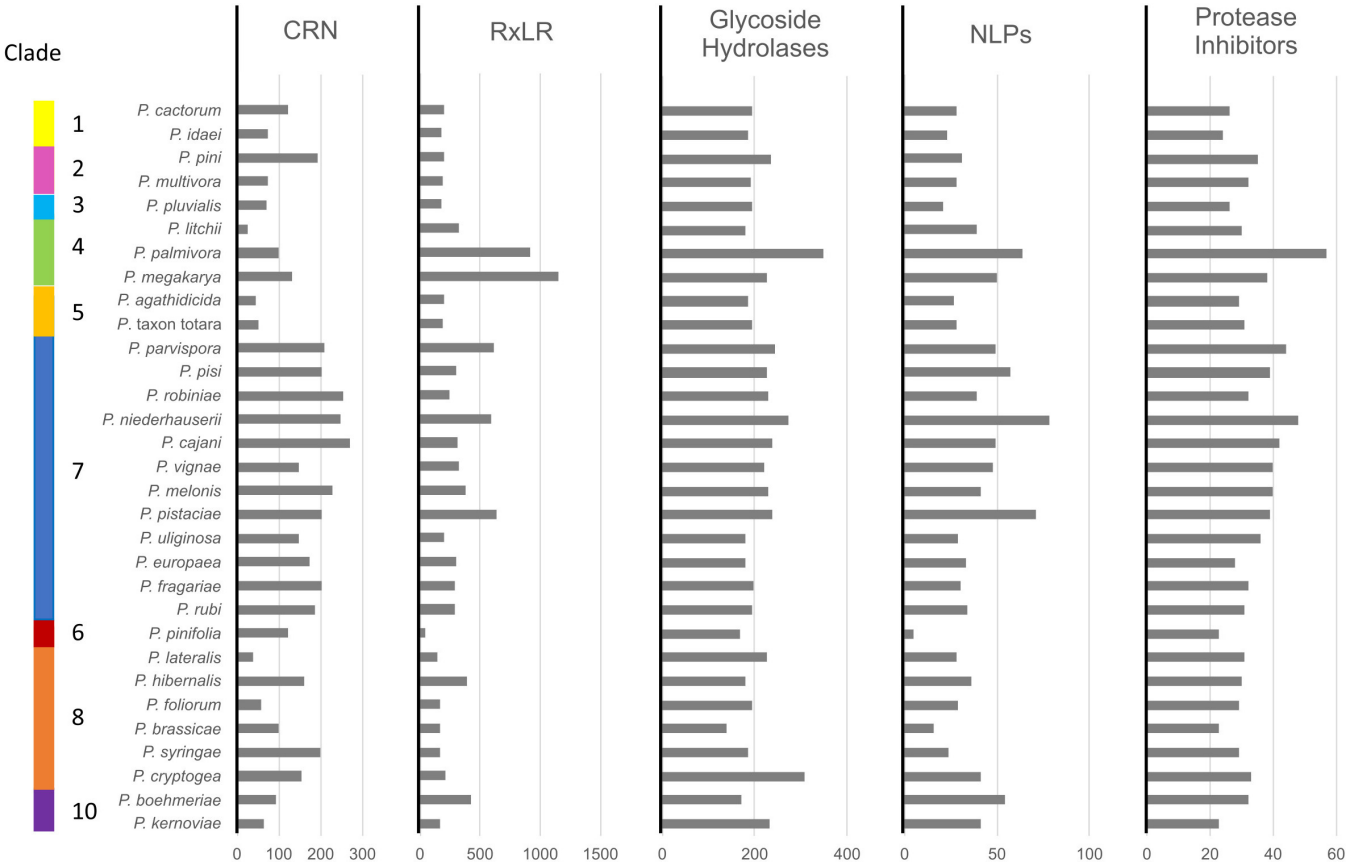

Fig. 3. The number of predicted effectors varied across the *Phytophthora* genomes. Bar chart representing amounts of effector genes found in 31 *Phytophthora* species for crinkler (CRN), RxLR, glycoside hydrolases, necrosis-inducing proteins (NLPs), and protease inhibitors. Species are shown in phylogenetic clade order, clade designations are shown on the left.

diction to the expected phylogenetic relationships of the respective organisms. The 1,110 other candidates left also branched within a clade of fungal or bacterial genes; however, in these cases, placement of the *Phytophthora* transcript with Fungi or Bacteria was caused by the absence of homologs in one or more of the intermediate clades Viridiplantae, Alveolata, and Amoebozoa. The number candidates that passed this filter ranged from 36 in *Phytophthora europea* to 127 in *P. palmivora*. There was a significant correlation between the total number of genes and the number of HGT candidates across the analyzed species ($r^2 = 0.49$; $P < 0.0001$) (Supplementary Fig. S2).

Sequence identity filtering of HGT candidates. The 2,214 HGT candidates retained after the phylogenetic filtering were submitted to a sequence identity discrepancy filter. A total of 1,688 candidates were rejected after the first "identity test," resulting in a "relaxed" set of 526 HGT candidates for which the sequence identity between the *Phytophthora* candidate HGT sequence and its closest homolog sequences in the putative donor species was shorter than the average identity between the two species; an average of 14.6 (±6.2) candidates were retained per *Phytophthora* genome, with a maximum of 33 for *P. palmivora* and *P. niederhauserii* (Supplementary Fig. S2). A gene enrichment analysis of this candidate set showed a significant enrichment for GO terms related to oxidoreductase activity and hydrolysis and metabolism of carbohydrates (cutinase activity, carbohydrate metabolic process) and proteins (Supplementary Table S3).

Among the 526 candidates of the relaxed set, 44 passed the second identity test, constituting a "strict" set of HGT candidates. For 28 of them (56.0%), BLAST search results indicated a strong homology with the clade of the putative donor where the BLAST E-values with species of the putative donor clade were lower than the E-values observed with species from non-donor clades. We then looked at their physical location on their respective scaffold to eliminate potential contaminants. All the candidates were found on scaffolds that had at least two gene models predicted on them. GO term enrichment analysis of this strict set indicated significant enrichment for GO terms related to oxidoreductase activity (GO:0055114, GO:0016491, GO: 0008670), and carbohydrate activity and cell-wall modification (GO:0000272, GO:0045490, GO: 0042545, GO:0045493, GO:0031176, GO:0030599) (Table 2).

Phylogenetic reevaluation of the strict set of HGT candidates. The 44 candidates of the strict set of HGT candidates were subjected to reevaluation by sampling additional taxa within the oomycetes. Their amino-acid sequences were first clustered into closely related groups of sequences by assigning them to the 21,303 orthologous clusters previously defined (discussed above). This process reduced the set of 44 candidates into 28 orthologous clusters that were then searched against the sequences of the 31 genome assemblies, five *Phytophthora* species sequenced in previous studies, and 36 oomycete genomes (Supplementary Table S4). Protein members of six of these clusters had homologs (BLASTP E-values $\leq$ 1e-025) (Supplementary Table S4) in the set of 21 strongly supported HGT candidates identified in the genome of *P. ramorum*, *P. infestans*, and *P. sojae* by Richards et al. (2011). Following these searches, we reconstructed maximum likelihood phylogenies for 19 of these can-

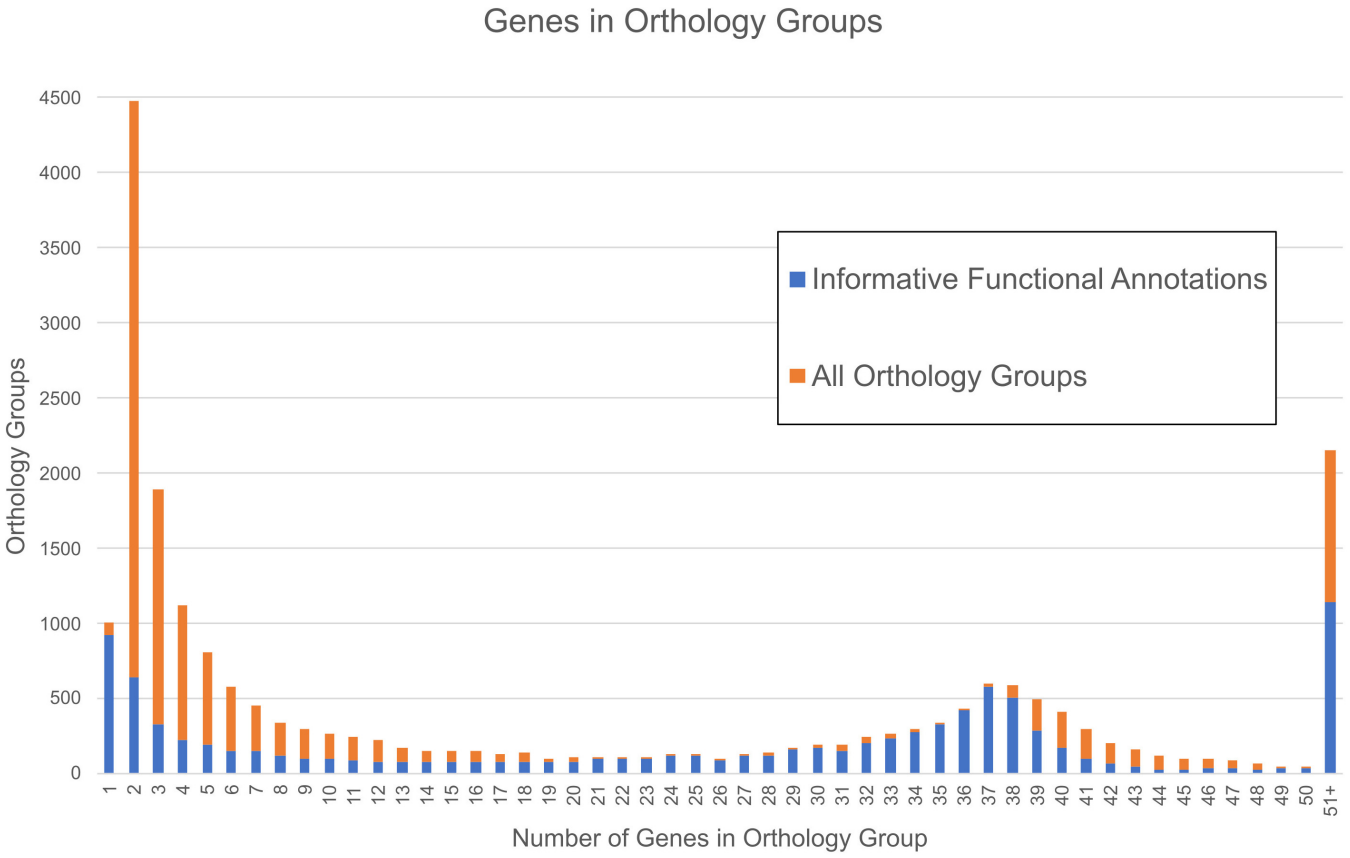

Fig. 4. Distribution of genes in *Phytophthora* orthology shows a bimodal frequency distribution, highlighting genes that are conserved in only a few genomes and orthology groups that have one gene per species. Numbers of genes assigned to orthology clusters with OrthoMCL (Li 2003) are shown. The 31 sequenced *Phytophthora* spp. and six additional previously sequenced *Phytophthora* genomes (*P. capsici*, *P. cinnanomi*, *P. infestans*, *P. parasitica*, *P. ramorum*, *P. sojae*) are included. Bars representing all orthology groups (orange) show all genes assigned into orthology groups; those representing informative functional annotations show genes assigned into orthology groups that have useful functional definitions and exclude genes labeled as 'uncharacterized', 'hypothetical', or similar.

didates that had putative functions related to the modification of compounds of the plant cell wall (e.g., pectin esterase, xylulose reductase, tannase, and endo-1,4-beta-xylanase), peptidases, oxidoreductases, and putative elicitors such as a NPP1 protein and an ATP-binding cassette (ABC) transporter (Supplementary Fig. S3). In 15 cases, the HGT candidate was found nested within a group of fungi or bacteria, as expected under the hypothesis of a transfer from one of these groups through a horizontal transfer event; comparative topology analysis of alternative tree hypotheses (expected phylogeny and transfer from an oomycete donor to a fungus or bacterium) using the Shimodaira-Hasegawa test provided support for this observation in 14 cases (Table 3). In four other cases, the topology test was significant for the opposite relationship, in which a transfer occurred from the oomycetes to fungi or bacteria. For HGT9, taxon sampling was not sufficient to accurately infer with confidence the putative HGT donor and enable tree topology testing (Table 3).

The distribution of sequence homologs of the strict set of HGT candidates among the oomycete phylogeny was strongly variable (Fig. 7; Supplementary Table S4). The majority of transfer events to oomycetes appear to have occurred relatively recently; three candidates had strong statistical support for trans-

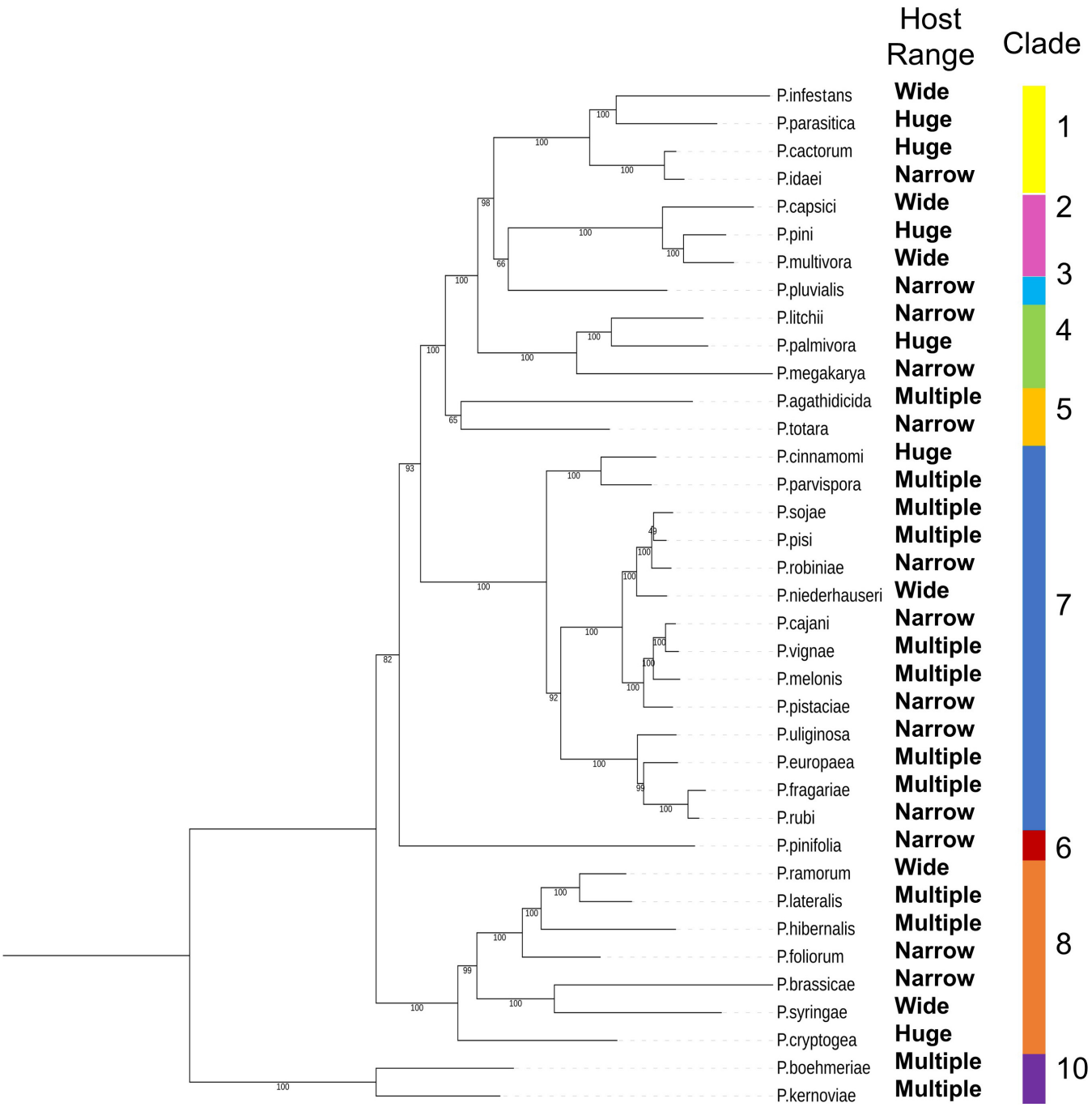

Fig. 5. Phylogenetic relationships of the 31 sequenced *Phytophthora* spp. and six additional previously sequenced *Phytophthora* spp. Sixty-one single-copy core orthologous proteins shared across 37 species were used to create a RAxML (Stamatakis 2014) phylogenetic tree, using each gene as an independent partition with its own substitution model and bootstrapped 1,000 times. Ranges of infected hosts are shown next to the phylogenetic tree species, defined as narrow (host species confined to one plant host genus), multiple (host species confined within two to nine host genera), wide (host species spanning 16 to 55 host genera), and huge (host species spanning 107 to 327 host genera). Clade assignments are shown on the right.

fer from bacteria or fungi to a common ancestor of the *Phytophthora* genus (HGT5, HGT7, and HGT10), two candidates to members of genus *Phytophthora* and order Peronosporales with hemibiotrophic lifestyle (HGT2 and HGT15), and two to members of order Peronosporales with a hemibiotrophic or an obligate biotrophic lifestyle (HGT12 and HGT20) (Fig. 7A). Four of these transfers reached close to gene fixation within the *Phytophthora* genus, as they were found in more than 80% of the species surveyed and in the nine *Phytophthora* phylogenetic clades considered (Fig. 7A). However, fixation was not the general rule accompanying recent transfers. For instance, two HGT candidates with functions related to plant pathogenicity (NPP1 protein and peptidase S9) were unique to *Phytophthora* clade 8 and did not have a homolog in any other *Phytophthora* clade or oomycete species (HGT5 and HGT7). Several HGT events with strong statistical support (Fig. 7A) appear to have occurred following major lifestyle transitions within the oomycetes, i.e., necrotrophy in order Pythiales (HGT6, HGT13 and HGT14) to obligate biotrophy and hemibiotrophy in order Peronosporales (HGT2, HGT12, HGT15, and HGT20) and transition to parasitism with three events trackable to a common ancestor of orders Saprolegniales, Pythiales, and Peronosporales (HGT1, HGT4 and HGT8). Eight of these ten genes had homologs (BLASTP E-values from 1e-137 to 1e-012) with pathogenicity, virulence, and effector genes of the Pathogen Host Interaction database (PHI-base) (Urban et al. 2019). Transition to the necrotrophic lifestyle involved transfers of genes encoding enzymes potentially involved in redox activity and toxin production (2,4-dienoyl-CoA reductase and phenol acid decarboxylase), while two of the four genes transfer at the transition to hemibiotrophic lifestyle comprehended have putative functions related to the degradation of the cell wall (xylulose reductase and endo-1,4-beta-xylanase).

Finally, the four significant transfers for the opposite relationship (oomycetes to bacteria or fungi) were all for genes fixed in the *Phytophthora* genus and mapped within order Peronosporales (three candidates) or orders Peronosporales and Pythiales (one candidate), suggesting relatively recent transfer events (Fig. 7B). Annotation of these genes indicates that they are potentially involved in the plant-pathogen interaction as they encode proteins involved in protection against plant defensive molecules (tannase and ABC transporter) and the oxidative stress occurring during the plant defense response (quinone oxidoreductase) and remodeling of the plant cell wall (pectinesterase) (Table 3; Supplementary Table S5).

Discussion

In this comparative genome study of 37 *Phytophthora* spp., we sequenced and assembled 31 genomes *de novo*. We investigated these genomes for evidence of HGT, phylogenetic relationships of genome structure and effectors, and association of host ranges. HGT has been identified as a significant source of variation in connection with the evolution of pathogenicity in *Phytophthora* spp. So far, genome-wide analyses of HGT impact on oomycete and *Phytophthora* genome evolution has identified putative transfers from fungi (Richards et al. 2006, 2011) and bacteria (McCarthy and Fitzpatrick 2016), many of which involve functions related to carbohydrate metabolism and

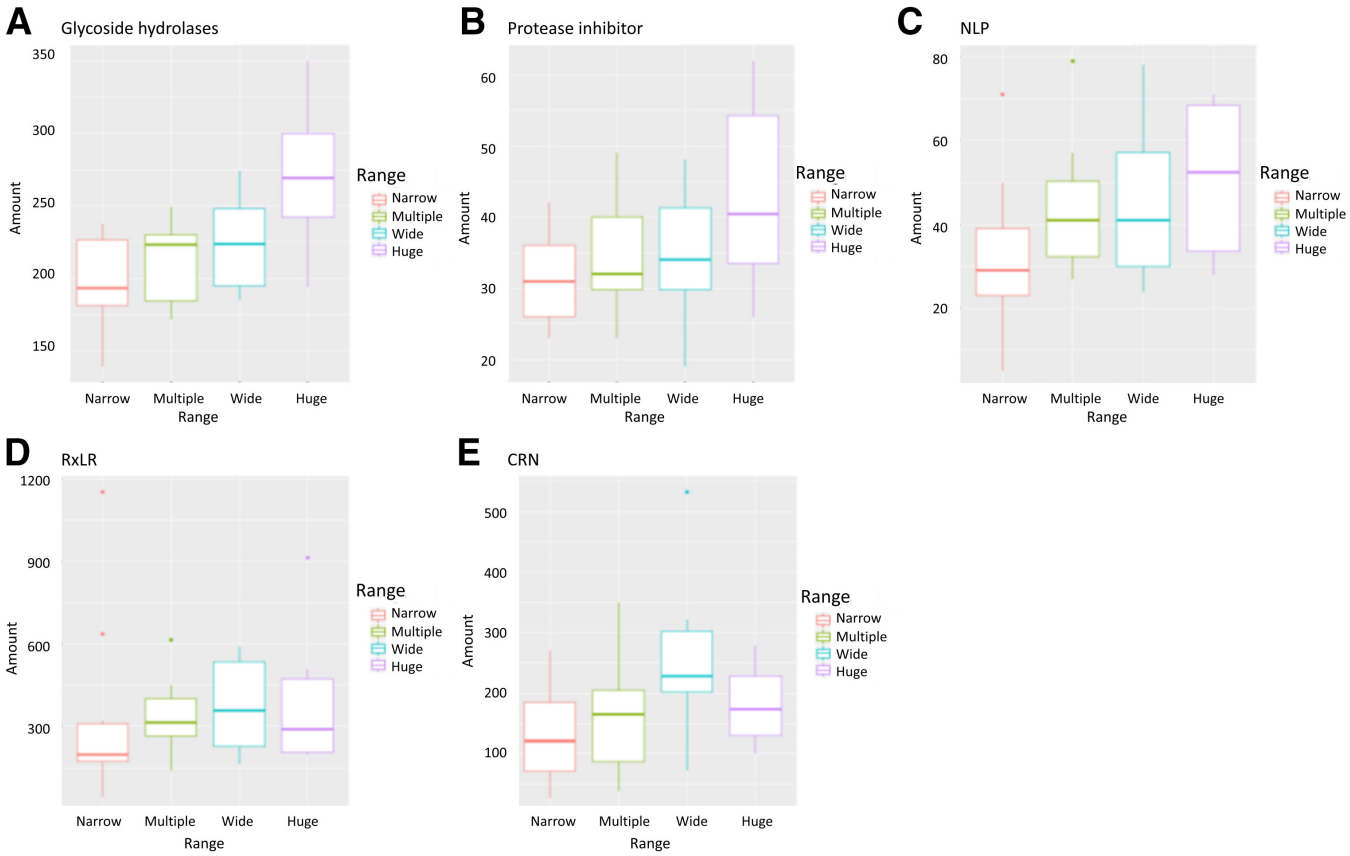

Fig. 6. Generally, *Phytophthora* spp. with a larger host range showed a greater predicted number of effector genes. Box plots showing numbers of effectors found per *Phytophthora* species, categorized into narrow (one plant genus), multiple (two to nine plant genera), wide (16 to 55 host genera), and huge (107 to 327 host genera) host ranges. **A,** Effectors glycoside hydrolases, **B,** protease inhibitors, **C,** necrosis-inducing proteins (NLPs), **D,** RxLRs, and **E,** crinklers (CRNs) are shown. Near identical paralogs were removed; proteins with greater than 95% amino acid identity over the full sequence length were reduced to one representative effector sequence. Statistically significant classifications were seen between the narrow-wide and narrow-huge comparisons in glycoside hydrolases (A).

pathogenicity. Our analysis supports these findings, identifying a set of 44 HGT candidates in *Phytophthora* species associated with enzymes putatively involved in the deconstruction of plant cell-wall components, evasion and protection against host defenses (Table 3; Supplementary Table S5). Our machine learning approach to identify HGT candidates aimed to identify genes likely inherited from bacteria or fungi; validation of these candidates with classical methods based on the identification of topological incongruence in phylogenies and the detection of discrepancies between gene and species distances resulted in a more conservative list of candidates than those previously proposed for *Phytophthora* species (McCarthy and Fitzpatrick 2016; Richards et al. 2006, 2011). Such a stringent approach had the power of rejecting the alternative evolutionary scenario in which the gene was present in the last common ancestor of the donor and recipient and was lost in intermediate lineages. Despite such a conservative approach, 30% of HGT candidates identified in a previous study (Richards et al. 2011) that included only three *Phytophthora* genomes were retrieved in our analysis.

An underlying hypothesis related to laterally transferred genes is that they may have functional or ecological roles, allowing the recipient to adapt to a novel lifestyle or to exploit a new ecological niche (Keeling and Palmer 2008). Using the comprehensive collection of *Phytophthora* genomes sequenced in this study and the oomycetes for which genome assemblies were available, we have been able to assess the extent of distribution of homologs of these candidates across the oomycete phylum. We confirmed that most of the HGTs into the oomycetes have occurred coincident with the emergence of major lifestyle innovations, such as the acquisition of plant parasitism or biotrophy (obligate or hemibiotrophy). Many candidates were detected in a large majority of the *Phytophthora* genomes sequenced, for example, seven candidates (HGT6, HGT11, HGT17, HGT18, HGT19, HGT21, and HGT22) were found in 94% or more of the 36 genomes surveyed. In some instances, homologs have been retained in distinct genetic lineages among the oomycetes, suggesting that these candidate genes may confer a function conserved across lineages with different lifestyles; for example, the endonuclease-encoding gene HGT1 was likely transferred before the radiation of the oomycetes and was retained in the five oomycete orders surveyed in this study. On the other hand, in some cases, the acquisition or retention of specific key pathogenicity genes appears to be restricted to some specific clades within genus *Phytophthora* (e.g., HGT5 and HGT7, NPP1) (Fig. 7), suggesting recent transfers following divergence of *Phytophthora* clades. In cases in which a putative HGT gene is present in a limited number of *Phytophthora* species from diverse clades, e.g., HGT13, gene loss by drift in species in which there was little benefit may be an explanation. Rapid diversification of the HGT gene under positive selection might make the gene undetectable to our algorithm in some species.

Oomycete-derived transfers to other kingdoms have been identified in a few rare instances, usually with limited statistical support (Richards et al. 2011). With the comprehensive genome sampling of our study, we found strong support for four transfers (of 44) from oomycetes to either bacteria or fungi, indicating bi-directional exchanges across kingdoms. By providing a source of novel genetic material that can increase the fitness of micro-organisms to their environments or their hosts (Feurtey and Stukenbrock 2018; Husnik and McCutcheon 2018; van Etten and Bhattacharya 2020), genes transferred horizontally have the potential to be traded back and forth across kingdoms. In the context of an ecological system in which a host plant interacts with a multitude of micro-organisms (microbiota), we can hypothe-

Table 2. Over-represented Gene Ontology (GO) terms for a set of candidate horizontal gene transfer (HGT) transcripts found in *Phytophthora* genomes

GO	Term	No. of terms in Full set[a]	No. of terms in HGT set[b]	Pr(*X*) = *k*[c]
Biological process				
GO:0055114	Obsolete oxidation-reduction process	1,543	5	<0.001
GO:0034079	Butanediol biosynthetic process	28	5	<0.001
GO:0045493	Xylan catabolic process	151	1	<0.01
GO:0008152	METABOLISM	3,663	3	<0.01
GO:0042545	Cell wall modification	493	1	<0.02
GO:0000272	Polysaccharide catabolism	635	1	<0.02
GO:0002084	Protein depalmitoylation	374	1	<0.02
GO:0006118	Electron transport	1,389	1	<0.05
GO:0045490	Pectin catabolic process	865	1	<0.05
Molecular function				
GO:0016491	Oxidoreductase activity	10,881	11	<0.001
GO:0000721	(R,R)-butanediol dehydrogenase activity	28	5	<0.001
GO:0003939	L-iditol 2-dehydrogenase activity	11	1	<0.001
GO:0016831	Carboxy-lyase activity	201	6	<0.001
GO:0008080	*N*-acetyltransferase activity	1082	3	<0.001
GO:0008270	Zinc ion binding	26,982	8	<0.001
GO:0005488	Binding	2,200	2	<0.01
GO:0031176	Endo-1,4-beta-xylanase activity	111	1	<0.01
GO:0008670	2,4-Dienoyl-CoA reductase (NADPH) activity	15	1	<0.01
GO:0004022	Alcohol dehydrogenase activity	102	1	<0.01
GO:0030599	Pectinesterase activity	498	1	<0.02
GO:0008474	Palmitoyl-(protein) hydrolase activity	457	1	<0.02
GO:0051213	Dioxygenase activity	460	1	<0.02
GO:0045330	Aspartyl esterase activity	423	1	<0.02
GO:0030570	Pectate lyase activity	657	1	<0.02
GO:0015267	Channel activity	742	1	<0.05
Cellular component				
GO:0005576	Extracellular region	4,624	2	<0.02

[a] Transcriptome of 37 *Phytophthora* genomes.
[b] Set = 44 HGT candidates.
[c] Probability (*q* value) of obtaining the same number of transcripts (*k*) or more by chance as given by a hypergeometric probability distribution.

size that some of the evolutionary innovations that are generated during the coevolutionary arms race between a pathogen and the host could be shared within the microbiota. For host-associated micro-organisms sharing the same ecological niche, the transfer of genetic material from those that are fit to the shared environment should represent a straightforward mechanism that will drive rapid adaptation of others to this environment (Soucy et al. 2015).

We investigated multiple aspects of *Phytophthora* genome structure and how this relates to the genus phylogeny. Genome size, gene amounts, and counts of orthologous genes varied within phylogenetic clades, highlighting the great diversity within the *Phytophthora* genus, and likely reflect the large observed differences in repeat content, some of which resulted from genome duplication. While some variation in repeat content may be due to differences in repeats collapsed in the genome assembly, repeat types and lengths identified in this study and sequences used for assembly are generally consistent across the sequenced genomes and should therefore collapse in assembly at similar rates. Interestingly, we noted that none of the genomes sequenced in this study approaches the 73% repeat content reported in the Sanger-assembled *P. infestans* genome (Haas et al. 2009) and is possibly due to the differences in sequencing technologies, including longer sequence lengths. Greater repeat content

Table 3. Summary of 19 horizontal gene transfer (HGT) candidates with strong phylogenetic support identified among *Phytophthora* spp.

Candidate	Putative function	Best hit on PHI-base (E-value)[a]	HGT identification[b]	No. of *Phytophthora*		Closest clade[c]	Most likely donor[d]	Most likely recipient[d]
				Species	Clades			
Transfer from other groups to oomycetes								
HGT_1	Endonuclease	PHI:5754, endonuclease, *Fusarium graminearum* (3.0e-024)	SVM, IP, D, T	29	8	Fungi	Fungi	Oomycetes
HGT_2	Xylulose reductase	PHI:2256, xylitol dehydrogenase, *Parastagonospora nodorum* (1.0e-128)	SVM, IP, D, T	12	5	Fungi	Fungi	Peronosporales (HB)
HGT_15	Zinc-binding dehydrogenase, polyketide synthase, enoylreductase domain	PHI:8321, gluconate 5-dehydrogenase, *Salmonella enterica* (1.0e-016)	SVM, IP, D, T	21	8	Bacteria	Bacteria	Peronosporales (HB)
HGT_4	Aquaporin	PHI:7047, water channel protein aquaporin, *Cryptococcus neoformans* (4.0e-012)	SVM, IP, D, T	30	8	Bacteria	Bacteria	Oomycetes
HGT_5	NPP1	–	SVM, IP, D, T	4	1	Bacteria	Bacteria	*Phytophthora* spp.
HGT_6	Phenol acid carboxylase	–	SVM, MC, D, T	35	9	Fungi	Fungi	Peronosporales/ Pythiales (plants)
HGT_7	Peptidase S9	–	SVM, IP, D, T	8	1	Fungi	Fungi	*Phytophthora* spp.
HGT_8	UDP-*N*-acetylglucosamine-peptide *N*-acetylglucosaminyl-transferase	PHI:4921, flagellin glycosyltransferase, *Burkholderia cenocepacia* (1.0e-021)	SVM, IP, D, T	2	2	Bacteria	Bacteria	Oomycetes
HGT_9	Alternative oxidase	–	SVM, IP, D	10	4	Fungi	Fungi (ns)	Oomycetes (ns)
HGT_10	Dioxygenase	–	SVM, MC, D, T	33	9	Fungi	Fungi	*Phytophthora*
HGT_11	Thioesterase	PHI:4988, sfp-type 4'-phosphopantetheinyl transferase, *Bipolaris maydis* (3.0e-005)	SVM, IP, D, T	35	8	Amoebozoa	Amoebozoa	Peronosporales/ Pythiales/ Lagediniales
HGT_12	Endo-1,4-beta-xylanase GH10	PHI:7912, endo-beta-1,4-xylanase *Phytophthora parasitica* (1.0e-137)	SVM, IP, D, T	33	9	Fungi	Fungi	Peronosporales
HGT_13	4-coumarate CoA ligase	PHI:10606, long-chain-fatty-acid–Co Aligase, *Pseudomonas aeruginosa* (1.0e-032)	SVM, IP, D, T	4	4	Fungi/ Bacteria	Bacteria/ Fungi	Peronosporales/ Pythiales
HGT_14	2,4-dienoyl-CoA reductase	PHI:8134, 3-Oxoacyl-[acyl-carrier-protein] reductase, *Salmonella enterica* (1.0e-014)	SVM, IP, D	28	7	Bacteria/ Archaea	Bacteria/ Archaea	Peronosporales/ Pythiales
HGT_20	Ribosomal-protein-alanine acetyltransferase	–	SVM, IP, D	29	9	Bacteria	Bacteria	Peronosporales
Transfer from oomycetes to other groups								
HGT_3	Quinone oxidoreductase	–	SVM, IP, D, T	23	9	Fungi	Peronosporales (HB)	Fungi (*Fusarium*)
HGT_16	Putative tannase	PHI:10222, feruloyl esterase, *Valsa mali* (4.0e-030)	SVM, IP, D, T	33	8	Bacteria	Peronosporales/ Pythiales	Bacteria
HGT_17	Putative pectinesterase CE8	PHI:278, pectin methylesterase, *Botrytis cinerea* (4.0e-077)	SVM, IP, D, T	34	9	Fungi	Peronosporales	Fungi
HGT_18	ATP-binding cassette	–	SVM, IP, D	34	9	Fungi	Peronosporales	Fungi

[a] Best BLASTp hit on the Pathogen Host Interaction database PHI-base (Urban et al. 2019).
[b] SVM = support vector machine, IP = incongruent phylogeny, MC = missing clades, D = distance, T = Alternate topology test (Shimodaira-Hasegawa test) significant.
[c] As reported in BLASTp analysis.
[d] ns = not significant according to Shimodaira-Hasegawa tests between the observed tree and alternative tree topologies; HB = only hemibiotrophic Peronosporales, i.e., genera *Phytophthora*, *Phytopythium*, and *Nothophytophthora*.

was not necessarily an indicator of large genome assembly size. While the five species with reported repeat content of greater than 30% (*P. pinifolia*, *P. megakarya*, *P. palmivora*, *P. fragariae*, and *P. hibernalis*) were within the nine largest assembly sizes, other species with large assembly sizes had less repeat content. For example, *P. niederhauserii* had an assembly size of 90 Mb but only 20.9% repeat content, which was considerably lower than expected when compared with even moderately repetitive *Phytophthora* genomes such as *P. sojae* (40%) (Tyler 2006). Genome size estimation using *k*-mer analysis also shows assemblies are shorter than expected (Supplementary Table S1). These observations suggest that repeat content may be underestimated in short-read genome assemblies and would expand with improved assembly and may also indicate missing repeat sequences from the *de novo* repeat identification process.

The numbers of predicted genes per species, genes per orthologous groups, and effector genes per species were consistent with those previously reported for *Phytophthora* species. Both the numbers of genes and the average sizes of genes were well within the ranges of the six previously sequenced *Phytophthora* species, for example, *P. sojae* was shown to have 26,000 genes with an average size of 1,181 bp (Tyler 2006). This supports our observation that the smaller assembly sizes of the *Phytophthora* genomes presented in this study were mainly associated with an overall reduction in repetitive regions, while the gene-containing sequences are relatively consistent in size. BUSCO analysis of the sequenced core ortholog content also showed similar results to previous *Phytophthora* studies. Some genome assemblies, including *P. palmivora*, which underwent whole-genome duplication, had lower single-copy gene numbers due to duplications. But overall, censuses of single-copy orthologs showed that both the genome assemblies and gene predictions were quite complete and comprised the majority of genes in each individual species sequenced. This suggests, that while genomes were small due

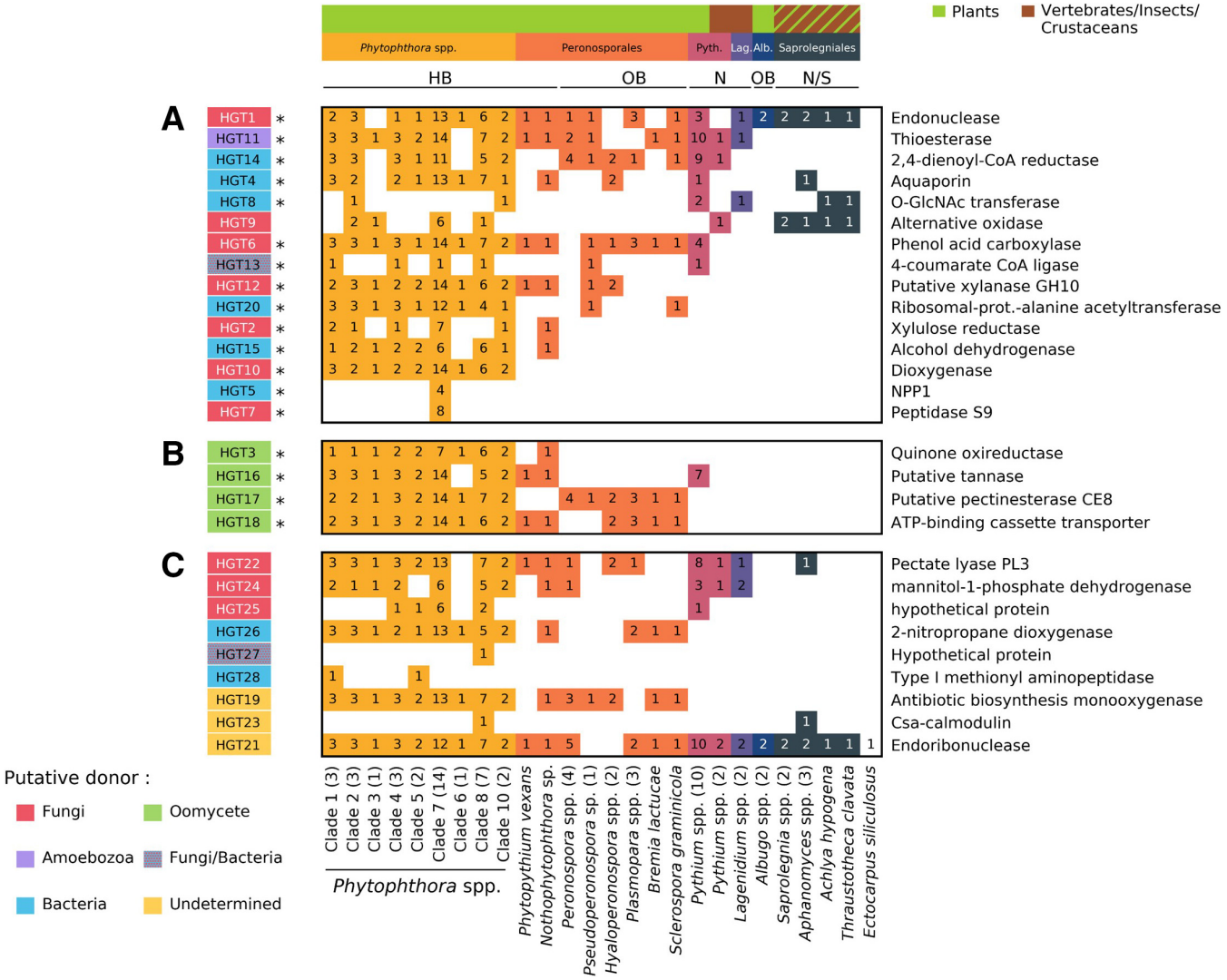

Fig. 7. Conservation level of the 44 *Phytophthora* HGT candidates in oomycetes. Set of 19 horizontal gene transfer (HGT) candidates for which a maximum likelihood phylogeny was reconstructed and alternate tree topologies were tested with the Shimodaira-Hasegawa test (asterisks indicate significant topological difference [$P < 0.05$] between the constrained alternate topology and the observed topology. **A,** HGT to oomycetes, **B,** HGT for the opposite relationships, i.e., transfers from oomycetes to fungi or bacteria, **C,** HGT candidates with no maximum likelihood phylogeny support. For each HGT candidate the number of sequence homologs identified among 37 *Phytophthora* and 30 oomycete transcriptomes (identified by reciprocal DIAMOND BLASTp, minimum E value of 1e-03, sequence subject coverage of 50% and sequence query coverage of 50%; a dash indicates the absence of a one-to-one ortholog) is reported. The filamentous brown alga *Ectocarpus siliculosus* (Ectocarpales, Ectocarpaceae) was used as an outgroup. Putative functions are indicated on the right. Top rows: Pyth. = Pythiales, Lag. = Lagenidiales, Alb. = Albuginales, HB = hemibiotrophic lifestyle, OB = obligate biotrophic, S = saprotrophic, N = necrotrophic. Species names, *Phytophthora* clade names and the group of species names are indicated on the bottom; numbers between brackets indicate the number of species considered in a group.

to collapsed repeat regions, the majority of core orthologs were captured in the assemblies and it can be extrapolated that the majority of the gene regions are assembled.

Effectors are proteins produced by pathogens that assist in host infection. Effectors are considered rapidly evolving genes that are usually conserved among a few closely related species and quickly diverge along the phylogeny. Two types of ortholog groups were identified in our *Phytophthora* genus analysis supporting this hypothesis; one group corresponds to well-conserved genes among all *Phytophthora* species responsible for core cellular functions, while a second group includes rapidly evolving gene families likely responsible for host infection and adaptation. We also investigated how the amounts of predicted apoplastic and cytoplasmic effectors related to the host range of each *Phytophthora* species. There were large differences in the numbers of effector genes identified per species. We did not observe a correlation between the number of predicted effectors with phylogenetic relationships or genome size. However, a clear relationship with host range was observed. Species with smaller host ranges had, on average, fewer predicted effectors than those with larger host ranges. When the *Phytophthora* species in this study were separated into four host-range categories, the distribution of apoplastic effectors increased as the range of infected hosts increased, from narrow to multiple, multiple to wide, and wide to huge. Cytoplasmic effectors showed a similar pattern; however, both the RxLR and CRN effector numbers dropped from wide to huge host ranges. Our study does not include a detailed measurement of gene expression levels of these effector genes during infection of numerous hosts, so there are limitations in how these correlations can be interpreted. In the absence of that information, we speculate that a large diversity of apoplastic effectors may be important for successfully overcoming the apoplastic defenses of a large diversity of host plants.

There may be a similar requirement for larger numbers of cytoplasmic effectors, but expression of very large numbers of cytoplasmic effectors may limit host range due to plant immune surveillance mechanisms. Detection of a single cytoplasmic effector by an NLR resistance protein may be sufficient to prevent infection, therefore a limited number of cytoplasmic effectors may result in a greatly expanded host range. Our observation of reduced cytoplasmic effector complements in huge host range species may also be indicative of cryptic host-specialization within these *Phytophthora* species. Recent work in *P. cactorum*, commonly considered a broad host–range pathogen, has shown genomic signatures of host specificity (Nellist et al. 2021). In this case, high-resolution phylogenetics demonstrated that host adaptation was associated with effector gene gain or loss between strawberry and apple infecting clades. Where such cryptic host-adaptation is present, pangenomic analysis may be a useful tool to infer broad or narrow host range and provide insight to associations of effector diversity across *Phytophthora* spp.

Materials and Methods

Collection and isolation of a genus-wide *Phytophthora* collection.

Mycelium samples were isolated for all *Phytophthora* species in this study, as well as germinated cyst samples for RNA sequencing of a subset of the species. For mycelium tissue, plugs of mycelium grown on standard V8 agar plates were added to a flask with 20% liquid V8 media clarified with calcium carbonate and were incubated with shaking at 45 rpm at 25°C for 1 week. Agar plugs were removed from the mycelium mass and the tissue was ground to a powder with liquid nitrogen, followed by DNA extraction using the methods described by Möller et al. (1992), except using 1% CTAB and phenol/chloroform treatment or total RNA extraction (Johansen and Carrington 2001) using TRIzol (Invitrogen), following manufacturer instructions.

For *Phytophthora robiniae* and *Phytophthora vignae*, similar to the protocol for *P. sojae* (Tyler 2006), zoospores were produced by repeated washing of 11-day-old V8-200 plates of mycelium with sterile double distilled water, followed by overnight incubation at 14°C. Germinated cysts were produced by exposing collected zoospores to cleared V8 broth for 1 h. For *P. parvispora*, mycelium mats were grown in liquid V8 for 5 days, then, the liquid V8 was changed out for soil extract (soil collected with stream water, mixed, and filter-sterilized), and zoospores were collected after 3 days, followed by germinated cyst induction as above. For *P. cajani*, *Phytophthora europaea*, *Phytophthora foliorum*, *P. hibernalis*, *P. pistaciae*, and *Phytophthora uliginosa* species that did not readily yield zoospores, mycelium was grown in Plich medium (Kamoun 1993) for RNA sequencing to compare against V8 medium growth. Known intergenic transcribed spacer and CoxII sequences for each species were used to confirm species identification before high-throughput sequencing. DNA and RNA quality were checked with electrophoresis (DNA), Bioanalyzer (RNA), and NanoDrop.

Isolate P414 of the strawberry crown rot pathogen *P. cactorum* and isolate SCRP371 of the raspberry root rot pathogen *P. idaei* were sequenced at the National Institute of Agricultural Botany at East Malling Research (NIAB EMR). P414 and SCRP371 were isolated from symptomatic strawberry and raspberry plants, respectively. DNA extraction was performed on freeze-dried mycelium, using a GenElute plant genomic DNA miniprep kit (Sigma), following the manufacturer protocol with the following modifications. The RNase A digestion step was not performed and samples were eluted using 2 × 100 µl elution buffer for P414 and 2 × 75 µl for SCRP371. Genomic libraries were prepared using a Nextera XT library preparation kit (Illumina) or TruSeq DNA LT kit (Illumina) for *P. cactorum* and *P. idaei*, respectively.

P. kernoviae, *P. lateralis*, *Phytophthora cryptogea*, and *P. pinifolia* were collected and isolated as described (Feau et al. 2016). *P. agathidicida*, *Phytophthora multivora*, *P. pluvialis*, and *P. taxon totara* were collected and isolated as described (Studholme et al. 2016).

Genome sequencing and assembly.

Thirty-one *Phytophthora* species were sequenced by our consortium. Genomes of 21 *Phytophthora* species were sequenced by BGI Genomics (Shenzhen, China) (*P. boehmeriae*, *P. brassicae*, *P. cajani*, *P. pini*, *P. europaea*, *P. foliorum*, *P. fragariae*, *P. hibernalis*, *P. litchii*, *P. megakarya*, *Phytophthora melonis*, *P. niederhauserii*, *P. palmivora*, *P. parvispora*, *Phytophthora pisi*, *P. pistaciae*, *P. robiniae*, *P. rubi*, *Phytophthora syringae*, *P. uliginosa*, and *P. vignae*), using 90-bp paired-end reads produced on the Illumina HiSeq2000 platform. Four *Phytophthora* species were sequenced by the University of British Columbia (*P. kernoviae*, *P. lateralis*, *P. pinifolia*, and *Phytophthora crypogea*), using Illumina HiSeq 2000 100-bp paired-end reads (Feau et al. 2016). Genomes from four *Phytophthora* species isolated from New Zealand (*P. agathidicida*, *P. multivora*, *P. pluvialis*, and *P. taxon totara*) were sequenced by Scion (New Zealand Forest Research Institute, Ltd.), using primarily Illumina HiSeq 100-bp paired-end reads (Studholme et al. 2016). Two *Phytophthora* genomes were sequenced by NIAB EMR (*P. idaei* and *P. cactorum*), using 250-bp paired-end reads produced on a MiSeq Benchtop Analyser (Illumina). BGI-sequenced genomes were adapter trimmed to remove Illumina adapters and quality trimmed to remove Phred scores of less than Q20 from the ends of reads (Martin 2011). Genome sequences were assembled with SOAPdenovo2 (Luo et al. 2012). Several initial

assemblies were done to identify an optimal *k*-mer length for each genome. Gap filling and single base proofreading were conducted with SOAPAligner (Gu et al. 2013). The University of British Columbia (UBC) genomes were quality trimmed and assembled using ABySS (Simpson et al. 2009) and a range of *k* values from 32 to 96 (Feau et al. 2016). Scion genomes were assembled using SPAdes (Bankevich et al. 2012) and contigs were extended using SSPACE (Boetzer et al. 2011; Studholme et al. 2016). Genomes sequenced at NIAB EMR were trimmed and adapters were removed using fastq-mcf (Aronesty 2013), prior to *de novo* assembly of the data, using Velvet (Zerbino and Birney 2008), at *k*-mer lengths of 61 and 41 bp for *P. cactorum* and *P. idaei*, respectively.

Genome size was estimated using *k*-mer counts of the raw Illumina sequence; *k*-mers were counted using Jellyfish count (version 2.2.6, -m 32) (Marçais and Kingsford 2011) Histograms created using Jellyfish hist were plotted using R (R Core Team 2019) to identify the apex and boundaries of the single copy *k*-mer peak. Genome size was calculated by dividing the total of unique *k*-mers by the mean coverage (peak *k*-mer frequency).

To separate mitochondrial genome contigs from the nuclear genome assembly, full-length mitochondrial genome sequences were collected from GenBank (Clark et al. 2016) for the following nine *Phytophthora* species: *Phytophthora andina*, *P. infestans*, *Phytophthora ipomoeae*, *Phytophthora mirabilis*, *P. parasitica*, *Phytophthora phaseoli*, *Phytophthora polonica*, *P. ramorum*, and *P. sojae*. The 31 consortium-assembled *Phytophthora* genomes were aligned with Blat (Kent 2002) to identify mitochondrial contigs. Blat alignments were filtered to return alignments greater than 50% of the aligned contig length, greater than 100 bp, and with gaps less than 50% of the contig length. Contigs identified as mitochondrial were removed from the genome assembly and are a part of a different study.

Transcriptome sequencing and assembly.

Twenty-four of the genome-sequenced *Phytophthora* species underwent RNA sequencing (*P. brassicae*, *P. cactorum*, *P. cajani*, *P. pini*, *P. europaea*, *P. foliorum*, *P. fragariae*, *P. hibernalis*, *P. kernoviae*, *P. lateralis*, *P. litchii*, *P. megakarya*, *P. melonis*, *P. niederhauserii*, *P. palmivora*, *P. parvispora*, *P. pinifolia*, *P. pisi*, *P. pistaciae*, *P. robiniae*, *P. rubi*, *P. syringae*, *P. uliginosa*, and *P. vignae*). Two samples, V8-grown mycelia and either Plich-grown mycelia or germinated cysts, were sequenced for each. Twenty-one species were sequenced by BGI Genomics, using custom library construction protocol. Random hexamer-primers were used to synthesize the first-strand cDNA; second-strand cDNA was synthesized using buffer, dNTPs, RNase H, and DNA polymerase I; short fragments were purified with QiaQuick PCR extraction kit, were resolved with EB buffer, and were connected with sequencing adaptors. Each *Phytophthora* transcriptome received 90-bp paired-end reads. Two of the above species were sequenced by UBC (*P. kernoviae* and *P. lateralis*). RNA of *P. cactorum* was sequenced by NIAB EMR.

To create a transcriptome reference for gene predictions, Trinity assemblies (Grabherr et al. 2011) were made for each *Phytophthora* species, using both RNA sequence samples (–seqType fq –min_contig_length 200). The *de novo* transcriptome assemblies were cleaned using the three-step TransDecoder process (Haas et al. 2013). A set of predicted protein sequences was made by combining *Phytophthora* protein sequences from GenBank (Clark et al. 2016) with protein sequences from the six previously sequenced and annotated *Phytophthora* species (*P. sojae* [Tyler 2006], *P. ramorum* [Tyler 2006], *P. infestans* [Haas et al. 2009], *P. capsici* [Lamour et al. 2012], *P. cinnanomi* [JGI PhycoCosm], and *P. parasitica* [NCBI database]). Transdecoder.LongOrfs (Haas et al. 2013) was used to identify the longest open reading frames (ORFs) in the Trinity assembly. BLASTP (Altschul et al. 1990) was used to align the longest ORFs to the set of constructed proteins identified from GenBank (parameters: -max_target_seqs 1 -evalue 1e-5). Finally, Transdecoder.Predict was used to predict the gene structure from the transcriptome assembly. The resulting cleaned transcriptome assemblies were used in the subsequent gene prediction methods.

***De novo* repeat identification.**

Each genome was repeat-masked to create a genome assembly ready for gene prediction, as described below. Repeat elements were *de novo* identified separately by species. *De novo* predictions were combined along with previously identified *Phytophthora* repeats for species-specific repeat identification.

To identify *de novo* discovered LTR retrotransposons, LTRharvest (Ellinghaus et al. 2008) and LTR_(Xu and Wang 2007) finder were run on each genome assembly. By species, LTR retrotransposon predictions from both LTRharvest and LTRfinder were condensed by coordinates and reduced by Blat alignments. LTR retrotransposons, non-LTR retrotransposons, DNA transposons, and other repeat elements were identified following the MAKER 'Repeat Library Construction-Advanced' (Cantarel et al. 2007) method (available online). This process utilizes the following programs: MITE-Hunter (with default parameters) (Han and Wessler 2010); GenomeTools suffixorator, LTRharvest, LTRdigest (run with 99% and 85% identity) (Gremme et al. 2013); RepeatModeler (Smit and Hubley 2015a), which calls RECON (Bao 2002), RepeatScout (Price et al. 2005), TRF (Benson 1999), NSEG (Wootton and Federhen 1996) and RMBlast; and sequence databases provided by the MAKER 'Repeat Library Construction-Advanced' method.

De novo repeat identification was further supplemented by LTR_retriever (Ou and Jiang 2018), using results from LTRharvest and LTRfinder. LTR_retriever retrotransposons were classified into subfamilies by species.

For each assembled genome, predicted repetitive elements identified in the above methods were combined with GIRI RepBase (volume 18, issue 9) (Bao et al. 2015) *Phytophthora* repeats. Genomes were repeat-masked using RepeatMasker (v 4.0.6, run with the described combined custom *Phytophthora* library and default parameters) (Smit et al. 2015b) and the combined repeat database to create gene prediction-ready genome assemblies.

Gene prediction and annotation in 31 *Phytophthora* species.

For each species sequenced, gene training models were made with both AUGUSTUS (Stanke and Morgenstern 2005) and SNAP (Korf 2004). AUGUSTUS was trained using the genome assembly and the set of previously sequenced *Phytophthora* proteins described in the transcriptome sequencing and assembly process in the text above. To train SNAP, for each species, BUSCO (Seppey et al. 2019) was run, using the Alveolata_Stramenopiles database in genome mode on each genome assembly to identify core orthologs. BUSCO gff files were converted to zff using maker2zff (Cantarel et al. 2007), SNAP tools fathom (-categorize 1000, -export 1000), forge, and hmm-assembler were used to create a training HMM. Genes identified as single copy core orthologs were combined and were used as the SNAP training set.

Several supplementary files were created to run the MAKER gene prediction pipeline. For the 'EST Evidence' section of MAKER, the transcriptome result of the three-step TransDecoder process (Haas et al. 2013) was used as the expressed sequence tag (EST) field. In the seven cases where RNA was not sequenced and, therefore, the TransDecoder transcriptome was not created, the phylogenetically closest species with RNA sequences was used. A concatenation of the gene sequences

of previously sequenced *Phytophthora* species and TransDecoder transcriptome assemblies was used for the alt-est field. For each genome assembly, *P. sojae*, *P. infestans*, and *P. ramorum* gene predictions were combined with five representative TransDecoder-cleaned transcriptome assemblies. To create species-specific alt-est sets, the five representative species were selected from within the *Phytophthora* phylogenetic clade (excluding the species of interest). If fewer than five species received RNA sequencing, nearby clades were used until five transcriptome assemblies were combined. This set of eight gene predictions and transcriptome assemblies was used in the alt-est field. For the 'Protein Homology' section of MAKER, all previously identified *Phytophthora* proteins were combined from GenBank together with the six previously sequenced *Phytophthora* species.

To predict genes, MAKER was run on the 31 genomes. The repeat-masked genome for each assembly was run using the AUGUSTUS and SNAP training models described above and the EST and protein evidence sequence sets as described above.

For validation purposes, three sets of BUSCO analyses were run. BUSCO 3.02 was run on the 31 genomic assemblies, on the 24 transcriptome assemblies, and on the 31 sets of predicted proteins. In all cases, the Alveolata_Stramenopiles BUSCO database of 234 single-copy orthologs was used.

Predicted genes were functionally annotated using BLASTX (Altschul et al. 1990) to align against known proteins. First, all predicted proteins were aligned against all *Phytophthora* species proteins obtained from the Reference Sequence collection (RefSeq) nonredundant protein database in NCBI (Clark et al. 2016; O'Leary et al. 2016) and the UniProt TrEMBL (UniProt Consortium 2019) database. Second, all predicted proteins were aligned against all Stramenopile proteins from NCBI RefSeq and UniProt. Third, all predicted proteins were aligned against all Fungi proteins from NCBI RefSeq and UniProt. BLASTX alignments were generated with the following parameter settings: -evalue 1e-5, -max_target-seqs 50 and were further filtered to return only hits that were at least 50% identical for 50% of the length of the subject protein. Protein functional annotation was made from the consensus of the top five protein alignments for each taxonomic classification. All results per query protein were screened to return proteins with informative functional annotations ranked by the two rounds of alignments.

Predicted proteins from MAKER were screened using InterProScan version 5.20-59.0 (Jones et al. 2014) to identify functional domains. GO (Ashburner et al. 2000; Gene Ontology Consortium 2019) terms were obtained from UniProt BLASTX alignments and the InterProScan runs.

Identification of effector proteins.

The prediction of cytoplasmic effectors of the RxLR and CRN families was performed on the six-frame translations of the *Phytophthora* whole-genome assemblies, using the application getorf (EMBOSS suite) (Rice et al. 2000). We searched for evidence of the presence of the motifs of interest (RxLR+EER motif for RxLR effectors [Win et al. 2007] and LxLAK for CRN effectors [Haas et al. 2009; Stam et al. 2013]) in each ORF translation by using a combination of regular expressions, using effectR (Tabima and Grünwald 2019).

To identify additional effector proteins that may not include one the canonical motifs, thus, may not be recognized by the RxLR or CRN regular expression, we searched against a profile hidden Markov model (HMM) (Eddy 1998). We built the HMM profile using an intersect of each set of candidate effectors predicted, using regular expressions for each sequenced species in the consortium and the previously predicted effectors from the reference genomes of *P. infestans*, *P. ramorum*, and *P. sojae* (Haas et al. 2009; Tyler 2006). We searched for additional effectors in all ORF translations against each HMM profile, using the hmmsearch program in HMMER (available online), using default threshold parameters. Predicted effectors from the motif method and the HMM method were examined for signal peptides, using SignalP 3.0 (Bendtsen et al. 2004).

Putative apoplastic protease inhibitors were annotated by batch BLASTP (E-value < 1e-30) against the MEROPS database (Rawlings et al. 2012). The glycoside hydrolase proteins were annotated using the carbohydrate-active enzyme database (CAZy) annotation web server dbCAN (Yin et al. 2012). The HMM profile of the NLP family (PF05630) was downloaded from Pfam database (El-Gebali et al. 2019). The hmmsearch program (with default threshold parameters) was used to search for NLP proteins in each genome assembly.

Classification into orthologous groups.

Predicted proteins from the 31 sequenced *Phytophthora* and from the six previously sequenced *Phytophthora* species were combined into orthologous groups, using OrthoMCL (Li 2003). Due to the large dataset of 37 full genomes, a two-stage process was used. In stage one, proteins were assigned to the online pre-constructed OrthoMCL orthology groups (Chen 2006). Predicted proteins from the MAKER process were uploaded to the OrthoMCL web site. This returned a file of proteins assigned to OrthoMCL groups. In stage two, all unassigned proteins were assigned to groups using the stand-alone version of OrthoMCL. All unassigned proteins from all 37 species were combined into a single FASTA sequence file. The protein FASTA file was aligned against itself using BLASTP. The BLASTP output was converted for input into OrthoMCL, which was run in mode 4.

Phylogenetic analysis of single-copy orthologs across *Phytophthora* species.

To estimate phylogenetic relationships across the 31 genomes and the six previously sequenced *Phytophthora* species, we first identified the single-copy, core orthologous genes shared across all sequenced species. We selected each of the orthology groups that contain exactly one gene from each of the 37 genomes in the orthology construction.

We constructed a phylogenetic tree using 61 genes from each species. Each set of orthologous proteins were multiply aligned using MAFFT ver. 7.271 (Nakamura et al. 2018; Katoh and Standley 2013). The phylogenetic tree was reconstructed using RAxML (Stamatakis 2014), using each gene as an independent partition with its own substitution model, bootstrapped 1,000 times. Only one tree was calculated, using all partitions.

Effector distribution across plant host ranges.

The United States Department of Agriculture fungal database (available online) was used to define the number of known plant hosts infected by each *Phytophthora* species considered in this study. With this information, host ranges were classified into four categories defined by the number of host genera containing known hosts. The narrow category encompasses host species confined to one plant host genus (one to three host species total), multiple encompasses host species confined within two to nine host genera (two to 32 host species total), wide encompasses host species spanning 16 to 55 host genera (22 to 119 host species total), and the category designated huge encompasses host species spanning 107 to 327 host genera (163 to 718 host species total).

Numbers of predicted cytoplasmic and apoplastic effector genes were plotted for each host range. To reduce errors caused by genome assembly artifacts, and to limit counts of functionally identical effector genes, near-identical paralogs were removed from the counts. To identify near-identical paralogs, effector amino acid sequences were aligned to one another, using the Smith-Waterman local aligner from the EMBOSS package

(Rice et al. 2000) to identify similarity. Effectors with greater than 95% amino acid identity over the full sequence length were reduced to one representative effector sequence. The resulting reduced set of effector predictions was used for analysis of the relationship of effector repertoires to host range. Host ranges were plotted for each effector type using ggplot2 (Wickham 2016) in R (R Core Team 2019).

HGT.

We used a two-step process to identify HGT gene candidates in *Phytophthora* genomes. In the first step, we used an SVM classifier to predict HGT candidates. In the second step, we applied two filters to screen out false-positive candidates and assess the likelihood that the candidates were acquired through HGT.

SVM classifier for prediction of HGT candidates.

We hypothesized that DNA sequence-composition features such as G+C content, codon bias, and codon usage frequency (Sharp and Matassi 1994) can be used to identify genes of recent bacterial or fungal origin in *Phytophthora* genomes. We constructed a multiclass SVM (Boser et al. 1992) for composition-based analysis of *Phytophthora* protein-coding genes and classification as either *Phytophthora*, bacterial, or fungal origin. SVM is well-suited for sequence-composition classification because of the availability of SVM libraries that perform well with large datasets with numerous variables and the ability of SVM to minimize unimportant features (Pedregosa et al. 2011). The SVM algorithm was implemented in a custom Python script using the SVC function, available from Scikit-learn Python library (Pedregosa et al. 2011).

Training sets consisted of 15,000 each of ascomycete, *Phytophthora* and bacterial transcripts, for a total of 45,000 transcripts. Ascomycete transcripts were selected by submitting a collection of complete transcript sets predicted from the genomes of representative species of eight main ascomycete classes, namely, *Tuber melanosporum* (Pezizomycetes, GCA_000151645) (Martin et al. 2010), *Arthrobotrys oligosporus* (Orbiliomycetes, GCA_000225545) (Yang et al. 2011), *Penicillium chrysogenum* (Eurotiomycetes, GCA_000226395) (van den Berg et al. 2008), *Leptosphaeria maculans* (Dothideomycetes, GCA_000230375) (Rouxel et al. 2011), *Cladonia grayi* (Lecanoromycetes) (Yang et al. 2011), *Sclerotinia sclerotiorum* (Leotiomycetes, GCA_000146945) (Amselem et al. 2011), *Fusarium graminearum* (Sordariomycetes, GCA_000240135.3) (Cuomo et al. 2007), and *Xylona heveae* (Xylonomycetes, GCA_001619985) (Gazis et al. 2016). Potential genes that underwent HGT were discarded from each transcript set by applying the following protocol: i) transcripts were translated into proteins and clustered using OrthoMCL (coverage and identity of at least 50%, E-value cut-off of 1e-05, inflation parameter = 2.5) (Li 2003); ii) one protein from each cluster was then queried against the NCBI Nonredundant (nr) database (max target sequences = 500); iii) clusters with at least one hit in any other organisms than a fungal taxon were discarded; iv) for each remaining cluster, each protein was queried against the nr database (max target sequences = 500) and the previous step was re-applied. *Phytophthora* genes were selected by the same process, using transcripts from *P. syringae* (this study), *P. sojae* (GCA_000149755) (Tyler 2006), *P. ramorum* (GCA_000149735) (Tyler 2006), *P. lateralis* (GCA_000500205) (Feau et al. 2016), *P. pinifolia* (GCA_000500225) (Feau et al. 2016), *P. cryptogea* (GCA_000468175) (Feau et al. 2016), *P. infestans* (GCA_000142945) (Haas et al. 2009), *P. brassicae* (this study), and *P. kernoviae* (GCA_000448265) (Feau et al. 2016) and eliminating clusters with any protein match other than with an oomycete taxon. Bacterial transcripts were selected following the same filtering approach on 21,096 transcripts retrieved from GenBank (representing 23 bacterial classes).

Sequence-composition features were used as input vectors to an SVM classifier and the curated training sets (discussed above) were used as model data. Following a preliminary analysis, codon usage frequency and GC content were selected as the sequence features, as they resulted in a higher prediction accuracy than codon bias (0.976 ± 0.002 vs. 0.973 ± 0.004, $t = 10.4$, $P < 0.0001$) (data not shown). This is consistent with the point that codon use frequency is inherently the fusion of both codon usage bias and amino acid composition signals (Wu 2007). To choose the best kernel for the SVM, we first used principal component analysis to explore the relationships among the three different classes (Supplementary Fig. S4). Radial basis function (rbf) kernel parameters (*C* and *gamma*) were systematically varied to optimize prediction accuracy, using a two-dimensional grid on which both parameters were chosen from the set $\{10^{-3}, 10^{-2}, \ldots, 10^{6}\}$. All these optimizations were performed with fivefold cross-validation of the training set (randomly withholding one-fifth of the training data as a testing dataset; 100 random draws for each pair of parameters tested) (Supplementary Fig. S4). Accuracy as defined by $(TP + TN)/(TP + TN + FP + FN)$ was used as a measure of the quality of the classification. Best classification accuracy (98.3%) was obtained with rbf kernel parameters of $C = 1{,}000.0$ and $gamma = 1.0$ (Supplementary Fig. S4).

Phytophthora transcript classification for HGT.

The 618,240 transcripts predicted from the 31 genomes and 103,992 transcripts predicted from five previously sequenced *Phytophthora* species (i.e., *P. sojae*, *P. ramorum*, *P. infestans*, *P. capsica*, and *P. cinnamomi* [discussed above]) were submitted to the classifier and were sorted into *Phytophthora*-origin, bacterial-origin, or fungal-origin classes, depending on the probability returned by the classifier for each of these classes. To generate a confidence score, we repeated the training of the classifier 100 times before running the classification on each genome. To maximize the training process of the classifier without increasing computing time and overloading memory, we used a random subsample of 45,000 transcripts (15,000 genes in each of the three classes) as a training set each time. Preliminarily, we determined the minimum threshold number of bootstrap replicates in which an HGT candidate was found that would minimize the probability that this candidate was a false positive (e.g., misclassifying a *Phytophthora* or a bacteria sequence as deriving from a fungal donor via HGT). This was done by submitting a subsample of 1,500 sequences randomly picked in the training set (500 transcripts in each class) to the classifier with the bootstrap procedure; then, false-positive and true-positive rates were calculated for incremental values of bootstrap replicates. Based on this test, we determined that the chance of misclassifying a fungal transcript as a *Phytophthora* or a bacterial sequence (i.e., a false positive) was <0.2% (1/500) if it was classified as fungal in at least 89/100 bootstrap replicates; in such case, the true positive rate [recall; $TP/(TP + FN)$] would be 92.3% (Supplementary Fig. S5A). For the bacterial sequences, this value was ≥79/100 bootstrap replicates; this corresponded to a true-positive rate (recall) of 97.5% (Supplementary Fig. S5B). These two bootstrap replicate thresholds were then used for the identification of HGT candidates in *Phytophthora* species.

HGT candidate false-positive filtering.

HGT candidates predicted with the SVM classifier were submitted to a phylogenetic filtering step by assessing the congruence of the gene phylogeny with the organism phylogeny. Each candidate transcript was translated into a protein sequence and searched using DIAMOND BLASTP (minimum

BLASTP E-value of 1e-03, sequence subject coverage of 50%, and sequence query coverage of 50%) (Buchfink et al. 2015) for closest homologs against protein-coding sequences downloaded from the NCBI RefSeq collection (O'Leary et al. 2016) for *Phytophthora* species (72,639 sequences) and the following clades: heterokonts (excluding *Phytophthora* spp., 164,619 sequences), Alveolata (1,527,928 sequences), Amoebozoa (113,408 sequences), Viridiplantae (5,556,940 sequences), Fungi (2,912,973 sequences), Archaea (1,830,006 sequences), and Bacteria (131,971,793 sequences). Candidates with no hits in the Bacteria or Fungi clades were directly rejected. Protein sequences for the top three DIAMOND BLASTP hits within each of the above clades were retrieved and aligned with the query protein, using MAFFT ver. 7.271 (Katoh and Standley 2013). Amino acid sites with a gap in more than one third of the sequences were removed. IQ-TREE was used to determine the best-fitting substitution model and reconstruct a maximum-likelihood tree for each of the alignments (Kalyaanamoorthy et al. 2017; Nguyen et al. 2015) with node support assessed by using the ultrafast bootstrap approximation method (Hoang et al. 2018). Each phylogenetic tree was exported in Newick format, was automatically rooted with sequences from the Archaea or Bacteria clades, and was exported into a .png file, using the BioPython package Phylo. For facilitating visual examination of trees, png files were gathered into one single pdf catalog with the Python library PyFPDF. Tree nodes were visually inspected to identify phylogenetic discordance (Stöver and Müller 2010). Two types of discordance were examined: i) "complete incongruence", when the HGT candidate sequence clusters with the fungi or bacteria clade with bootstrap support $\geq$50%, resulting in a phylogenetic tree completely discordant with the expected organism phylogeny, i.e., ((((((Phytophthora, Heterokonta), Alveolata), Viridiplantae), (Fungi, Amoebozoa)), Archaea), Bacteria) (Burki et al. 2020; Keeling and Burki 2019) and ii) "missing clades", when the HGT candidate sequence clusters with the fungi or bacteria clade because other clades are missing (i.e., the HGT candidate sequence did not have orthologs in intermediate clades such as Viriplantae, Amoebozoa, and Alveolata).

HGT candidates that passed phylogenetic filtering were submitted to a sequence identity discrepancy filter. Assuming a molecular clock, the sequence identity between a pair of orthologous genes should be in the same range as the average sequence identity between their respective species. However, for a pair of sequences related through an HGT event between two species, the proportionality should be broken, leading to an identity discrepancy when compared with the pairwise species identity (Novichkov et al. 2004). To identify such discrepancies, we performed "identity tests." We calculated the nucleotide sequence identity between the candidate HGT sequence in *Phytophthora* spp. and its closest homolog sequences in the putative donor species in Bacteria or Fungi. Then, the full transcriptome of the putative donor species was downloaded and searched with BLASTN for 1,000 random transcripts from the *Phytophthora* species to identify one-to-one orthologs and plot a distribution of expected nucleotide identity values. In a first identity test, discrepancies were identified by comparing the observed nucleotide sequence identity found between the HGT candidate in *Phytophthora* spp. and its homolog sequence in the putative donor to the expected distribution using a Wilcoxon sign-rank test. Candidates were rejected if the difference between the observed value (sequence identity between the *Phytophthora* HGT candidate and its homolog sequence in the putative donor) and the average of the expected distribution was not significant and lower than an arbitrary "discrepancy cutoff" of 8.56 (corresponding to the top quartile of the distribution of the difference between the observed values and the expected values). Proteins retained at this step were included into a relaxed list of candidates. To ascertain that the discrepancy was not caused by a high conservation of the gene among the different clades, we validated this list with a second identity test that consisted of examining if the nucleotide sequence identity between the HGT candidate and the closest species in the non-donor clades was not significantly higher than the average identity expected between the two species. Only candidates for which the difference in nucleotide sequence identity between the *Phytophthora* HGT candidate and its homolog sequence in the putative donor was lower than 52% (corresponding to the uppermost quartile of the distribution of the difference between the observed values and the expected values of nucleotide sequence identity between pairs of homologs from the two species) were retained after this stage. When the second identity test could not be performed due to the absence of sequence homologs in non-donor clades (i.e., Viriplantae, Amoebozoa, and Alveolata), the discrepancy cutoff, i.e., the difference between the nucleotide sequence identity for the *Phytophthora* HGT candidate and its homolog sequence in the putative donor and the average of the expected distribution was raised to 80.83% identity corresponding to the 5% upper quantile of the distribution of the difference between the observed values and the expected values. Proteins that passed this second filter were kept in a strict list of candidates.

We assessed if the HGT candidate could have been a bacterial or fungal contaminant mistakenly sequenced and assembled with a genome assembly generated in this study. HGT candidates were considered as putative contaminant candidates if they were the only coding sequences to map to a given scaffold.

Data availability.

Genome assemblies and genome and transcriptome sequences that were created in this study have been deposited in the NCBI database under BioProjects PRJNA746351, PRJNA714689, and PRJNA702516. *P. cactorum* and *P. idaei* data are available under BioProjects PRJNA383548 and PRJNA391273. Other genomic resources including assembly files, gene/protein predictions and annotations, differential expression analysis, and orthology analysis can be accessed on the Phytophthora Genome Sequencing Consortium download website.

Acknowledgments

We thank G. Browne and D. Kluepfel for the *P. pini* isolate and valuable inputs.

Author-Recommended Internet Resources

HMMER program: www.hmmer.org
JGI PhycoCosm: http://genome.jgi.doe.gov/Phyci1/Phyci1.home.html
MAKER Repeat Library Construction-Advanced site: http://weatherby.genetics.utah.edu/MAKER/wiki/index.php/Repeat_Library_Construction-Advanced
NCBI database: https://www.ncbi.nlm.nih.gov/genome/11752?genome_assembly_id=49439
Phytophthora Genome Sequencing Consortium download website: https://phyto-seq.cqls.oregonstate.edu
OrthoMCL web site: www.orthomcl.org
RepeatMasker and RepeatModeler programs: http://www.repeatmasker.org
United States Department of Agriculture fungal database: https://nt.ars-grin.gov/fungaldatabases

Literature Cited

Ali, S. S., Shao, J., Lary, D. J., Kronmiller, B. A., Shen, D., Strem, M. D., Amoako-Attah, I., Akrofi, A. Y., Begoude, B. A. D., ten Hoopen, G. M., Coulibaly, K., Kebe, B. I., Melnick, R. L., Guiltinan, M. J., Tyler, B. M., Meinhardt, L. W., and Bailey, B. A. 2017. *Phytophthora megakarya* and *Phytophthora palmivora*, closely related causal agents of cacao black pod rot, underwent increases in genome sizes and gene numbers by different mechanisms. Genome Biol. Evol. 9:536-557.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. 1990. Basic local alignment search tool. J Mol Biol. 215:403-410.

Amselem, J., Cuomo, C. A., van Kan, J. A. L., Viaud, M., Benito, E. P., Couloux, A., Coutinho, P. M., de Vries, R. P., Dyer, P. S., Fillinger, S., Fournier, E., Gout, L., Hahn, M., Kohn, L., Lapalu, N., Plummer, K. M., Pradier, J.-M., Quévillon, E., Sharon, A., Simon, A., ten Have, A., Tudzynski, B., Tudzynski, P., Wincker, P., Andrew, M., Anthouard, V., Beever, R. E., Beffa, R., Benoit, I., Bouzid, O., Brault, B., Chen, Z., Choquer, M., Collémare, J., Cotton, P., Danchin, E. G., da Silva, C., Gautier, A., Giraud, C., Giraud, T., Gonzalez, C., Grossetete, S., Güldener, U., Henrissat, B., Howlett, B. J., Kodira, C., Kretschmer, M., Lappartient, A., Leroch, M., Levis, C., Mauceli, E., Neuvéglise, C., Oeser, B., Pearson, M., Poulain, J., Poussereau, N., Quesneville, H., Rascle, C., Schumacher, J., Ségurens, B., Sexton, A., Silva, E., Sirven, C., Soanes, D. M., Talbot, N. J., Templeton, M., Yandava, C., Yarden, O., Zeng, Q., Rollins, J. A., Lebrun, M.-H., and Dickman, M. 2011. Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. PLoS Genet. 7:1002230.

Armitage, A. D., Lysøe, E., Nellist, C. F., Lewis, L. A., Cano, L. M., Harrison, R. J., and Brurberg, M. B. 2018. Bioinformatic characterisation of the effector repertoire of the strawberry pathogen *Phytophthora cactorum*. PLoS One 13:e1002230e0202305.

Aronesty, E. 2013. Comparison of sequencing utility programs. Open Bioinforma J. 7:1-8.

Asai, S., and Shirasu, K. 2015. Plant cells under siege: Plant immune system versus pathogen effectors. Curr. Opin. Plant Biol. 28:1-8.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. 2000. Gene Ontology: Tool for the unification of biology. Nat. Genet. 25:25-29.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. v., Sirotkin, A. v., Vyahhi, N., Tesler, G., Alekseyev, M. A., and Pevzner, P. A. 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19:455-477.

Bao, W., Kojima, K. K., and Kohany, O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mobile DNA 6:11.

Bao, Z. 2002. Automated de novo identification of repeat sequence families in sequenced genomes. Genome Res. 12:1269-1276.

Bendtsen, J. D., Nielsen, H., von Heijne, G., and Brunak, S. 2004. Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. 340: 783-795.

Benson, G. 1999. Tandem repeats finder: A program to analyze DNA sequences. Nucleic Acids Res. 27:573-580.

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

Boetzer, M., Henkel, C. v., Jansen, H. J., Butler, D., and Pirovano, W. 2011. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics 27:578-579.

Boser, B. E., Guyon, I. M., and Vapnik, V. N. 1992. A training algorithm for optimal margin classifiers. Pages 144-152 in: COLT '92: Proceedings of the Fifth Annual Workshop on Computational Learning Theory, July 1992. Association for Computing Machinery, New York.

Bozkurt, T. O., Schornack, S., Win, J., Shindo, T., Ilyas, M., Oliva, R., Cano, L. M., Jones, A. M. E., Huitema, E., van der Hoorn, R. A. L., and Kamoun, S. 2011. *Phytophthora infestans* effector AVRblb2 prevents secretion of a plant immune protease at the haustorial interface. Proc. Natl. Acad. Sci. U.S.A. 108:20832-20837.

Brasier, C., Scanu, B., Cooke, D., and Jung, T. 2022. *Phytophthora*: An ancient, historic, biologically and structurally cohesive and evolutionarily successful generic concept in need of preservation. IMA Fungus 13:12.

Buchfink, B., Xie, C., and Huson, D. H. 2015. Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12:59-60.

Burki, F., Roger, A. J., Brown, M. W., and Simpson, A. G. B. 2020. The new tree of eukaryotes. Trends Ecol. Evol. 35:43-55.

Caillaud, M., Piquerez, S. J. M., Fabro, G., Steinbrenner, J., Ishaque, N., Beynon, J., and Jones, J. D. G. 2012. Subcellular localization of the *Hpa* RxLR effector repertoire identifies a tonoplast-associated protein HaRxL17 that confers enhanced plant susceptibility. Plant J. 69:252-265.

Cantarel, B. L., Korf, I., Robb, S. M. C., Parra, G., Ross, E., Moore, B., Holt, C., Sanchez Alvarado, A., and Yandell, M. 2007. MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res. 18:188-196.

Chen, F. 2006. OrthoMCL-DB: Querying a comprehensive multi-species collection of ortholog groups. Nucleic Acids Res. 34:D363-D368.

Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Sayers, E. W. 2016. GenBank. Nucleic Acids Res. 44:D67-D72.

Cuomo, C. A., Guldener, U., Xu, J.-R., Trail, F., Turgeon, B. G., di Pietro, A., Walton, J. D., Ma, L.-J., Baker, S. E., Rep, M., Adam, G., Antoniw, J., Baldwin, T., Calvo, S., Chang, Y.-L., DeCaprio, D., Gale, L. R., Gnerre, S., Goswami, R. S., Hammond-Kosack, K., Harris, L. J., Hilburn, K., Kennell, J. C., Kroken, S., Magnuson, J. K., Mannhaupt, G., Mauceli, E., Mewes, H.-W., Mitterbauer, R., Muehlbauer, G., Munsterkotter, M., Nelson, D., O'Donnell, K., Ouellet, T., Qi, W., Quesneville, H., Roncero, M. I. G., Seong, K.-Y., Tetko, I. v., Urban, M., Waalwijk, C., Ward, T. J., Yao, J., Birren, B. W., and Kistler, H. C. 2007. The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. Science 317:1400-1402.

Dick, M. W. 2001. The Peronosporomycetes. Pages 39-72 in: Systematics and Evolution, D. J. McLaughlin, E. G. McLaughlin, P. A. Lemke, eds. Springer, Berlin.

Dong, S., Kong, G., Qutob, D., Yu, X., Tang, J., Kang, J., Dai, T., Wang, H., Gijzen, M., and Wang, Y. 2012. The NLP toxin family in *Phytophthora sojae* includes rapidly evolving groups that lack necrosis-inducing activity. Mol. Plant-Microbe Interact. 25:896-909.

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

Drake, B., and Jones, G. 2017. Public value at risk from *Phytophthora ramorum* and *Phytophthora kernoviae* spread in England and Wales. J. Environ. Manage. 191:136-144.

Eddy, S. R. 1998. Profile hidden Markov models. Bioinformatics 14:755-763.

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., Qureshi, M., Richardson, L. J., Salazar, G. A., Smart, A., Sonnhammer, E. L. L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S. C. E., and Finn, R. D. 2019. The Pfam protein families database in 2019. Nucleic Acids Res. 47:D427-D432.

Ellinghaus, D., Kurtz, S., and Willhoeft, U. 2008. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC Bioinformatics 9:18.

Erwin, D. C., and Ribeiro, O. K. 1996. *Phytophthora* diseases worldwide. American Phytopathological Society, St. Paul, MN, U.S.A.

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

Feurtey, A., and Stukenbrock, E. H. 2018. Interspecific gene exchange as a driver of adaptive evolution in fungi. Annu. Rev. Microbiol. 72: 377-398.

Gazis, R., Kuo, A., Riley, R., LaButti, K., Lipzen, A., Lin, J., Amirebrahimi, M., Hesse, C. N., Spatafora, J. W., Henrissat, B., Hainaut, M., Grigoriev, I. V., and Hibbett, D. S. 2016. The genome of *Xylona heveae* provides a window into fungal endophytism. Fungal Biol. 120:26-42.

Gene Ontology Consortium. 2019. The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res. 47:D330-D338.

Gijzen, M. 2009. Runaway repeats force expansion of the *Phytophthora infestans* genome. Genome Biol. 10:241.

Goheen, E. M., Hansen, E. M., Kanaskie, A., McWilliams, M. G., Osterbauer, N., and Sutton, W. 2002. Sudden oak death caused by *Phytophthora ramorum* in Oregon. Plant Dis. 86:441-441.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., and Regev, A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29:644-652.

Gremme, G., Steinbiss, S., and Kurtz, S. 2013. GenomeTools: A comprehensive software library for efficient processing of structured genome annotations. IEEE ACM T. Comput. Bi. 10:645-656.

Grunwald, N. J., Goss, E. M., and Press, C. M. 2008. *Phytophthora ramorum*: A pathogen with a remarkably wide host range causing sudden oak death on oaks and ramorum blight on woody ornamentals. Mol. Plant Pathol. 9.

Gu, S., Fang, L., and Xu, X. 2013. Using SOAPaligner for short reads alignment. Curr. Protoc. Bioinformatics 44:11.11.1-17.

Gunderson, J. H., Elwood, H., Ingold, A., Kindle, K., and Sogin, M. L. 1987. Phylogenetic relationships between chlorophytes, chrysophytes, and oomycetes. Proc. Natl. Acad. Sci. U.S.A. 84:729-7405823-5827.

Haas, B. J., Kamoun, S., Zody, M. C., Jiang, R. H. Y., Handsaker, R. E., Cano, L. M., Grabherr, M., Kodira, C. D., Raffaele, S., Torto-Alalibo, T., Bozkurt, T. O., Ah-Fong, A. M. V., Alvarado, L., Anderson, V. L., Armstrong, M. R., Avrova, A., Baxter, L., Beynon, J., Boevink, P. C., Bollmann, S. R., Bos, J. I. B., Bulone, V., Cai, G., Cakir, C., Carrington, J. C., Chawner, M., Conti, L., Costanzo, S., Ewan, R., Fahlgren, N., Fischbach, M. A., Fugelstad, J., Gilroy, E. M., Gnerre, S., Green, P. J., Grenville-Briggs, L. J., Griffith, J., Grünwald, N. J., Horn, K., Horner, N. R., Hu, C.-H., Huitema, E., Jeong, D.-H., Jones, A. M. E., Jones, J.

D. G., Jones, R. W., Karlsson, E. K., Kunjeti, S. G., Lamour, K., Liu, Z., Ma, L., MacLean, D., Chibucos, M. C., McDonald, H., McWalters, J., Meijer, H. J. G., Morgan, W., Morris, P. F., Munro, C. A., O'Neill, K., Ospina-Giraldo, M., Pinzón, A., Pritchard, L., Ramsahoye, B., Ren, Q., Restrepo, S., Roy, S., Sadanandom, A., Savidor, A., Schornack, S., Schwartz, D. C., Schumann, U. D., Schwessinger, B., Seyer, L., Sharpe, T., Silvar, C., Song, J., Studholme, D. J., Sykes, S., Thines, M., van de Vondervoort, P. J. I., Phuntumart, V., Wawra, S., Weide, R., Win, J., Young, C., Zhou, S., Fry, W., Meyers, B. C., van West, P., Ristaino, J., Govers, F., Birch, P. R. J., Whisson, S. C., Judelson, H. S., and Nusbaum, C. 2009. Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. Nature. 461:393-398.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., MacManes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., Henschel, R., LeDuc, R. D., Friedman, N., and Regev, A. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8:1494-1512.

Han, Y., and Wessler, S. R. 2010. MITE-Hunter: A program for discovering miniature inverted-repeat transposable elements from genomic sequences. Nucleic Acids Res. 38:e199-e199.

Hardham, A. R. 2005. *Phytophthora cinnamomi*.. Mol. Plant Pathol. 6: 589-604.

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., and Vinh, L. S. 2018. UFBoot2: Improving the ultrafast bootstrap approximation. Mol. Biol. Evol. 35:518-522.

Husnik, F., and McCutcheon, J. P. 2018. Functional horizontal gene transfer from bacteria to eukaryotes. Nat. Rev. Microbiol. 16:67-79.

Jiang, R. H. Y., Tripathy, S., Govers, F., and Tyler, B. M. 2008. RXLR effector reservoir in two *Phytophthora* species is dominated by a single rapidly evolving superfamily with more than 700 members. Proc. Natl. Acad. Sci. U.S.A. 105:4874-4879.

Jiang, R. H. Y., and Tyler, B. M. 2012. Mechanisms and evolution of virulence in oomycetes. Annu. Rev. Phytopathol. 50:295-318.

Johansen, L. K., and Carrington, J. C. 2001. Silencing on the spot. induction and suppression of RNA silencing in the *Agrobacterium*-mediated transient expression system. Plant Physiol. 126:930-938.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., and Hunter, S. 2014. InterProScan 5: Genome-scale protein function classification. Bioinformatics 30:1236-1240.

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

Jung, T., Jung, M. H., Cacciola, S. O., Cech, T., Bakonyi, J., Seress, D., Mosca, S., Schena, L., Seddaiu, S., Pane, A., di San Lio, G. M., Maia, C., Cravador, A., Franceschini, A., and Scanu, B. 2017. Multiple new cryptic pathogenic Phytophthora species from Fagaceae forests in Austria, Italy and Portugal. IMA Fungus 8:219-244.

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

Kamoun, S. 1993. Extracellular protein elicitors from *Phytophthora*: Host-specificity and induction of resistance to bacterial and fungal phytopathogens. Mol. Plant-Microbe Interact. 6:15-25.

Kao, C. W., and Leu, L. S. 1980. Sporangium germination of *Peronophythora litchii*, the causal organism of litchi downy blight. Mycologia 72: 15737-748.

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

Keeling, P. J., and Burki, F. 2019. Progress towards the tree of eukaryotes. Curr. Biol. 29:808-817.

Keeling, P. J., and Palmer, J. D. 2008. Horizontal gene transfer in eukaryotic evolution. Nat. Rev. Genet. 9:605-618.

Kennedy, D. M., Duncan, J. M., Dugard, P. I., and Topham, P. H. 1986. Virulence and aggressiveness of single-zoospore isolates of *Phytophthora fragariae*. Plant Pathol. 35:344-354.

Kent, W. J. 2002. BLAT—The BLAST-like alignment tool. Genome Res. 12:656-664.

Knaus, B. J., Tabima, J. F., Shakya, S. K., Judelson, H. S., and Grünwald, N. J. 2020. Genome-wide increased copy number is associated with emergence of dominant clones of the Irish potato famine pathogen *Phytophthora infestans*. mBio. 11:e00326-20.

Korf, I. 2004. Gene finding in novel genomes. BMC Bioinformatics 5:59.

Kroon, L. P. N. M., Brouwer, H., de Cock, A. W. A. M., and Govers, F. 2012. The genus *Phytophthora* anno 2012. Phytopathology 102:348-364.

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

Li, L. 2003. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. Genome Res. 13:2178-2189.

Liu, T., Song, T., Zhang, X., Yuan, H., Su, L., Li, W., Xu, J., Liu, S., Chen, L., Chen, T., Zhang, M., Gu, L., Zhang, B., and Dou, D. 2014. Unconventionally secreted effectors of two filamentous pathogens target plant salicylate biosynthesis. Nat. Commun. 5.

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C., Cheung, D. W., Yiu, S.-M., Peng, S., Xiaoqian, Z., Liu, G., Liao, X., Li, Y., Yang, H., Wang, J., Lam, T.-W., and Wang, J. 2012. SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler. Gigascience 1:18.

Marçais, G., and Kingsford, C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. Bioinformatics 27:764-770.

Martin, F., Kohler, A., Murat, C., Balestrini, R., Coutinho, P. M., Jaillon, O., Montanini, B., Morin, E., Noel, B., Percudani, R., Porcel, B., Rubini, A., Amicucci, A., Amselem, J., Anthouard, V., Arcioni, S., Artiguenave, F., Aury, J.-M., Ballario, P., Bolchi, A., Brenna, A., Brun, A., Buée, M., Cantarel, B., Chevalier, G., Couloux, A., da Silva, C., Denoeud, F., Duplessis, S., Ghignone, S., Hilselberger, B., Iotti, M., Marçais, B., Mello, A., Miranda, M., Pacioni, G., Quesneville, H., Riccioni, C., Ruotolo, R., Splivallo, R., Stocchi, V., Tisserant, E., Viscomi, A. R., Zambonelli, A., Zampieri, E., Henrissat, B., Lebrun, M.-H., Paolocci, F., Bonfante, P., Ottonello, S., and Wincker, P. 2010. Périgord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. Nature. 464: 1033-1038.

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

Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 17:10-12.

McCarthy, C. G. P., and Fitzpatrick, D. A. 2016. Systematic search for evidence of interdomain horizontal gene transfer from prokaryotes to oomycete lineages. mSphere 1:e00195-16.

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

McGowan, J., and Fitzpatrick, D. A. 2017. Genomic, network, and phylogenetic analysis of the oomycete effector arsenal. mSphere 2:e00408-17.

Möller, E. M., Bahnweg, G., Sandermann, H., and Geiger, H. H. 1992. A simple and efficient protocol for isolation of high molecular weight DNA from filamentous fungi, fruit bodies, and infected plant tissues. Nucleic Acids Res. 20:6115-6116.

Morgan, W., and Kamoun, S. 2007. RXLR effectors of plant pathogenic oomycetes. Curr. Opin. Microbiol. 10:332-338.

Nakamura, T., Yamada, K. D., Tomii, K., and Katoh, K. 2018. Parallelization of MAFFT for large-scale multiple sequence alignments. Bioinformatics 34:2490-2492.

Nellist, C. F., Armitage, A. D., Bates, H. J., Sobczyk, M. K., Luberti, M., Lewis, L. A., and Harrison, R. J. 2021. Comparative analysis of host-associated variation in *Phytophthora cactorum*. Front Microbiol. 12.

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

Novichkov, P. S., Omelchenko, M. v., Gelfand, M. S., Mironov, A. A., Wolf, Y. I., and Koonin, E. V. 2004. Genome-wide molecular clock and horizontal gene transfer in bacterial evolution. J. Bacteriol. 186: 6575-6585.

Oh, S.-K., Young, C., Lee, M., Oliva, R., Bozkurt, T. O., Cano, L. M., Win, J., Bos, J. I. B., Liu, H.-Y., van Damme, M., Morgan, W., Choi, D., van der Vossen, E. A. G., Vleeshouwers, V. G. A. A., and Kamoun, S. In 2009. Planta expression screens of *Phytophthora infestans* RXLR effectors reveal diverse phenotypes, including activation of the *Solanum bulbocastanum* disease resistance protein Rpi-blb2. Plant Cell 21: 2928-2947.

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V.,

Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., Farrell, C. M., Goldfarb, T., Gupta, T., Haft, D., Hatcher, E., Hlavina, W., Joardar, V. S., Kodali, V. K., Li, W., Maglott, D., Masterson, P., McGarvey, K. M., Murphy, M. R., O'Neill, K., Pujar, S., Rangwala, S. H., Rausch, D., Riddick, L. D., Schoch, C., Shkeda, A., Storz, S. S., Sun, H., Thibaud-Nissen, F., Tolstoy, I., Tully, R. E., Vatsan, A. R., Wallin, C., Webb, D., Wu, W., Landrum, M. J., Kimchi, A., Tatusova, T., DiCuccio, M., Kitts, P., Murphy, T. D., and Pruitt, K. D. 2016. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 44:D733-D45.

Ou, S., and Jiang, N. 2018. LTR_retriever: A highly accurate and sensitive program for identification of long terminal repeat retrotransposons. Plant Physiol. 176:1410-1422.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, É. 2011. Scikit-Learn: Machine learning in Python. J. Mach. Learn. Res. 12:2825-2830.

Price, A. L., Jones, N. C., and Pevzner, P. A. 2005. De novo identification of repeat families in large genomes. Bioinformatics 21:i351-i358.

R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. https://www.R-project.org

Raffaele, S., and Kamoun, S. 2012. Genome evolution in filamentous plant pathogens: Why bigger can be better. Nat. Rev. Microbiol. 10: 417-430.

Rawlings, N. D., Barrett, A. J., and Bateman, A. 2012. MEROPS: The database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res. 40:D343-D350.

Rice, P., Longden, I., and Bleasby, A. 2000. EMBOSS: The European molecular biology open software suite. Trends in Genet. 16:276-277.

Richards, T. A., Dacks, J. B., Jenkinson, J. M., Thornton, C. R., and Talbot, N. J. 2006. Evolution of filamentous plant pathogens: Gene exchange across eukaryotic kingdoms. Curr. Biol. 16:1857-1864.

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

Rizzo, D. M., Garbelotto, M., Davidson, J. M., Slaughter, G. W., and Koike, S. T. 2002. *Phytophthora ramorum* as the cause of extensive mortality of *Quercus* spp. and *Lithocarpus densiflorus* in California. Plant Dis. 86:205-214.

Rizzo, D. M., Garbelotto, M., and Hansen, E. M. 2005. *Phytophthora ramorum*: Integrative research and management of an emerging pathogen in California and Oregon Forests. Annu. Rev. Phytopathol. 43:309-335.

Rouxel, T., Grandaubert, J., Hane, J. K., Hoede, C., van de Wouw, A. P., Couloux, A., Dominguez, V., Anthouard, V., Bally, P., Bourras, S., Cozijnsen, A. J., Ciuffetti, L. M., Degrave, A., Dilmaghani, A., Duret, L., Fudal, I., Goodwin, S. B., Gout, L., Glaser, N., Linglin, J., Kema, G. H. J., Lapalu, N., Lawrence, C. B., May, K., Meyer, M., Ollivier, B., Poulain, J., Schoch, C. L., Simon, A., Spatafora, J. W., Stachowiak, A., Turgeon, B. G., Tyler, B. M., Vincent, D., Weissenbach, J., Amselem, J., Quesneville, H., Oliver, R. P., Wincker, P., Balesdent, M.-H., and Howlett, B. J. 2011. Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by Repeat-Induced Point mutations. Nat. Commun. 2:202.

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

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

Sharp, P. M., and Matassi, G. 1994. Codon usage and genome evolution. Curr. Opin. Genet. Dev. 4:851-860.

Shen, D., Liu, T., Ye, W., Liu, L., Liu, P., Wu, Y., Wang, Y., and Dou, D. 2013. Gene duplication and fragment recombination drive functional diversification of a superfamily of cytoplasmic effectors in *Phytophthora sojae*. PLoS One 8:e70036.

Simpson, J. T., Wong, K., Jackman, S. D., Schein, J. E., Jones, S. J. M., and Birol, I. 2009. ABySS: A parallel assembler for short read sequence data. Genome Res. 19:1117-1123.

Smit, A., and Hubley, R. 2015a. RepeatModeler Open-1.0. Institute for Systems Biology, Seattle, WA, U.S.A. http://www.repeatmasker.org

Smit, A., Hubley, R., and Green, P. 2015b. RepeatMasker Open-4.0. Institute for Systems Biology, Seattle, WA, U.S.A. http://www.repeatmasker.org

Soucy, S. M., Huang, J., and Gogarten, J. P. 2015. Horizontal gene transfer: Building the web of life. Nat. Rev. Genet. 16:472-482.

Stam, R., Jupe, J., Howden, A. J. M., Morris, J. A., Boevink, P. C., Hedley, P. E., and Huitema, E. 2013. Identification and characterisation CRN effectors in *Phytophthora capsici* shows modularity and functional diversity. PLoS One 8:e59517.

Stamatakis, A. 2014. RAxML version 8: 472-482 A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30: 1312-1313.

Stanke, M., and Morgenstern, B. 2005. AUGUSTUS: A web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. 33:W465-W467.

Stöver, B. C., and Müller, K. F. 2010. TreeGraph 2: Combining and visualizing evidence from different phylogenetic analyses. BMC Bioinformatics 11:7.

Studholme, D. J., McDougal, R. L., Sambles, C., Hansen, E., Hardy, G., Grant, M., Ganley, R. J., and Williams, N. M. 2016. Genome sequences of six *Phytophthora* species associated with forests in New Zealand. Genom. Data. 7:54-56.

Tabima, J. F., and Grünwald, N. J. 2019. *effectR*: An expandable R package to predict candidate RxLR and CRN effectors in oomycetes using motif searches. Mol. Plant-Microbe Interact. 32:1067-1076.

Tabima, J. F., Kronmiller, B. A., Press, C. M., Tyler, B. M., Zasada, I. A., and Grünwald, N. J. 2017. Whole genome sequences of the raspberry and strawberry pathogens *Phytophthora rubi* and *P. fragariae*. Mol. Plant-Microbe Interact. 30:767-769.

Thines, M. 2014. Phylogeny and evolution of plant pathogenic oomycetes—A global overview. Eur. J. Plant Pathol. 138:431-447.

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

Tian, M., Win, J., Song, J., van der Hoorn, R., van der Knaap, E., and Kamoun, S. 2007. A *Phytophthora infestans* cystatin-like protein targets a novel tomato papain-like apoplastic protease. Plant Physiol. 143: 364-377.

Torto, T. A. 2003. EST mining and functional expression assays identify extracellular effector proteins from the plant pathogen *Phytophthora*. Genome Res. 13:1675-1685.

Torto-Alalibo, T., Collmer, C. W., Gwinn-Giglio, M., Lindeberg, M., Meng, S., Chibucos, M. C., Tseng, T.-T., Lomax, J., Biehl, B., Ireland, A., Bird, D., Dean, R. A., Glasner, J. D., Perna, N., Setubal, J. C., Collmer, A., and Tyler, B. M. 2010. Unifying themes in microbial associations with animal and plant hosts described using the gene ontology.. Microbiol. Mol. Biol. Rev. 74:479-503.

Turner, R. S. 2005. After the famine: Plant pathology, *Phytophthora infestans*, and the late blight of potatoes, 1845-1960. Hist. Stud. Phys. Biol. 35:341-370.

Tyler, B. M. 2006. *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. Science 313:1261-1266.

Tyler, B. M. 2007. *Phytophthora sojae*: Root rot pathogen of soybean and model oomycete. Mol. Plant Pathol. 8:1-8.

UniProt Consortium. 2019. UniProt: A worldwide hub of protein knowledge. Nucleic Acids Res. 47:D506-D515.

Urban, M., Cuzick, A., Seager, J., Wood, V., Rutherford, K., Venkatesh, S. Y., De Silva, N., Martinez, M. C., Pedro, H., Yates, A. D., Hassani-Pak, K., and Hammond-Kosack, K. E. 2019. PHI-base: The pathogen-host interactions database. Nucleic Acids Res. 48:D613-D620.

van den Berg, M. A., Albang, R., Albermann, K., Badger, J. H., Daran, J.-M., Driessen, M., A., J., Garcia-Estrada, C., Fedorova, N. D., Harris, D. M., Heijne, W. H. M., Joardar, V., Kiel, W., J., A. K., Kovalchuk, A., Martín, J. F., Nierman, W. C., Nijland, J. G., Pronk, J. T., Roubos, J. A., van der Klei, I. J., van Peij, N. N. M. E., Veenhuis, M., von Döhren, H., Wagner, C., Wortman, J., and Bovenberg, R. A. L. 2008. Genome sequencing and analysis of the filamentous fungus *Penicillium chrysogenum*. Nat Biotechnol. 26:1161-1168.

van der Lee, T., de Witte, I., Drenth, A., Alfonso, C., and Govers, F. 1997. AFLP linkage map of the oomycete *Phytophthora infestans*. Fungal Genet. Biol. 21:278-291.

van Etten, J., and Bhattacharya, D. 2020. Horizontal gene transfer in eukaryotes: Not if, but how much?. Trends Genet. 36:915-915.

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

Wang, Q., Han, C., Ferreira, A. O., Yu, X., Ye, W., Tripathy, S., Kale, S. D., Gu, B., Sheng, Y., Sui, Y., Wang, X., Zhang, Z., Cheng, B., Dong, S., Shan, W., Zheng, X., Dou, D., Tyler, B. M., and Wang, Y. 2011. Transcriptional programming and functional interactions within the *Phytophthora sojae* RXLR effector repertoire. Plant Cell 23:2064-2086.

Wang, S., Boevink, P. C., Welsh, L., Zhang, R., Whisson, S. C., and Birch, P. R. J. 2017. Delivery of cytoplasmic and apoplastic effectors from *Phy-*

tophthora infestans haustoria by distinct secretion pathways. New Phytol. 216:205-215.
Wang, W., and Jiao, F. 2019. Effectors of *Phytophthora* pathogens are powerful weapons for manipulating host immunity. Planta 250:413-425.
Wang, Y., Tyler, B. M., and Wang, Y. 2019. Defense and counterdefense during plant-pathogenic oomycete infection. Annu. Rev. Microbiol. 73: 667-696.
Wickham, H. 2016. ggplot2: Elegant graphics for data analysis. Springer-Verlag, Berlin.
Win, J., Kanneganti, T.-D., Torto-Alalibo, T., and Kamoun, S. 2006. Computational and comparative analyses of 150 full-length cDNA sequences from the oomycete plant pathogen *Phytophthora infestans*. Fungal Genet. Biol. 43:20-33.
Win, J., Morgan, W., Bos, J., Krasileva, K. v., Cano, L. M., Chaparro-Garcia, A., Ammar, R., Staskawicz, B. J., and Kamoun, S. 2007. Adaptive evolution has targeted the C-terminal domain of the RXLR effectors of plant pathogenic oomycetes. Plant Cell 19:2349-2369.
Wootton, J. C., and Federhen, S. 1996. Analysis of compositionally biased regions in sequence databases. Methods Enzymol. 266:554-571.
Wu, X.-M. 2007. The analysis method and progress in the study of codon bias. Hereditas 29:420.
Xu, Z., and Wang, H. 2007. LTR_FINDER: An efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res. 35: W265-W268.
Yang, J., Wang, L., Ji, X., Feng, Y., Li, X., Zou, C., Xu, J., Ren, Y., Mi, Q., Wu, J., Liu, S., Liu, Y., Huang, X., Wang, H., Niu, X., Li, J., Liang, L., Luo, Y., Ji, K., Zhou, W., Yu, Z., Li, G., Liu, Y., Li, L., Qiao, M., Feng, L., and Zhang, K.-Q. 2011. Genomic and proteomic analyses of the fungus *Arthrobotrys oligospora* provide insights into nematode-trap formation. PLoS Pathog. 7:e1002179.
Yang, X., Tyler, B. M., and Hong, C. 2017. An expanded phylogeny for the genus *Phytophthora*. IMA Fungus 8:355-384.
Ye, W., Wang, Y., Shen, D., Li, D., Pu, T., Jiang, Z., Zhang, Z., Zheng, X., Tyler, B. M., and Wang, Y. 2016a. Sequencing of the litchi downy blight pathogen reveals it is a *Phytophthora* species with downy mildew-like characteristics. Mol. Plant-Microbe Interact. 29:573-583.
Ye, W., Wang, Y., Tyler, B. M., and Wang, Y. 2016b. Comparative genomic analysis among four representative isolates of *Phytophthora sojae* reveals genes under evolutionary selection. Front. Microbiol. 7:1547.
Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., and Xu, Y. 2012. dbCAN: A web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 40:W445-W451.
Zerbino, D. R., and Birney, E. 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18:821-829.
Zhang, M., Li, Q., Liu, T., Liu, L., Shen, D., Zhu, Y., Liu, P., Zhou, J.-M., and Dou, D. 2014. Two cytoplasmic effectors of *Phytophthora sojae* regulate plant cell death via interactions with plant catalases. Plant Physiol. 167:164-175.