Draft Genome Sequence for the Tree Pathogen *Phytophthora plurivora*

Ramesh R. Vetukuri[1,*], Sucheta Tripathy[2,3,*,†], Mathu Malar C[2,3,†], Arijit Panda[2,3], Sandeep K. Kushwaha[4,5], Aakash Chawade[4], Erik Andreasson[1], Laura J. Grenville-Briggs[1], and Stephen C. Whisson[6]

[1]Department of Plant Protection Biology, Swedish University of Agricultural Sciences, Alnarp, Sweden

[2]Computational Genomics Laboratory, Structural Biology and Bioinformatics Division, Council of Scientific and Industrial Research, Indian Institute of Chemical Biology, Kolkata, India

[3]Academy of Scientific and Innovative Research, Ghaziabad, Uttar Pradesh, India

[4]Department of Plant Breeding, Swedish University of Agricultural Sciences, Alnarp, Sweden

[5]National Bioinformatics Infrastructure Sweden (NBIS), Department of Biology, Lund University, Sweden

[6]Cell and Molecular Sciences, The James Hutton Institute, Dundee, United Kingdom

[†]Joint first authors and these authors contributed equally to this work.

*Corresponding authors: E-mails: ramesh.vetukuri@slu.se; tsucheta@gmail.com.

Accepted: July 28, 2018

Data deposition: This project has been deposited at DDBJ/ENA/GenBank under the accessions NMPK00000000 and SRP132452. The Genome version described in this paper is version NMPK01000000. Raw sequence data can be found in the NCBI Sequence Read Archive, with accession number SRP132452.

Abstract

Species from the genus *Phytophthora* are well represented among organisms causing serious diseases on trees. *Phytophthora plurivora* has been implicated in long-term decline of woodland trees across Europe. Here we present a draft genome sequence of *P. plurivora*, originally isolated from diseased European beech (*Fagus sylvatica*) in Malmö, Sweden. When compared with other sequenced *Phytophthora* species, the *P. plurivora* genome assembly is relatively compact, spanning 41 Mb. This is organized in 1,919 contigs and 1,898 scaffolds, encompassing 11,741 predicted genes, and has a repeat content of approximately 15%. Comparison of allele frequencies revealed evidence for tetraploidy in the sequenced isolate. As in other sequenced *Phytophthora* species, *P. plurivora* possesses genes for pathogenicity-associated RXLR and Crinkle and Necrosis effectors, predominantly located in gene-sparse genomic regions. Comparison of the *P. plurivora* RXLR effectors with orthologs in other sequenced species in the same clade (*Phytophthora multivora* and *Phytophthora capsici*) revealed that the orthologs were likely to be under neutral or purifying selection.

Key words: *Phytophthora plurivora*, two speed genome, root rot pathogen, *P. plurivora* genome, purifying selection.

Introduction

Species of the genus *Phytophthora* are all known to be pathogens of plants, some with a narrow host range, and others able to infect hundreds of different plant species (Erwin and Ribeiro 1996). *Phytophthora* species superficially resemble fungi but are hyphae-forming oomycetes, and are placed in the stramenopiles (syn. Heterokonta; kingdom Chromista, SAR supergroup), along with brown algae and diatoms (Cavalier-Smith 2018). *Phytophthora plurivora* is a soil-borne root rot pathogen infecting a broad range of woody plants, such as *Quercus* spp., *Acer* spp., *Alnus* spp., *Vaccinium* spp., *Rhododendron* spp., and *Fagus* spp. (Jung and Burgess 2009; Schoebel et al. 2014). It is predominantly implicated in widespread declines of European beech (*Fagus sylvatica*) and oak species (*Quercus* spp.) (Jung et al. 2000; Jung 2009). Common symptoms of *P. plurivora* disease include collar rots, bark cankers, extensive damage to fine roots, and crown dieback on young and mature trees (Jung et al. 2005; Orlikowski et al. 2011). Stem inoculations to test the relative susceptibility of conifers and broadleaved tree species common in Sweden demonstrated *P. plurivora* to be highly aggressive on pedunculate oak (*Quercus robur* L.), European

beech and black cottonwood (*Populus trichocarpa*), highlighting the overall risk of different *Phytophthora* species to forest trees (Cleary et al. 2017).

The most recent phylogeny of the genus *Phytophthora* places *P. plurivora* in Clade 2, subclade 2c, together with related pathogens *Phytophthora acerina*, *Phytophthora pachyleura*, *Phytophthora capensis*, *Phytophthora pini*, *Phytophthora multivora*, and *Phytophthora citricola* (Yang et al., 2017). *P. plurivora* is proposed to be most likely native to Europe based on haplotype and microsatellite data analysis (Schoebel et al. 2014). It is now distributed worldwide, aided by dissemination of diseased plant material through the plant nursery trade (Schoebel et al. 2014). In southern Sweden, *P. plurivora*, along with *P. cactorum*, has been recognized as an increasing threat to cultivated plantation forests (Cleary et al. 2017; Grenville-Briggs et al. 2017).

The great majority of studies on *Phytophthora* species have focused on those that infect important annual crop plants, for example, *Phytophthora infestans* that infects potato and tomato, and *Phytophthora sojae* that infects soybean (Kamoun et al., 2015). Here, we present a draft genome sequence assembly for a *P. plurivora* strain recently isolated from Sweden, as a resource to enable future studies that aim to develop a greater understanding of how *Phytophthora* species such as *P. plurivora* cause disease on woody host plants. We then used this resource to examine the complement of effector coding genes to identify rapidly evolving candidate RXLR and Crinkler effector coding genes (Schornack et al. 2009), collinearity of genomic regions, and whether the effector genes reside in gene poor genomic regions as in other *Phytophthora* species.

Materials and Methods

Sample Collection and Sequencing

P. plurivora strain AV1007 was isolated from a bleeding canker on a diseased European beech tree (*F. sylvatica*) in Malmö, Sweden in 2016. *P. plurivora* was cultured in liquid V8 juice media and DNA extracted as described (Löbmann et al. 2016). Paired-end reads of 2 × 250 bp (37 million reads from each library) were sequenced using the Illumina Hiseq2000 rapid mode sequencing platform at MR DNA Molecular Research Laboratory, USA. Adapter removal was performed using Trimmomatic, and FastQC tools were used for raw data assessment (Bolger et al. 2014). To prepare mycelium samples for RNA isolation, *P. plurivora* was grown in liquid V8 medium for three days at 20 °C, collected by gravity filtration, snap frozen in liquid nitrogen, and stored at −70 °C until used for RNA extraction. Three independent biological replicates were prepared. Total RNA for RNA sequencing was extracted from frozen mycelium samples using the RNeasy Plant mini kit (QIAGEN) following the manufacturer's protocol. Yield and integrity of the RNA was assessed using a NanoDrop Micro Photometer (NanoDrop Technologies), Agilent Bioanalyzer (Agilent Research Laboratories) and agarose gel electrophoresis, respectively. Library preparation for RNA sequencing was carried out using Illumina TruSeq RNA preparation kit using poly-A selection. Illumina HiSeq2500 2 × 126 bp paired-end sequencing was performed at SciLifeLab, Sweden. The reads obtained (Library 1: 30193560, Library 2: 36829808 and Library 3: 38506144) were mapped onto the reference genome, and the Tuxedo suite (Trapnell et al. 2012) was used for analysis. Expression levels of genes were assessed using the Tuxedo pipeline.

Genome Assembly and Annotation

The Illumina sequence reads were assembled *de novo* using SPAdes 3.5.0, a de-Bruijn graph-based assembler which incorporates error correction and removal of poor quality reads (Bankevich et al. 2012). The assembled genome was assessed using QUAST (Gurevich et al. 2013). To evaluate if the assembled genome size for *P. plurivora* is a marked underestimate of true genome size, such as may occur due to a high repeat content, we calculated a genome size estimate using k-mer analysis of our sequence data (Schell et al. 2017). We used Jellyfish (Marçais and Kingsford 2011) for generating the k-mer histograms with a k-mer value of 25 and minimum phred quality of reads at 20. The histograms generated by Jellyfish were fed into GenomeScope (github.com/schatzlab/genomescope; last accessed August 23, 2018). We also computed genome sizes for *Phytophthora infestans* and *P. multivora* for comparison purposes using the same method. The raw sequence reads for *P. infestans* and *P. multivora* (isolate NZFS 3378) were downloaded from NCBI (https://trace.ncbi.nlm.nih.gov/Traces/sra/? run=ERR1990236 and https://trace.ncbi.nlm.nih.gov/Traces/sra/? run=SRR2126502). Augustus (Stanke et al. 2004) was used for gene prediction using the previously sequenced plant pathogenic relative, *Phytophthora capsici*, as a training model. The gene space assessment of *P. plurivora* was evaluated by BUSCO 2.0 (Benchmarking Universal Single-Copy Orthologs [BUSCOs]) for completeness based on a set of common stramenopile genes (Simão et al. 2015). For comparison, BUSCO analysis was also carried out for genome assemblies for *P. infestans*, *P. capsici*, *P. ramorum*, and *P. multivora*. The predicted protein sequences were submitted to Interproscan 5 (Jones et al. 2014) on our local server to annotate Pfam domain functions in predicted protein sequences.

Repeat and Effector Analysis

Analysis of repetitive DNA sequences was carried out using RepeatModeler 1.0.10 (http://www.repeatmasker.org/; Last accessed August 23, 2018) for *de novo* repeat identification, using oomycete genomes as a model repeat library. Repeat finding in *P. plurivora* was carried out using RepeatMasker

(http://www.repeatmasker.org) (Smit et al. 2013–2015), using the consensus file generated by RepeatModeler.

Four different pipelines were used for RXLR effector prediction. The most stringent pipeline was based on an in-house protocol (supplementary fig. 1*a*, Supplementary Material online, Effector Prediction Pipeline A) and the second pipeline was based on EffectorP 1.0, a fungal-based pipeline (supplementary fig. 1*a*, Supplementary Material online, Effector Prediction Pipeline B) (Sperschneider et al. 2016). The other two prediction pipelines involved using earlier RXLR effector prediction methods (Whisson et al. 2007; Win et al. 2007) from a Galaxy server (Afgan et al. 2016) (supplementary fig. 1*a*, Supplementary Material online, Effector Prediction Pipelines C and D) (Cock et al. 2013). The in-house effector prediction method was used to predict effectors for *P. plurivora* and other closely related oomycetes such as *P. multivora* (two isolates) and *P. capsici* for comparative genomic analysis.

Prediction of Crinkle and Necrosis (CRN) effectors was performed as described previously (Haas et al. 2009; Yin et al. 2017). SignalP 3.0 and Phobius 1.01 searches to predict secretion signal peptides were performed as described by Yin et al. (2017).

Intergenic Distance and Synteny Analysis

Analysis of intergenic distances was carried as described in Saunders et al. (2014). Briefly, the 5′ and 3′ intergenic distances for all genes, effectors, and BUSCOs were 2D binned and plotted using R packages GenomicRanges, rtracklayer, Rsamtools, and ggplots. The Nucmer program from Mummer 3.0 (Kurtz et al. 2004) was used with maxgap 50 and breaklen 400 for comparing the genomes of *P. plurivora* with related species *P. multivora* and *P. capsici*.

Computing dN/dS Ratio for RXLR Effectors

RXLR effectors predicted by our in-house pipeline were analyzed using OrthoMCL (Li et al. 2003) for predicting clusters of orthologous RXLR effector genes from the genomes analyzed here. The clustered groups of effectors from six genomes (*P. plurivora*, *P. multivora* isolates NZFS 3378 and NZFS 3448, *P. capsici, P. ramorum*, and *P. cinnamomi*) were used for calculation of non-synonymous/synonymous codon substitution (dN/dS). Only OrthoMCL orthologous clusters containing at least three genes in a cluster were used for dN/dS ratio calculation using the maximum likelihood method PAML in the CODEML package (Yang 2007).

Pathogen–Host Interactions, CAZY, Secondary Metabolite, and Ploidy Analyses

To identify pathogenicity factors in common with other pathogens, BLASTP searches of all *P. plurivora* protein sequences were performed against the pathogen–host interactions (PHIs) database (Urban et al. 2015). Predicted protein sequences from the *P. plurivora* genome were explored for carbohydrate active enzymes by BLASTP sequence similarity search against the dbCAN database (http://csbl.bmb.uga.edu/dbCAN/; Last accessed August 23, 2018) using a $1e^{-10}$ e-value cut-off (Lombard et al. 2014). The entire genome assembly was uploaded to the antiSMASH 3.0 server (Weber et al. 2015; Blin et al. 2017) for identification of regions potentially involved in secondary metabolite production. To estimate the ploidy of the sequenced *P. plurivora* genome, we used ploidyNGS (Corrêa dos Santos et al. 2017) that derives ploidy information from allele frequencies present in the Illumina short reads.

Results and Discussion

Overview of the *P. Plurivora* Draft Genome Assembly

Here we have presented the first draft genome assembly for the plant pathogen, *P. plurivora*, the fourth species sequenced from Clade 2 of the *Phytophthora* genus. *P. plurivora* is a species recently separated from the *P. citricola* species complex (Jung and Burgess 2009), together with *P. multivora* and *P. pini* (Hong et al. 2009; Scott et al. 2009). Using SPAdes 3.5.0, a total of 41 Mb of *P. plurivora* genome was assembled into 1,919 contigs and 1,898 scaffolds (mean coverage, 220×). Contigs below 2 kb were removed from the assembly, and mitochondrial genome sequences were not screened out; 9% of sequence reads were unassembled and were discarded. This represents one of the smallest draft genome assemblies for a *Phytophthora* species (McGowan and Fitzpatrick 2017); *P. multivora*, *P. kernoviae*, and *P. agathidicida* have similar genome assembly sizes of 40, 43, and 37 Mb, respectively (supplementary fig. 2*a*, Supplementary Material online). Other *Phytophthora* genomes sequenced to date are typically in excess of 50 Mb and up to 240 Mb (Haas et al. 2009). Assessment of assembly quality revealed: N50 = 48,620 bp; N75 = 21,603 bp; L50 = 242; L75 = 547; longest contig = 294,496 bp; number of contigs >25 kb = 489; number of contigs >10 kb = 921). Using the previously sequenced plant pathogenic relative, *P. capsici* as a training model, gene prediction using Augustus predicted 11,749 genes. Preliminary annotation of *P. plurivora* predicted genes revealed 6,353 sequences having Pfam domains. The gene sequences predicted from the *P. plurivora* genome are available in EumicrobeDB (www.eumicrobedb.org; Last accessed August 23, 2018) (Panda et al. 2018). From an RNAseq dataset generated from *in vitro* cultured mycelium of *P. plurivora*, over half the genes predicted were detected with a fragments per kilobase of transcript per million mapped reads (FPKM) value of 10 or greater, while approximately 74% of predicted genes were detected at an FPKM of 5 or greater. All the predicted genes (11,749) had overlaps with RNAseq data (Supplementary material 1, Supplementary Material online). Similar and lower proportions of predicted genes were

Table 1
Summary of Genomes Mentioned in this Study

Genome Described here	Total Genes	Genome Size (Mb)	Number of Predicted RXLRs	Total No of Contigs/Scaffolds	Host
P. plurivora	11,741	41	84 (Pipeline A)	1,898	*F. sylvatica*
P. multivora isolate 1 (NZFS 3378)	14,200	40	92 (Pipeline A)	2,844	*Idesia polycarpa*
P. multivora isolate 2 (NZFS 3448)	15,091	40	84 (Pipeline A)	2,840	*Metrosideros kermadecensis*
P. ramorum	16,134	65	370 (Jiang et al. 2008)	2,576	*Quercus agrifolia*
P. capsici	20,378	62	140 (Pipeline A)	917	Laboratory backcross progeny
P. cinnamomi	26,132	58	565 (Studholme et al. 2016; McGowan and Fitzpatrick 2017)	1,314	*Eucalyptus marginata*

expressed at these levels in *in vitro* cultured mycelium of *P. infestans* (Ah-Fong et al. 2017) and *P. capsici* (Chen et al. 2013), respectively. This suggests that the accuracy of our gene predictions for *P. plurivora* compares favorably to those in other *Phytophthora* species.

An assessment of gene space representation revealed that 226 out of the 234 stramenopile BUSCOs (96.6%) were represented as single complete copies in the *P. plurivora* genome. No complete duplicated BUSCOs were identified, while one was fragmented and seven were not found (3.0%). By comparison, 16 BUSCOs were not found in *P. capsici*, 10 and 12 BUSCOs were absent from the genome assemblies of two sequenced isolates of *P. multivora* (NZFS 3378 and NZFS 3448, respectively), four BUSCOs were absent from the *P. ramorum* genome, and nine BUSCOs were absent from the *P. infestans* genome. We compared the absent BUSCOs with each other and found none that were common to genome assemblies of *P. plurivora*, two isolates of *P. multivora*, *P. capsici*, *P. ramorum*, or *P. infestans* (supplementary fig. 3, Supplementary Material online and supplementary material 2, Supplementary Material online). The BUSCO analysis suggests that our genome assembly is highly representative of the gene space in *P. plurivora*, and compares favorably to other *Phytophthora* genome assemblies.

To further evaluate our draft *P. plurivora* genome, we compared the assembly statistics to those for *P. multivora* (two isolates), *P. kernoviae*, and *P. agathidicida*, which were also sequenced and assembled using similar strategies (Studholme et al. 2016). The *P. plurivora* assembly also compares favorably to these other species (table 1; supplementary fig. 2*a* and *b*, Supplementary Material online). Genome size estimates have not been evaluated by methods such as flow cytometry for many *Phytophthora* species (Jung et al. 2017), and prior to this study there was no genome size estimate available for *P. plurivora*. From k-mer analysis of our sequence data, we calculated a genome size estimate of 45 Mb (supplementary fig. 4*a*, Supplementary Material online). We calculated the genome size of *P. infestans* and *P. multivora* isolate NZFS 3378 for comparison. The genome sizes of *P. infestans* and *P. multivora* NZFS 3378 were calculated to be 145 and 49 MB respectively. The lower than expected value for *P. infestans* is likely due to the high repeat content of that genome. Our calculations thus suggest that our genome assembly for *P. plurivora* spans over 90% of the genome size estimate, and that repetitive sequences are not as prevalent as in some other *Phytophthora* genomes, especially *P. infestans* (Tyler et al. 2006; Haas et al. 2009).

Table 2
Repeat Elements Found in the *P. plurivora* Genome

Repeats	Number	Length Occupied	Percentage of Sequence
SINEs:	52	6,478	0.02%
ALUs	0	0	0.00%
MIRS	0	0	0.00%
LINEs:	181	87,398	0.22%
LINE1	114	43,446	0.11%
LINE2	0	0	0.00%
L3/CR1	25	20,942	0.05%
LTR elements:	1,543	945,917	2.34%
ERVL	0	0	0.00%
ERVL-MaLRs	0	0	0.00%
ERVL-class I	0	0	0.00%
ERVL-class II	0	0	0.00%
DNA elements:	2,430	1,115,747	2.76%
hAT-charlie	0	0	0.00%
TcMar-Tigger	3	887	0.00%
Unclassified	1,427	826,032	2.04%
Total interspersed repeats:		2,981,572	7.37%
Satellites:	0	0	0.00%
Simple repeats	4,134	186,013	0.46%
Low complexity:	500	26,340	0.07%

Phytophthora species secrete effector proteins that act either in the apoplastic space or inside host cells to aid infection and/or elicit defense responses (reviewed in Schornack et al. 2009; Whisson et al. 2016). Within this latter effector class are the RXLR and CRN (crinkler) effector families, which are characterized by a signal peptide and conserved peptide motifs required for translocation into the host (Whisson et al. 2007; Schornack et al. 2010). SignalP v3.0 analysis predicted 1,737 secreted proteins of which 84 were predicted by our HMM as RXLR class effector proteins (see later section Supplementary material 3, Supplementary Material online) and 60 proteins grouped as predicted secreted CRN class effectors. Within the

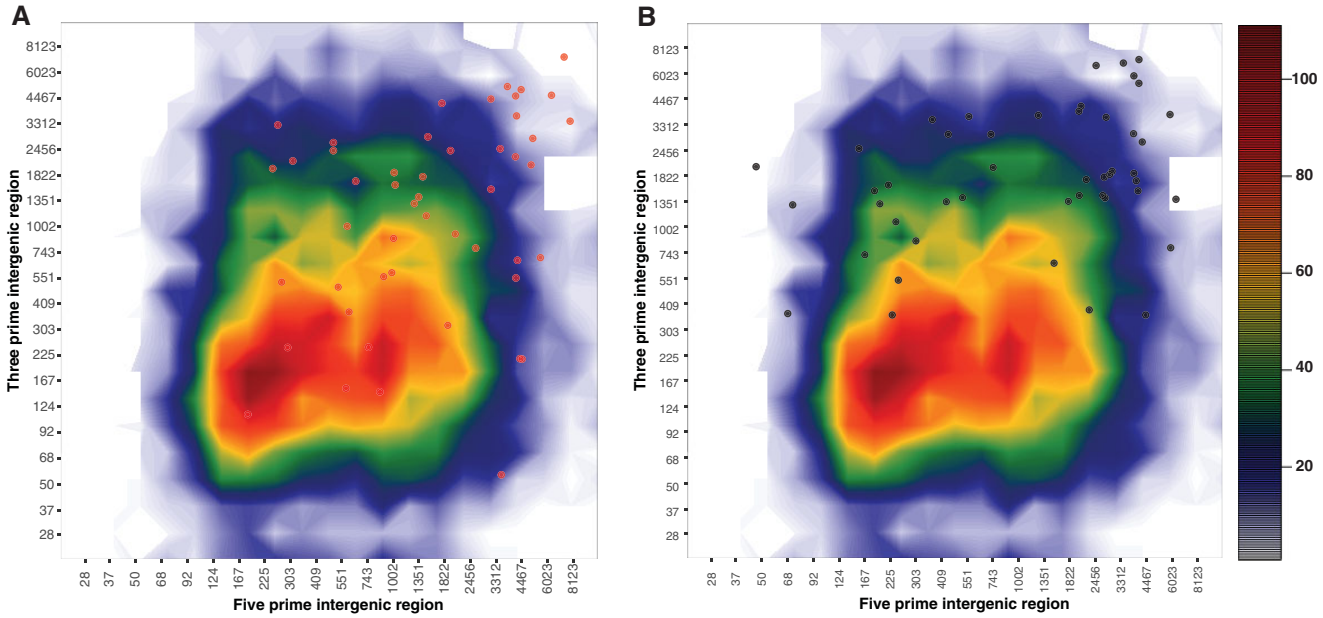

Fig. 1.—Plot showing the intergenic distance between genes in *P. plurivora*. (*A*) Red dots indicate the intergenic distances for RXLR effector-coding genes. (*B*) Black dots indicate the intergenic distances of CRN effector coding genes. The color scale representing the gene content per bin is shown to the right of the plots; values on x- and y-axes are nucleotides (nt). See supplementary material 4, Supplementary Material online and supplementary fig. 6, Supplementary Material online for data and boxplot analysis.

RXLR class, the following motifs were present; EER (80 proteins), WY (24 proteins), WYY (23 proteins), LYD (14 proteins) (Win et al. 2012; Ye et al. 2015).

Large numbers of CAZy proteins have been identified in other sequenced *Phytophthora* species (Ospina-Giraldo et al 2010; Grenville-Briggs et al. 2017). Here, we identified glycoside hydrolases (332), glycosyltransferases (271), carbohydrate binding modules (304), polysaccharide lyases (49), and carbohydrate esterases (43). The number of polysaccharide lyases predicted in *P. plurivora* is similar to that found in other *Phytophthora* species, but our primary analysis here shows elevated numbers of other CAZy proteins in *P. plurivora*, compared with other *Phytophthora* species (Ospina-Giraldo et al. 2010). Homologs were found in the PHI database for 2.6% of the total predicted *P. plurivora* proteome (308/11,749), predominantly with *P. sojae* and *Fusarium graminearum*. The secondary metabolite analysis on the *P. plurivora* genome revealed six genomic regions with potential to encode enzymes involved in secondary metabolite formation (supplementary table 1, Supplementary Material online). The six genomic regions comprise 56 genes that match with the antiSMASH 3.0 database (Weber et al. 2015). This included a non-ribosomal peptide synthase, and enzymes for ectoine and terpene biosynthesis. When compared with fungi, species of *Phytophthora* are not known to produce many secondary metabolites. The best characterized secondary metabolites from *Phytophthora* are the mating hormones α1 and α2, which are diterpene molecules (Tomura et al. 2017). It has also been shown that *Phytophthora* species produce a signal molecule derived from 4, 5-dihydroxy-2, 3-pentanedione which has quorum-sensing activity in bacteria (Kong et al. 2010). It is possible that the secondary metabolism predictions from the *P. plurivora* genome may be involved in the synthesis of these bioactive secondary metabolites.

Ploidy levels in *Phytophthora* species can be variable within and between species, but are at least diploid (Bertier et al. 2013; Li et al. 2017). Using ploidyNGS to analyze allele frequency distributions based on Kolmogorov-Smirnov distance, ploidyNGS suggested that the *P. plurivora* strain sequenced here was most likely tetraploid. *P. plurivora* is a homothallic (self-fertile, inbreeding) species, signifying that if its survival in the environment is via sexually derived oospores, then heterozygosity levels will be reduced with each generation, as has been observed in *P. plurivora* strains sampled from different countries (Schoebel et al. 2014). Despite this consideration, heterozygous loci with additional alleles were detected, suggestive of an elevated ploidy level (supplementary fig. 5, Supplementary Material online).

Phytophthora genomes often contain high levels of repetitive DNA sequences, such as *P. infestans* for which the genome contains over 75% repetitive sequences (Tyler et al. 2006; Haas et al. 2009). Approximately 15% of the *P. plurivora* genome is comprised of repetitive sequences, far less than many other *Phytophthora* genomes sequenced to date. The predominant repeat type is interspersed repeats, accounting for 50% of the total repeats. Repetitive DNA elements and long terminal repeat (LTR) retroelements comprise approximately 30% of the total repeats (table 2).

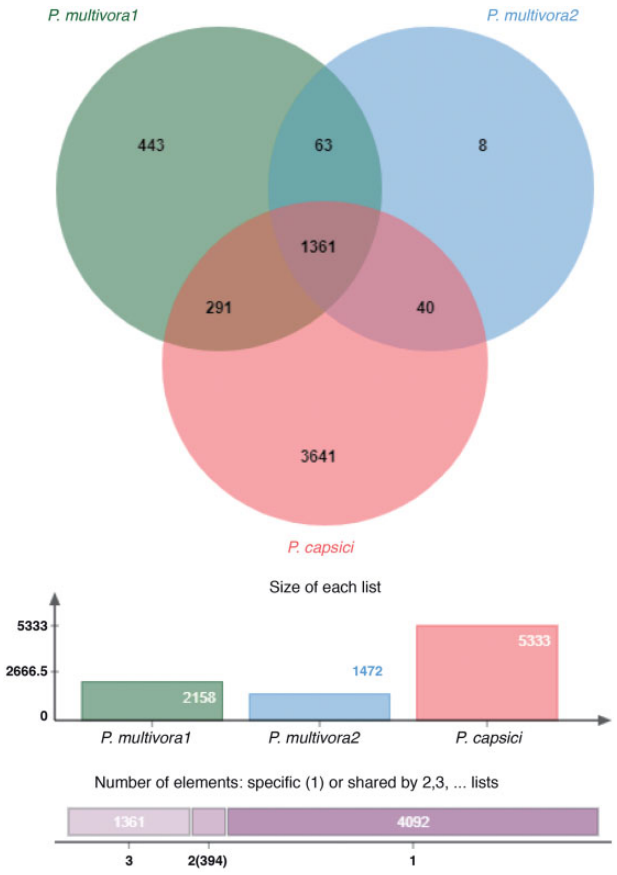

Fig. 2.—Venn diagram comparing *P. plurivora* genes not found in genome assemblies of *P. multivora* isolates and/or *P. capsici*. Bar graph below the Venn diagram shows the total number of genes specific to *P. plurivora* that were not found in the genome assemblies of *P. multivora* or *P. capsici*. See supplementary material 5, Supplementary Material online for lists of genes used for figure construction.

RXLR and CRN Effector Prediction

As in other *Phytophthora* species sequenced to date, the genome of *P. plurivora* contains many genes encoding RXLR class effectors. RXLR effectors have been extensively researched in *P. infestans* and *P. sojae* and found to be translocated inside plant cells during infection to facilitate the infection process (Whisson et al. 2016; Wang et al. 2017). These effectors are modular proteins that contain an N-terminal signal peptide, a conserved RXLR peptide motif typically within the next 40 amino acids, and often an EER motif near the RXLR. The functional effector peptide region is located between the RXLR–EER and the C-terminus. The majority of effectors in this class that have been functionally characterized contain all three of these features (Anderson et al. 2015; Whisson et al. 2016). We used four different methods to predict effectors from *Phytophthora* genomes: **A** is a modified prediction method described by Jiang et al. (2008); **B** is the effectorP program for fungal effector prediction, based on machine learning (Sperschneider et al. 2016) and **C** (Win et al. 2007) and **D** (Whisson et al. 2007) are older prediction methods using a Galaxy pipeline (supplementary fig. 1*a*, Supplementary Material online). The four different RXLR effector prediction pipelines yielded differing numbers of candidates (supplementary material 4, Supplementary Material online). We compared the RXLR effectors predicted by each of four pipelines and plotted a Venn diagram (supplementary fig. 1*b*, Supplementary Material online). Pipeline C (Win et al. 2007) produced the largest number of RXLR effectors (196), but half of the effectors predicted by this pipeline did not possess the canonical EER motif. By comparison, the HMM-based prediction methods predicted fewer candidate RXLR effectors. Pipeline A, with additional filtering steps, was the most stringent, yielding 84 candidate effectors, of which 80 contain the signal peptide, RXLR and EER motifs. As such, we regarded this as a 'high confidence' set of predicted effectors for further analysis. We predicted CRN effectors as described (Yin et al. 2017) by using CRN sequences from *P. infestans* as training material (Haas et al. 2009). We predicted 139 CRN proteins from *P. plurivora*, of which 60 had signal peptides for secretion, and these can be considered as candidate effectors. A large proportion of CRN proteins identified from other *Phytophthora* genomes, such as *P. infestans* and *P. capsici*, also do not possess a predicted signal peptide (Haas et al. 2009; Stam et al. 2013).

Effector-Coding Genes are Located in Gene-Sparse Regions of the *P. plurivora* Genome

Comparison of RXLR class effectors between *Phytophthora* species from different clades has typically revealed that these effectors have diverse sequences, with many having no homologs in other species (Quinn et al. 2013; McGowan and Fitzpatrick 2017), and thus are evolving rapidly. Comparisons of effectors from more closely related species can reveal the genomic processes acting on effector coding genes to drive their diversification. It has been proposed that *Phytophthora* species have two-speed genomes, where effector coding genes reside in gene-poor, repeat-rich regions that are prone to rapid evolution (Dong et al. 2015). The RXLR and CRN effectors in *P. plurivora* also reside in more gene-poor regions of the genome, as evidenced by greater intergenic distances between them and neighboring genes (fig. 1*A* and *B*). Specifically, the mean intergenic distance between all *P. plurivora* genes at the 5′ end is 1,107 bp whereas this distance is 3,117 bp for RXLR effector coding genes and 2,018 bp for CRN effector coding genes. At the 3′ end, the mean intergenic distance for the entire predicted gene set is 848 bp, whereas it is 2,128 bp for RXLR effector genes and 3,087 bp for CRN effector genes. The median intergenic distance between all predicted genes was significantly less than that for the effector gene classes analyzed. The median 5′ end intergenic distance between all predicted genes was 582 bp, whereas it was 1,920 and

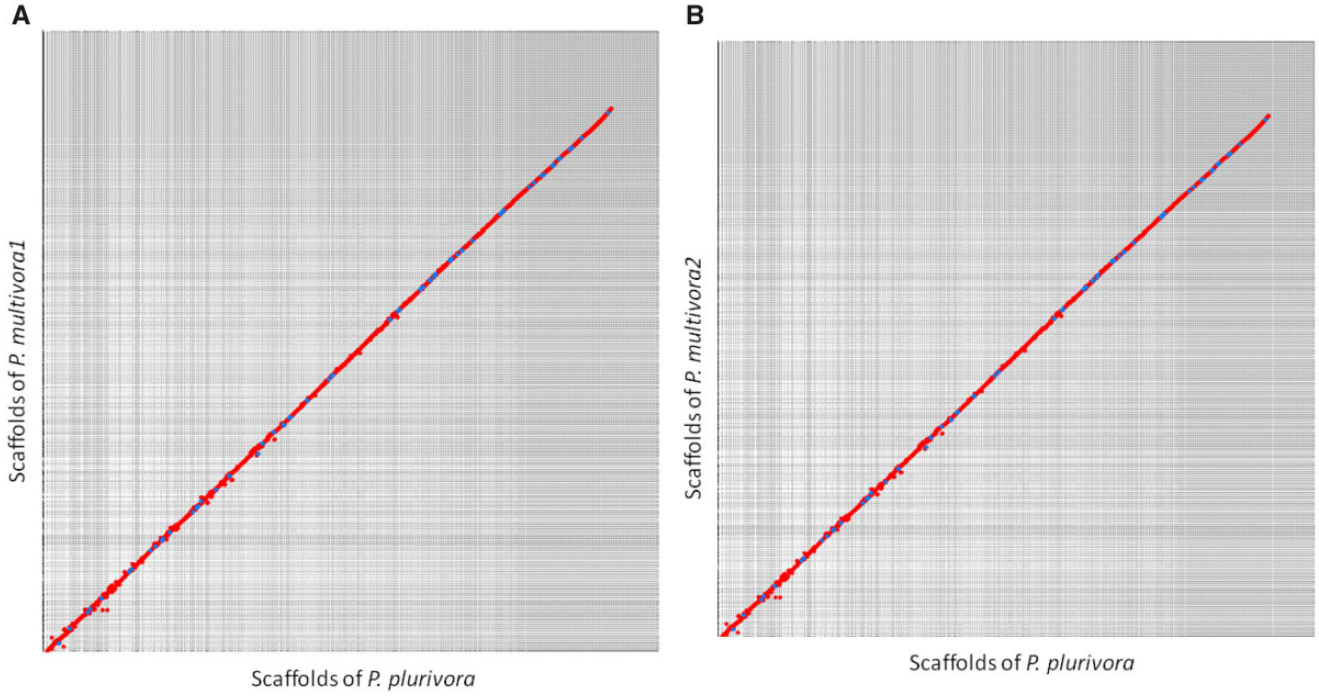

Fig. 3.—Collinearity between *P. plurivora* and *P. multivora* genomes. (*A*) Whole genome alignment (Mummer) between *P. plurivora* and *P. multivora* isolate 1 (NZFS 3378). (*B*) Whole genome alignment (Mummer) between *P. plurivora* and *P. multivora* isolate 2 (NZFS 3448). The X-axis is the reference genome of *P. plurivora* and the Y-axis is *P. multivora*. The blue dots represent reverse complement matches and red represents forward matches.

1,518 bp for the RXLR and CRN effector genes, respectively. Similarly, the median 3′ prime intergenic distance between all the genes was 316 bp, whereas it was 1,414 and 1,925 bp for RXLR and CRN effectors, respectively. A *t*-test on flanking intergenic distances of effectors and core genes showed that intergenic distances were significantly different from the core genes and effector genes (supplementary material 5, Supplementary Material online and supplementary fig. 6, Supplementary Material online). These results for *P. plurivora* follow a similar trend observed for other *Phytophthora* genomes (Dong et al. 2015).

Synteny Analysis

Draft genome sequences for two *P. multivora* isolates from Studholme et al. (2016) allowed us to compare them with our *P. plurivora* draft assembly. Genome assemblies of *P. plurivora* and both isolates of *P. multivora* are of similar size, at approximately 40 Mb each. Comparing the genes from *P. plurivora* with those from other Clade 2 species, the highest level of similarity was observed between *P. plurivora* and *P. multivora*. There were 2,158 and 1,472 *P. plurivora* genes without orthologs in *P. multivora* NZFS3378 and NZFS3348, respectively, of which 1,361 genes were common between NZFS3378 and NZFS3348 (fig. 2; supplementary material 6, Supplementary Material online).

Genome collinearity was studied between *P. plurivora* and both isolates of *P. multivora*, and *P. capsici*. The largest scaffold (scaffold_1 length of 294,496 bp containing 95 protein-coding genes [GenBank NMPK01000001.1]) was almost fully collinear with both *P. multivora* isolates, except for three missing genes: g65, g72, and g95. These three genes are highly conserved in the two *P. multivora* isolates and are also highly conserved in other *Phytophthora sp*. Blocks of genome collinearity have been identified between diverse species such as *P. infestans*, *P. sojae*, and *P. ramorum* previously (Haas et al. 2009). *P. plurivora* and *P. multivora* are more closely related than the species in these previous comparisons (Yang et al. 2017), and so it is unsurprising that most of their genomes exhibit strong collinearity (fig. 3*A* and *B*).

Through comparing *P. plurivora* and *P. multivora*, we identified evidence of localized gene duplication and sequence diversification in a cluster of RXLR effectors. *P. plurivora* Scaffold_267 contains seven RXLRs in its entire length of 45 kb, and we identified two scaffolds (LGSM010000246.1 and LGSM01000099.1) in *P. multivora* NZFS 3378 (isolate 1) and two scaffolds in *P. multivora* NZFS 3448 (isolate 2) (LGSL01000255.1 and LGSL01000075.1) having collinearity with it. All seven RXLR effectors found in this *P. plurivora* scaffold had collinear homologs in the *P. multivora* scaffolds; however, one RXLR of *P. plurivora* PlRXLR 39, is duplicated in *P. multivora* NZFS 3378 (isolate 1) and one RXLR of

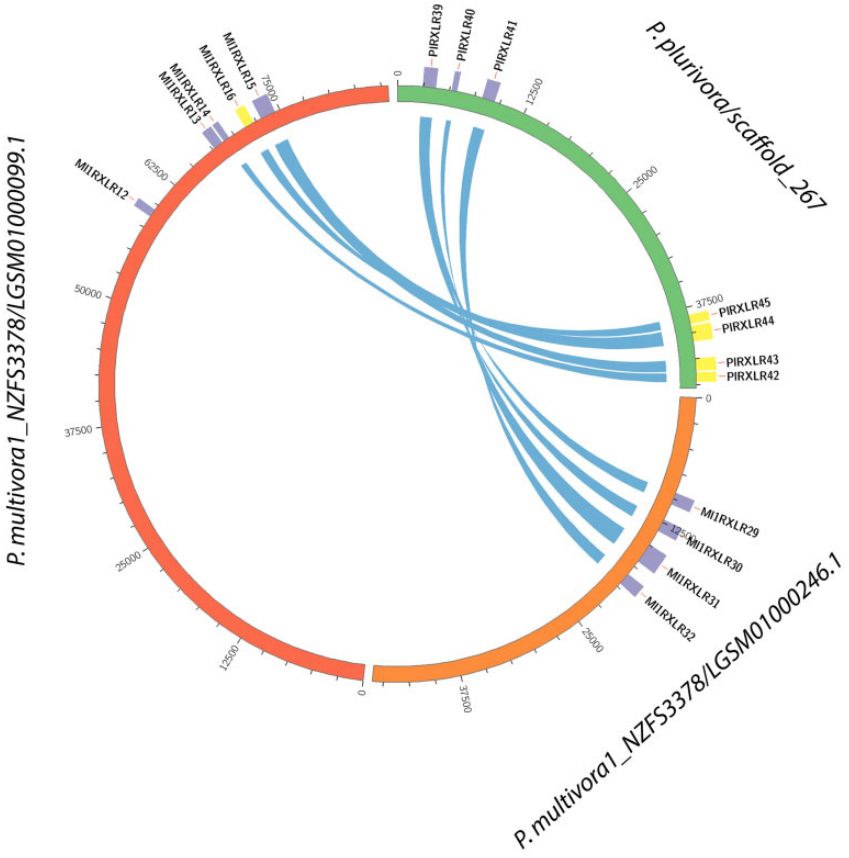

Fig. 4.—Scaffold_267 of *P. plurivora* containing seven RXLR-coding genes in its entire length of 45 kb has collinearity with two scaffolds (LGSM010000246.1 and LGSM01000099.1) in *P. multivora* NZFS 3378 (isolate 1) and two scaffolds in *P. multivora* NZFS 3448 (isolate 2) (LGSL01000255.1 and LGSL01000075.1 [not shown]). *P. plurivora* PlRXLR39 has undergone duplication in *P. multivora* into MIRXLR29 and MIRXLR30. The purple rectangles in the image indicate the plus strand and the yellow rectangles indicate the negative strand.

P. multivora NZFS 3378 (isolate 1), MIIRXLR15 is duplicated into two RXLRs in *P. plurivora* e.g., PlRXLR43 and PlRXLR44 (fig. 4).

Analysis of Synonymous and Non-synonymous Codon Substitution of *P. plurivora* RXLR Effectors Reveals a Subset under Neutral or Purifying Selection

To determine whether RXLR effector sequence diversification was selection-neutral or under positive selection, we compared the *P. plurivora* RXLR effector complement with those predicted from two isolates of *P. multivora*, *P. capsici, P. cinnamomi*, and *P. ramorum*. All six species are considered to be plant pathogens with a broad host range. *P. multivora* is closely related to *P. plurivora*, while *P. capsici* is more distantly related within Clade 2 (Yang et al. 2017). *P. cinnamomi* and *P. ramorum* are both pathogens of trees and are placed in Clade 7c and Clade 8c, respectively (Yang et al. 2017). Using pipeline A, we predicted effectors of two *P. multivora* isolates NZFS 3378 (isolate 1) and NZFS 3448 (isolate 2) (Studholme et al. 2016), *P. capsici* (Lamour et al. 2012), *P. cinnamomi* (Studholme et al. 2016) and *P. ramorum* (Jiang et al. 2008). For *P. capsici* 140 RXLR effectors were predicted, whereas for *P. multivora* there were 84 and 92 predicted for NZFS 3378 and NZFS 3448, respectively.

OrthoMCL analysis (default BLASTP parameters, 1.5 inflation value) identified 105 clusters that had at least three members in a group. A total of 48 groups contained members from a single species, 42 groups contained members from two species, and 15 groups contained members from three species. No clusters comprised RXLRs from more than three *Phytophthora* species. *P. plurivora* effectors were represented within 47 clusters, *P. multivora* was represented in 53 clusters and *P. capsici* within 33 clusters having three or more members. dN/dS ratios were calculated for all clusters containing at least three members (supplementary material 7, Supplementary Material online), as a previous study had shown that RXLR effector paralogs within a species could be under positive selection (Win et al. 2007). Values for dN/dS >1.0 are suggestive of positive selection, while values below 1.0 are suggestive of purifying selection. The 105

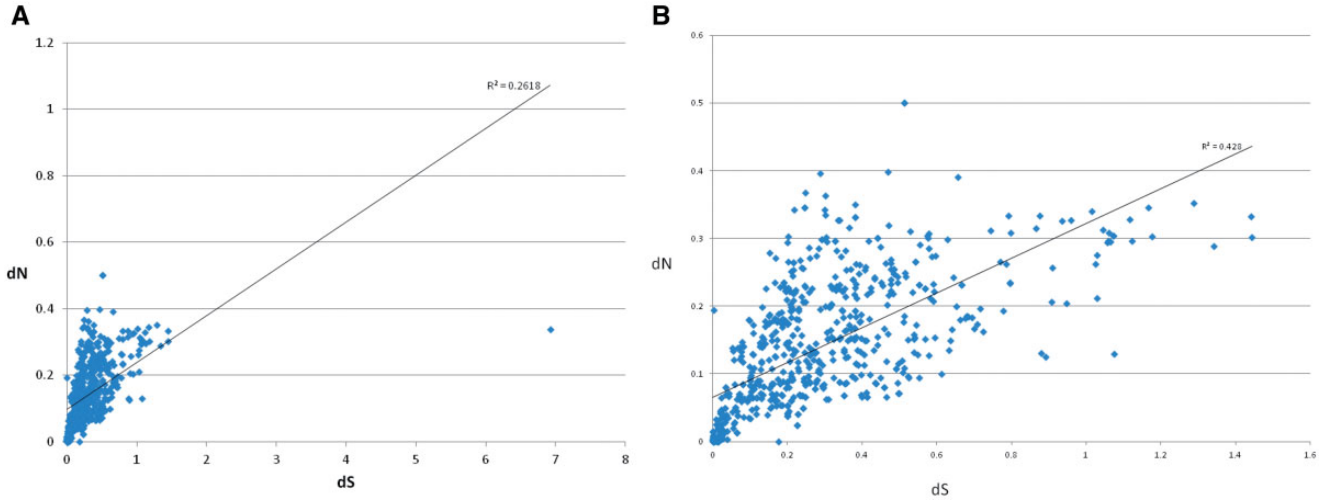

Fig. 5.—dN/dS analysis of RXLR effector coding genes from *P. plurivora, P. capsici, P. ramorum, P. multivora* isolates, *and P.cinnamomi*. (*A*) Scatter plot of all dN and dS values for RXLR effector coding genes from *P. plurivora, P. capsici, P. ramorum, P. multivora* isolates, *and P.cinnamomi*. (*B*) Scatter plot of dN and dS values after removal of the outlier pair PrAvh147 and PrAvh227 from *P. ramorum*. See supplementary material 6, Supplementary Material online for dN and dS values used for figure construction.

clusters resolved into 432 pairs of proteins. The higher dN/dS ratios ranged from 1.0 to 3.6 and included five effectors from *P. plurivora*, but only PlRXLR53 exhibited a markedly elevated dN/dS ratio (1.9) that suggested positive selection (fig. 5*A* and *B*).

Oomycete RXLR effectors are considered to be rapidly evolving and under positive selection, as homologs are often not found in species in different clades of *Phytophthora* (Quinn et al. 2013; McGowan and Fitzpatrick 2017). It was therefore unexpected that only PlRXLR53, grouped with Ml1RXLR62, Ml2RXLR50 and Ml2RXLR83 from *P. multivora*, showed a dN/dS ratio (1.9) that suggested positive selection. All other *P. plurivora* effectors in OrthoMCL groups of three or more effector genes were either not under selection (neutral), or under purifying selection. By comparison, effectors from *P. ramorum* showed higher dN/dS ratios (supplementary material 7, Supplementary Material online) as shown previously (Win et al. 2007). Seven *P. plurivora* RXLR effectors had no homologs in any of 26 sequenced *Phytophthora* genomes: PlRXLR4, PlRXLR6, PlRXLR18, PlRXLR27, PlRXLR36, PlRXLR52, PlRXLR58. Taken together, these results suggest that *P. plurivora* possesses RXLR effectors that are under diverse evolutionary pressures, with one subset showing evidence of purifying selection, and a further subset that have evolved rapidly and are specific only to *P. plurivora*. Sequencing of additional Clade 2 species, and species that are more closely related to *P. plurivora*, may provide more resolution in clarifying the mode of selection acting on this class of effectors. We have plotted multiple sequence alignments of PlRXLRs having the highest dN/dS ratios from group12 (PlRXLR9), group14 (PlRXLR20, 32, 76, 81) and group39 (PlRXLR53) in supplementary fig. 7*a–c*, Supplementary Material online.

Conclusion

The genome sequence presented here provides a resource that can underpin further investigation into the mechanisms of disease caused by *P. plurivora*, a prevalent but little researched pathogen of important tree species. Our genome sequence of *P. plurivora* is consistent with the genome architecture of other sequenced *Phytophthora* species, and we found evidence for elevated ploidy, as can occur in *Phytophthora* species. This genome resource can be used in future population genomic studies for identification of haplotypes and alleles, and in identifying which effectors may function in infection of woody host plants.

Supplementary Material

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

Acknowledgments

We are grateful to Laura Vetukuri for collecting and isolating *P. plurivora* strain AV1007. This work is supported by research funds from Swedish Research Council Formas (2015-430), the Swedish Foundation for Strategic Research (FFL5), Helge Ax: son Johnsons Stiftelse, Nordic Joint Committee for Agricultural and Food Research (NKJ), Nordic Forest Research (SNS) network and the Scottish Government Rural and Environment Science and Analytical Services Division (RESAS). Funding support to ST through a DBT-Ramalingaswamy fellowship is gratefully acknowledged. MMC was supported by a DST-INSPIRE AORC fellowship. The authors acknowledge support from the

National Genomics Infrastructure in Stockholm funded by Science for Life Laboratory, the Knut and Alice Wallenberg Foundation, Partnership Alnarp, Parvatha Vardhini foundation and the Swedish Research Council, and SNIC/Uppsala Multidisciplinary Center for Advanced Computational Science for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure.

Literature Cited

Afgan E, et al. 2016. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res. 44(W1):W3–W10.

Ah-Fong AM, Kim KS, Judelson HS. 2017. RNA-seq of life stages of the oomycete *Phytophthora infestans* reveals dynamic changes in metabolic, signal transduction, and pathogenesis genes and a major role for calcium signaling in development. BMC Genomics 18(1):198.

Anderson RG, Deb D, Fedkenheuer K, McDowell JM. 2015. Recent progress in RXLR effector research. Mol Plant-Microbe Interact. 28(10):1063–1072.

Bankevich A, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 19(5):455–477.

Bertier L, Leus L, D'hondt L, de Cock AWAM, Höfte M. 2013. Host adaptation and speciation through hybridization and polyploidy in *Phytophthora*. PLOS ONE 8(12):e85385.

Blin K, et al. 2017. antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification. Nucleic Acids Res. 45(W1):W36–W41.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15):2114–2120.

Cavalier-Smith T. 2018. Kingdom Chromista and its eight phyla: a new synthesis emphasising periplastid protein targeting, cytoskeletal and periplastid evolution, and ancient divergences. Protoplasma 255(1):297–357.

Chen XR, Xing YP, Li YP, Tong YH, Xu JY. 2013. RNA-Seq reveals infection-related gene expression changes in *Phytophthora capsici*. PLOS ONE 8(9):e74588.

Cleary MR, Blomquist M, Vetukuri RR, Böhlenius H, Witzell J. 2017. Susceptibility of common tree species in Sweden to *Phytophthora cactorum*, *P. cambivora* and *P. plurivora*. Forest Pathol. 47(3):e12329.

Cock PJA, Grüning BA, Paszkiewicz K, Pritchard L. 2013. Galaxy tools and workflows for sequence analysis with applications in molecular plant pathology. PeerJ. 1:e167.

Corrêa dos Santos R, Goldman GH, Riaño-Pachón DM. 2017. ploidyNGS: visually exploring ploidy with Next Generation Sequencing data. Bioinformatics 33(16):2575–2576.

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

Erwin DC, Ribeiro OK. 1996. Phytophthora Diseases Worldwide. St. Paul (MN): APS Press.

Grenville-Briggs LJ, et al. 2017. Draft genome of the oomycete pathogen *Phytophthora cactorum* strain LV007 isolated from European beech (*Fagus sylvatica*). Genomics Data 12:155–156.

Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29(8):1072–1075.

Haas BJ, et al. 2009. Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. Nature 461(7262):393–398.

Hong CX, et al. 2009. The avocado subgroup of *Phytophthora citricola* constitutes a distinct species, *Phytophthora mengei* sp. nov. Mycologia 101(6):833–840.

Jiang RHY, Tripathy S, Govers F, Tyler BM. 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 USA. 105(12):4874–4879.

Jones P, et al. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30(9):1236–1240.

Jung T, Blaschke H, Oßwald W. 2000. Involvement of soilborne *Phytophthora* species in Central European oak decline and the effect of site factors on the disease. Plant Pathol. 49(6):706–718.

Jung T, et al. 2005. Involvement of *Phytophthora* species in the decline of European beech in Europe and the USA. Mycologist 19(4):159–166.

Jung T. 2009. Beech decline in Central Europe driven by the interaction between *Phytophthora* infections and climatic extremes. Forest Pathol. 39(2):73–94.

Jung T, Burgess TI. 2009. Re-evaluation of *Phytophthora citricola* isolates from multiple woody hosts in Europe and North America reveals a new species, *Phytophthora plurivora* sp. nov. Persoonia 22(1):95–110.

Jung T, et al. 2017. Six new *Phytophthora* species from ITS Clade 7a including two sexually functional heterothallic hybrid species detected in natural ecosystems in Taiwan. Persoonia 38(1):100–135.

Kamoun S, et al. 2015. The Top 10 oomycete pathogens in molecular plant pathology. Mol Plant Pathol. 16(4):413–434.

Kong P, Lee BWK, Zhou ZS, Hong C. 2010. Zoosporic plant pathogens produce bacterial autoinducer-2 that affects *Vibrio harveyi* quorum sensing. FEMS Microbiol Lett. 303(1):55–60.

Kurtz S, et al. 2004. Versatile and open software for comparing large genomes. Genome Biol. 5:R12.

Lamour KH, et al. 2012. Genome sequencing and mapping reveal loss of heterozygosity as a mechanism for rapid adaptation in the vegetable pathogen *Phytophthora capsici*. Mol Plant-Microbe Interact. 25(10):1350–1360.

Li L, Stoeckert CJ, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res 13(9):2178–2189.

Li Y, et al. 2017. Changing ploidy as a strategy: the Irish potato famine pathogen shifts ploidy in relation to its sexuality. Mol Plant-Microbe Interact. 30(1):45–52.

Löbmann MT, et al. 2016. The occurrence of pathogen suppressive soils in Sweden in relation to soil biota, soil properties, and farming practices. Appl Soil Ecol. 107:57–65.

Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 42(D1):D490–D495.

Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 27(6):764–770.

McGowan J, Fitzpatrick DA. 2017. Genomic, network, and phylogenetic analysis of the oomycete effector arsenal. mSphere 2(6):e00408–17.

Orlikowski LB, et al. 2011. *Phytophthora* root and collar rot of mature *Fraxinus excelsior* in forest stands in Poland and Denmark. Forest Pathol. 41(6):510–519.

Ospina-Giraldo MD, Griffith JG, Laird EW, Mingora C. 2010. The CAZyome of *Phytophthora* spp.: a comprehensive analysis of the gene complement coding for carbohydrate-active enzymes in species of the genus *Phytophthora*. BMC Genomics 11(1):525.

Panda A, et al. 2018. EumicrobeDBLite: a lightweight genomic resource and analytic platform for draft oomycete genomes. Mol Plant Pathol. 19(1):227–237.

Quinn L, et al. 2013. Genome-wide sequencing of *Phytophthora lateralis* reveals genetic variation among isolates from Lawson cypress (*Chamaecyparis lawsoniana*) in Northern Ireland. FEMS Microbiol Lett. 344(2):179–185.

Saunders DGO, Win J, Kamoun S, Raffaele S. 2014. Two-dimensional data binning for the analysis of genome architecture in filamentous plant pathogens and other eukaryotes. In: Birch P, Jones JT, Bos JIB, editors.

Plant-pathogen interactions: methods and protocols. Totowa, NJ: Humana Press. p. 29–51.
Schell T, et al. 2017. An annotated draft genome for *Radix auricularia* (Gastropoda, Mollusca). Genome Biol Evol. 9(3):585–592.
Schoebel CN, Stewart J, Gruenwald NJ, Rigling D, Prospero S. 2014. Population history and pathways of spread of the plant pathogen *Phytophthora plurivora*. PLOS ONE 9(1):e85368.
Schornack S, et al. 2009. Ten things to know about oomycete effectors. Mol Plant Pathol. 10(6):795–803.
Schornack S, et al. 2010. Ancient class of translocated oomycete effectors targets the host nucleus. Proc Natl Acad Sci USA. 107(40):17421–17426.
Scott PM, et al. 2009. *Phytophthora multivora* sp. nov., a new species recovered from declining *Eucalyptus*, *Banksia, Agonis* and other plant species in Western Australia. Persoonia 22:1–13.
Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31(19):3210–3212.
Smit AFA, Hubley R, Green P. RepeatMasker Open-4.0. 2013-2015. Available from: http://www.repeatmasker.org; last accessed August 23, 2018.
Sperschneider J, et al. 2016. EffectorP: predicting fungal effector proteins from secretomes using machine learning. New Phytologist 210(2):743–761.
Stam R, et al. 2013. Identification and characterization of CRN effectors in *Phytophthora capsici* shows modularity and functional diversity. PLOS ONE 8(3):e59517.
Stanke M, Steinkamp R, Waack S, Morgenstern B. 2004. AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic Acids Res. 32(Web Server):W309–W312.
Studholme DJ, et al. 2016. Genome sequences of six *Phytophthora* species associated with forests in New Zealand. Genomics Data 7:54–56.
Trapnell C, et al. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc. 7(3):562–578.
Tomura T, Molli SD, Murata R, Ojika M. 2017. Universality of the *Phytophthora* mating hormones and diversity of their production profile. Sci Rep. 7(1):5007.
Tyler BM. 2006. *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. Science 313(5791):1261–1266.
Urban M, et al. 2015. The Pathogen-Host Interactions database (PHI-base): additions and future developments. Nucleic Acids Res. 43(D1):D645–D655.
Wang S, et al. 2017. Delivery of cytoplasmic and apoplastic effectors from *Phytophthora infestans* haustoria by distinct secretion pathways. New Phytologist. 216(1):205–215.
Weber T, et al. 2015. antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Res. 43(W1):W237–W243.
Whisson SC, et al. 2007. A translocation signal for delivery of oomycete effector proteins into host plant cells. Nature 450(7166):115–118.
Whisson SC, Boevink PC, Wang S, Birch PRJ. 2016. The cell biology of late blight disease. Curr Opin Microbiol. 34:127–135.
Win J, et al. 2012. Sequence divergent RXLR effectors share a structural fold conserved across plant pathogenic oomycete species. PLOS Pathogens 8(1):e1002400.
Win J, et al. 2007. Adaptive evolution has targeted the C-terminal domain of the RXLR effectors of plant pathogenic oomycetes. Plant Cell 19(8):2349–2369.
Yang X, Tyler BM, Hong C. 2017. An expanded phylogeny for the genus *Phytophthora*. IMA Fungus 8(2):355–384.
Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24(8):1586–1591.
Ye W, Wang Y, Wang Y. 2015. Bioinformatics analysis reveals abundant short alpha-helices as a common structural feature of oomycete RXLR effector proteins. PLOS ONE 10(8):e0135240.
Yin L, et al. 2017. Genome sequence of *Plasmopara viticola* and insight into the pathogenic mechanism. Sci Rep. 7(1):46553.

Associate editor: Richard Cordaux