MPMI Vol. 33, No. 5, 2020, pp. 742–753. https://doi.org/10.1094/MPMI-07-19-0211-R

e-**X**tra*

The Genome of *Peronospora belbahrii* Reveals High Heterozygosity, a Low Number of Canonical Effectors, and TC-Rich Promoters

Marco Thines,[1,2,3,†] Rahul Sharma,[1,2,3] Sander Y. A. Rodenburg,[4,5] Anna Gogleva,[6] Howard S. Judelson,[7] Xiaojuan Xia,[1,2] Johan van den Hoogen,[4] Miloslav Kitner,[8] Joël Klein,[9] Manon Neilen,[9] Dick de Ridder,[5] Michael F. Seidl,[4] Guido van den Ackerveken,[9] Francine Govers,[4] Sebastian Schornack,[6] and David J. Studholme[10]

[1] Institute of Ecology, Evolution and Diversity, Goethe University, Max-von-Laue-Str. 9, 60323 Frankfurt (Main), Germany
[2] Senckenberg Gesellschaft für Naturforschung, Senckenberganlage 25, 60325 Frankfurt (Main), Germany
[3] Integrative Fungal Research (IPF) and Translational Biodiversity Genomics (TBG), Georg-Voigt-Str. 14-16, 60325 Frankfurt (Main), Germany
[4] Laboratory of Phytopathology, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands
[5] Bioinformatics Group, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands
[6] University of Cambridge, Sainsbury Laboratory, 47 Bateman Street, Cambridge, CB2 1LR, U.K.
[7] Department of Microbiology and Plant Pathology, University of California, Riverside, CA 92521 U.S.A.
[8] Department of Botany, Faculty of Science, Palacký University Olomouc, Šlechtitelů 27, 78371 Olomouc, Czech Republic
[9] Plant-Microbe Interactions, Department of Biology, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands
[10] Biosciences, College of Life and Environmental Sciences, University of Exeter, Stocker Road, Exeter EX4 4QD, U.K.

Accepted 23 February 2020.

Along with *Plasmopara destructor*, *Peronosopora belbahrii* has arguably been the economically most important newly emerging downy mildew pathogen of the past two decades. Originating from Africa, it has started devastating basil production throughout the world, most likely due to the distribution of infested seed material. Here, we present the genome of this pathogen and results from comparisons of its genomic features to other oomycetes. The assembly of the nuclear genome was around 35.4 Mbp in length, with an N_{50} scaffold length of around 248 kbp and an L_{50} scaffold count of 46. The circular mitochondrial genome consisted of around 40.1 kbp. From the repeat-masked genome, 9,049 protein-coding genes were predicted, out of which 335 were predicted to have extracellular functions, representing the smallest secretome so far found in peronosporalean oomycetes. About 16% of the genome consists of repetitive sequences, and, based on simple sequence repeat regions, we provide a set of microsatellites that could be used for population genetic studies of *P. belbahrii*. *P. belbahrii* has undergone a high degree of convergent evolution with other obligate parasitic pathogen groups, reflecting its obligate biotrophic lifestyle. Features of its secretome, signaling networks, and promoters are presented, and some patterns are hypothesized to reflect the high degree of host specificity in *Peronospora* species. In addition, we suggest the presence of additional virulence factors apart from classical effector classes that are promising candidates for future functional studies.

Keywords: comparative genomics, downy mildew, evolutionary biology, metabolic pathways, oomycetes

Oomycetes comprise more than 2,000 species with a broad array of lifestyles. They are found in almost all ecosystems, efficiently colonizing decaying organic matter as well as living hosts and play important ecological roles as saprotrophs and

The sequence data and annotations of this study have been deposited at ENA under the BioProject accession numbers PRJEB15119 and PRJEB20871. The assembly scaffolds have been accessioned in ENA with the numbers CACTHD010000001 to CACTHD010002780. The Kyoto Encyclopedia of Genes and Genomes orthology hidden Markov models are deposited at the NARCIS data repository under doi:10.4121/uuid:9b6e6aa0-b815-409f-9e96-04828b03290b. The mitochondrial genome was deposited in GenBank under the accession number MN530986.

Current address for Johan van den Hoogen: Global Ecosystem Ecology, Institute of Integrative Biology, ETH Zürich, 8092 Zürich, Switzerland.

Current address for Michael F. Seidl: Theoretical Biology & Bioinformatics, Department of Biology, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands.

†Corresponding author: M. Thines; m.thines@thines-lab.eu

Funding: This work was supported by the research funding program LOEWE "Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz" of Hesse's Ministry of Higher Education, Research, and the Arts in the framework of LOEWE IPF and the LOEWE Center for Translational Biodiversity Genomics (TBG). Further funding was received by the Food-for-Thought campaign from the Wageningen University Fund, the Research Council Earth and Life Sciences of the Netherlands Organization of Scientific Research in the framework of a VENI grant (863.15.005), and a JSTP grant (833.13.002), the National Institute of Food and Agriculture and National Science Foundation of the United States (the Gatsby Charitable Foundation [GAT3395/GLD]), and the Royal Society (UF160413).

*The ***e*-X***tra logo stands for "electronic extra" and indicates that six supplementary figures and 14 supplementary tables are published online.

The author(s) declare no conflict of interest.

pathogens (Judelson 2012; Thines 2014). Many oomycete species are responsible for huge economic losses and pose a major threat to food security. Unlike fungi, oomycetes are diploid for most of their life cycle and, therefore, have the potential to propagate as heterozygous clones.

Most known oomycete species cause downy mildew and are obligate biotrophic. Among this group of more than 800 species, about 500 belong to the genus *Peronospora* (Thines and Choi 2016; Thines 2014). *Peronospora belbahrii* is a highly specialized species causing basil downy mildew disease (Thines et al. 2009). Since its introduction, probably from Africa, it has emerged as a major threat to production of the culinary herb basil (*Ocimum basilicum*) (Thines and Choi 2016; Thines et al. 2009), paralleled in its huge recent impact only by *Plasmopara destructor*, the downy mildew of ornamental balsamine (Görg et al. 2017).

The lifestyle of pathogens is reflected by its primary and secondary metabolism. Obligate biotrophs fully rely on their host for survival and utilize metabolites provided by their hosts (Judelson 2017). Therefore, obligate biotrophs have reduced sets of enzyme-encoding genes required for metabolic pathways (Baxter et al. 2010; Kemen et al. 2011; Sharma et al. 2015a and b; Spanu 2012). For example, in addition to the widespread auxotrophy for thiamine and sterol in oomycetes (Dahlin et al. 2017; Judelson 2017), obligate biotrophic oomycetes have lost the ability to reduce inorganic nitrogen and sulfur (Baxter et al. 2010; Kemen et al. 2011; Sharma et al. 2015a and b; Spanu 2012).

The first line of pathogen attack relies on hydrolytic enzymes such as cutinases, glycosyl hydrolases, lipases, and proteases that degrade plant cell-wall components to promote pathogen entry (Blackman et al. 2015). Besides these enzymes, oomycetes secrete effectors that target host processes (Schornack et al. 2009). After germination and penetration, downy mildews are nutritionally dependent on living host cells, from which they extract nutrients via haustoria, which are believed to represent the main interaction interface between host and pathogen (Ellis et al. 2007; Schornack et al. 2009; Soylu and Soylu 2003). Extracellular proteins secreted from hyphae and haustoria play key roles in establishing stable interactions by modulating plant immune responses and plant metabolism (Meijer et al. 2014). While the biological function of many oomycete effectors still remains to be resolved, they have been classified based on sequence similarity and other recognizable sequence features, such as the WY domain (Boutemy et al. 2011) in host-translocated (cytoplasmic) effectors and apoplastic effectors. The latter include protease inhibitors, small cysteine-rich proteins (SCPs), elicitins, and necrosis-inducing-like protein (NLPs). NLPs are widely distributed across plant-pathogenic oomycetes and share a highly conserved NPP1 domain (Feng et al. 2014; Seidl and Van den Ackerveken 2019). Most cytoplasmic effectors carry short motifs associated with their translocation, such as the characteristic RxLR effector motif that often co-occurs with a downstream EER motif (Dou et al. 2008; Tyler et al. 2013).

The availability of next-generation sequencing enables economical and facile analysis of genomes, leading to many important insights into the biology of many oomycetes (Ali et al. 2017; Baxter et al. 2010; Gaulin et al. 2018; Grünwald 2012; Haas et al. 2009; Jiang et al. 2013; Judelson 2012; Lamour et al. 2012; Lévesque et al. 2010; Links et al. 2011; McCarthy and Fitzpatrick 2017; Raffaele and Kamoun 2012; Schena et al. 2008; Soanes et al. 2007; Sharma et al. 2015; Tian et al. 2011; Tyler et al. 2006). Because of their economic importance, some downy mildews have had their genomes sequenced, including *Bremia lactucae*, *Hyaloperonospora arabidopsidis*, *Peronospora effusa*, *Peronospora tabacina*, *Plasmopara viticola*, *Plasmopara halstedii*, *Plasmopara muralis*, *Plasmopara destructor* (filed as *Plasmopara obducens*), *Pseudoperonospora cubensis*, *Pseudoperonospora humuli*, and *Sclerospora graminicola* (Baxter et al. 2010; Derevnina et al. 2015; Feng et al. 2018; Fletcher et al. 2018, 2019; Kobayashi et al. 2017; Nayaka et al. 2017; Sharma et al. 2015; Tian et al. 2011; Ye et al. 2016). Here, we augment this repertoire with the annotated genome sequence of *P. belbahrii*. Using comparative genomics, we uncover evidence of convergent evolution in loss of metabolic function between this species and phylogenetically distant obligate pathogens, we identify features of its signaling networks and secretome, and we identify sequence features of its promoters. Finally, we present a set of microsatellite repeats that will be useful as markers for diversity and population studies.

RESULTS

General features of the genome assembly.

We assembled Illumina HiSeq 2000 sequencing data from four genomic libraries into 2,780 scaffolds. The draft assembly was rather contiguous, with an N_{50} scaffold length of 246.53 kbp and an L_{50} scaffold count of 46 (Supplementary Table S1). The total length of 35.39 Mbp falls within the range observed for other *Peronospora* genome assemblies (Supplementary Table S1). In total, 9,049 protein-coding genes were predicted from the repeat-masked genome. The circular mitochondrial genome was 40,106 bp long (Supplementary Fig. 1).

Repeat sequences account for 5.71 Mbp, which is 16.15% of the total assembly size. Furthermore, 18.20% of the scaffold-level assembly consisted of gaps between contigs. Nevertheless, more than 92% of the short-insert paired sequence reads were successfully mapped against the genome assembly, suggesting a high degree of completeness (Supplementary Table S2). This was also confirmed with BUSCO3 (Simão et al. 2015). Of the 234 core genes of the BUSCO Alveolata/Straminipila-specific library, 207 (88%) are found as intact and single-copy. This is similar to other *Peronospora* genome assemblies (Supplementary Fig. 2), which contain between 94 (40%) and 221 (94%) complete single-copy genes. Of the core genes reported as missing by BUSCO3, most could be found by tBLASTNn searches (Altschul et al. 1990) against the genome. In addition, synteny with other oomycete genome drafts revealed a rather high degree of synteny. There are 4,574 blocks on the *P. belbahrii* genome having synteny with *Bremia lactucae*, with a mean length of 2,063 bp and a total length of 9,437,307 bp. Given that the length of the *P. belbahrii* assembly is 35,394,047 bp, the synteny coverage is 26.66%, which is remarkably high, given the huge divergence and differences in genome size between the two species (Supplementary Fig. 3).

We note that the size of our genome assembly is smaller than that of a publicly available but unpublished *P. belbahrii* assembly (deposited by the University of Michigan, accession GCA_002864105.1), which is 59.2 Mbp in length and was assembled using Abyss2 (Jackman et al. 2017). However, BUSCO3 analysis reveals that the Michigan assembly contains a high frequency of gene duplication (Supplementary Fig. S2). We also observed similar duplications when we assembled our data using Abyss2 instead of Velvet. The duplication level can be explained by assembly of distinct alleles into separate contigs. In line with this, the analysis of our genome data confirms substantial heterozygosity in the *P. belbahrii* genome (Fig. 1; Supplementary Table S3).

Heterozygosity.

The genome of *P. belbahrii* displays significant levels of heterozygosity (Supplementary Table S3) as also indicated by a clear peak near to 50% allelic frequency in Figure 1A. A similar peak is also observed for some other oomycete genomes investigated (Supplementary Table S4), including those of *P. tabacina*, *Phytophthora ramorum* and *Saprolegnia parasitica* (Fig. 1A). The distribution of heterozygosity over the *P. belbahrii* genome is not uniform and we identified several long stretches of the genome showing little or no heterozygosity. For instance, we identified a patchy distribution of heterozygosity over scaffold 882 (249 kbp), in which there is a region of 70 kbp that shows no detectable heterozygosity with normal sequencing depth, while other regions of this scaffold show more typical levels of heterozygosity as exemplified by another region of 72 kbp (Fig. 1B).

Gene spacing and promoter structure.

Intergenic regions upstream of *P. belbahrii* start codons span a wide range of sizes (Supplementary Fig. S4A), consistent with the organization of its genome into gene-dense and gene-sparse regions (GDR and GSR, respectively); the former are defined as those in which genes have their 5′ ends within 2 kb of another gene. Seventy-six percent of predicted *P. belbahrii* genes are in GDRs, which is more than in other oomycetes such as *Phytophthora infestans* (51%) and *Plasmopara halstedii* (67%) (Baxter et al. 2010; Haas et al. 2009; Sharma et al. 2015). However, GSRs are likely to be disproportionately located in the unassembled gaps in draft-quality genome assemblies, meaning that these percentages should be interpreted with care. The median intergenic distance within *P. belbahrii* GDRs is 530 bp, which is slightly larger than in *Plasmopara halstedii* (420 bp) and *Phytophthora infestans* (430 bp).

P. belbahrii promoters are relatively AT-rich, rising to 59% AT 50 bp upstream of the start codon (Supplementary Fig. S4B). *P. belbahrii* contains core promoter motifs at different frequencies as compared with other oomycetes (Supplementary Fig. S4C). Each core motif occurs mostly within 100 bp of the start codon, as illustrated for the INR+FPR supra-motif in Supplementary Figure 4D.

As 64% of *P. belbahrii* promoters lacked a known core promoter sequence, attempts were made to identify alternative motifs using MEME (Bailey et al. 2009). Several candidates were identified, all of which resembled microsatellites. Figure 2A shows the most significant motif ($P = 10^{-33}$ compared with shuffled promoters), which is composed of TC dinucleotide repeats and showed a strong orientation bias ($P = 10^{-40}$ [Fig. 2B]). A search of all *P. belbahrii* promoters for $(TC)_n$ repeats with $n = 3$ revealed significant over-representation of arrays having up to eight repeat units ($P < 10^{-3}$ compared with shuffled sequences). A search for other dinucleotide microsatellites indicated that $(TC/GA)_n$ was the most plentiful in *P. belbahrii*, with $Z = 5.7$ ($P < 10^{-6}$) for over-representation compared with a shuffled dataset (Supplementary Fig. S4C). Some other microsatellites were also

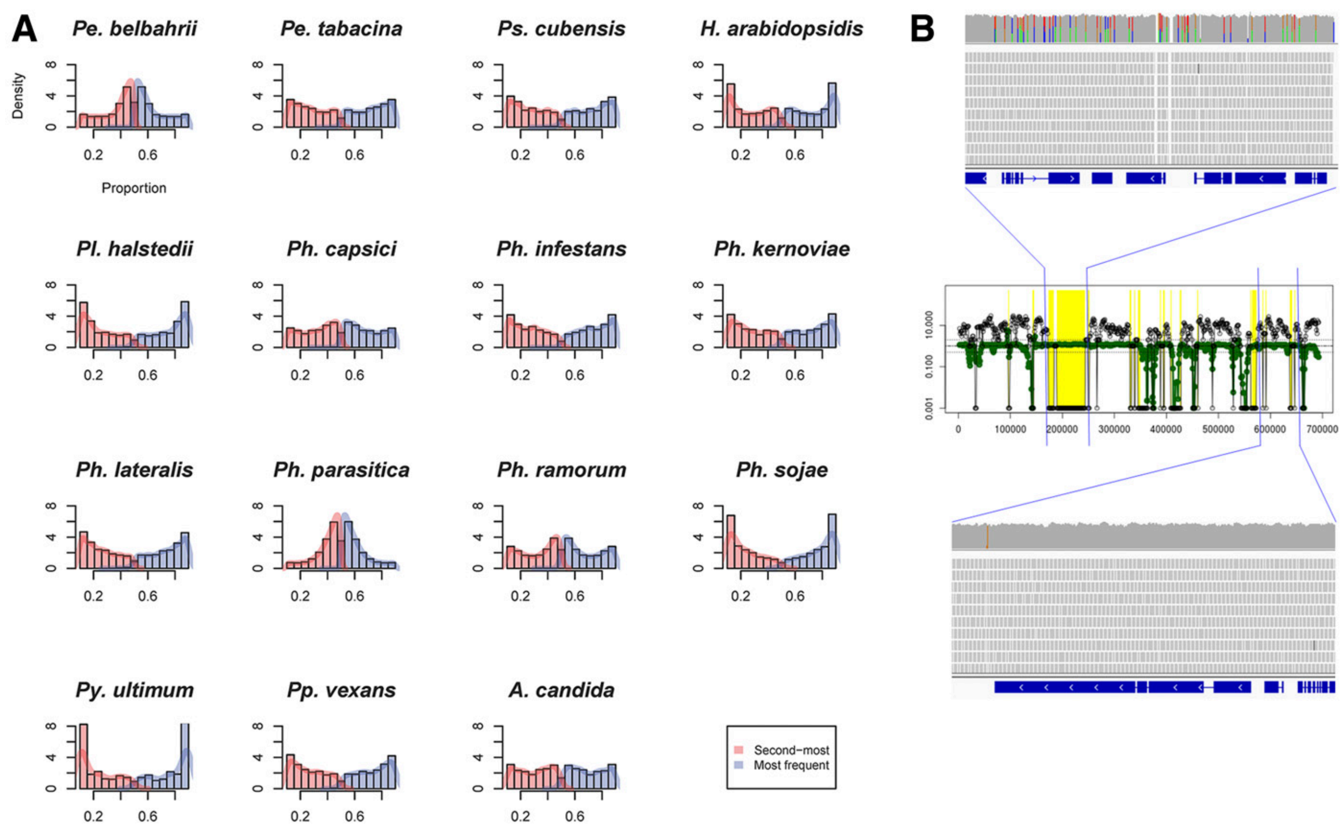

Fig. 1. Heterozygosity. **A,** Frequency distributions of relative abundances of the major allele over all genomic sites in sequenced oomycete genomes. For each species, raw Illumina sequence reads were aligned against the appropriate genome assembly using BWA-mem, discarding reads that align to more than one location. For each single-nucleotide polymorphism (SNP) site in each genome, the relative abundance of the most commonly occurring nucleotide (A, C, G, or T) was counted. The histogram shows the frequency distribution for the relative abundance of the most commonly occurring nucleotide. The histograms have been cropped at the top for clarity. A clear peak around 0.5 indicated diploidy and the presence of an increased number of heterozygous SNPs. **B,** Distribution of heterozygosity across scaffold 882 of the *Peronospora belbahrii* genome assembly (displayed in the center). In the upper panel, a heterozygous 72-kbp region is shown. Heterozygous sites are visible as two-colored vertical lines, with the relative lengths indicating the relative abundance of the two bases in the reads aligned at that site. The central panel shows an overview of the heterozygosity profile and the depth of coverage (*y* axis) over the whole scaffold. The black circles indicate the number of heterozygous sites per 5-kbp window. The green circles indicate the normalized depth of coverage (divided by the coverage median, 51.6). The dashed horizontal lines indicate 0.5, 1, and 2× median depth of coverage. The yellow shading indicates regions where no heterozygosity was detected and depth of coverage was at least 90% of the median. The lower panel shows a 72-kbp region that lacks detectable heterozygosity.

over-represented in promoters, including those with CT/AG, CA/TG, or AC/GT repeats.

Microsatellites.

Microsatellites have potential use in identification processes and population studies. We characterized the full microsatellite complement of *P. Belbahrii* and compared it with 34 other species of pathogens. The proportion of microsatellite sequences in the majority of oomycete genomes ranged from 0.023 (*Albugo candida*) to 0.331% (*Pythium irregulare*) when only two 6-bp arrays were considered. Higher ratios were observed for 1- to 6-bp simple sequence repeat (SSR) arrays (from 0.044% for *Plasmopara halstedii* to 0.341% for *Pythium irregulare*) (Supplementary Tables S5 to S7). The fraction of the genome comprised of SSR arrays in fungi with respectively similar lifestyle was higher (0.146% for *Blumeria graminis* f. sp. *tritici* to 0.668% for *Colletotrichum graminifolia*). The proportion of SSRs in genomes was not correlated with the size of the assembled genomes (R^2 = 0.0008). Interestingly, a low proportion of mononucleotides was observed in *Saprolegnia parasitica* (less than 1%), *P. belbahrii*, and *Pythium* species (3 to 10%). Dinucleotide and trinucleotide motifs were the most abundant types of repeats in the *P. belbahrii* genome, each representing more than 42% of all screened motifs (Supplementary Table S6). In coding regions of *P. belbahrii* di- and especially trinucleotide SSR arrays were frequently found, namely AAG/CTT, ATC/GAT, and AG/CT (Supplementary Table S7). In oomycete genomes, a higher proportion of AAG/CTT and AGC/GCT among other 3-bp arrays was observed, while the presence of 4- to 6-bp repeats was lower than in biotrophic fungi (Supplementary Tables S5 and S6). A total of 110 SSR markers (including 12 markers in coding regions) with potential use in population genetic studies were designed (Supplementary Table S8). These represent mainly di- and trinucleotide repeats (43 and 54 markers, respectively), while markers with tetra- to hexanucleotide repeats are less abundant (11, 1, and 1, respectively). All markers (109) are spread across 77 scaffolds and have been found in the alternative *P. belbahrii* assembly, and 19 of them are variable (i.e., we found an allele with a different size in the GCA_002864105.1 assembly).

Metabolism.

To compare metabolism of the obligate biotroph *P. belbahrii* with that of other oomycetes, Kyoto Encyclopedia of Genes and Genomes (KEGG)-based metabolic networks were constructed for 11 oomycete species with different lifestyles (Fig. 3). Overall, the metabolic networks of the obligate biotrophs *P. belbahrii*, *Plasmopara halstedii*, *Albugo laibachii*, and *Hyaloperonospora arabidopsidis* contain fewer genes, enzymes, reactions, and metabolites in comparison with the hemibiotrophs and the animal pathogen *Saprolegnia parasitica* (Fig. 3). The enzymes in each metabolic network were associated with reactions of 108 different KEGG pathways (Supplementary Table S9). Species clustered together according to their lifestyle when classified by numbers of supported reactions per pathway (Fig. 3), despite the fact that the non-peronosporalean *Albugo candida* and *Albugo labaichii* are phylogenetically not closely related to the downy mildews. As obligate biotrophy arose multiple times among the oomycetes (Thines 2014), this suggests convergent evolution of metabolic functions and defects.

Remarkably, the top ten most variant pathways (by coefficient of variation), contain three pathways classified in the KEGG BRITE category lipid metabolism (Supplementary Table S9). The steroid biosynthesis pathway was the most variable between species, in terms of supported reactions per pathway (Supplementary Table S9). *Peronospora belbahrii*, *Plasmopara halstedii*, and *Hyaloperonospora arabidopsidis* support only one reaction in this pathway, catalyzed by a cholesterol ester acylhydrolase (EC 3.1.1.13) and *Albugo laibachii* supports one additional reaction, catalyzed by an Acyl-CoA cholesterol *O*-acyltransferase (EC: 2.3.1.26). The obligate biotrophs contain generally fewer enzymes that function in the phenylpropanoid biosynthesis pathway (KEGG pathway map00940) than other oomycetes. Interestingly, *P. belbahrii* is an extreme case supporting only five reactions in this pathway, whereas the number of reactions in other oomycetes ranges from nine to 22. The analysis of primary metabolism pathways shows that all these mildews are unable to assimilate nitrate and nitrite and also lack an enzyme in the sulfur metabolism pathway generating cysteine (Fig. 4), which is in line with the previous studies (Baxter et al. 2010; Jiang et al. 2013).

In addition to the loss of metabolic pathways found in other oomycetes, *P. belbahrii*, like other members of the genus, does

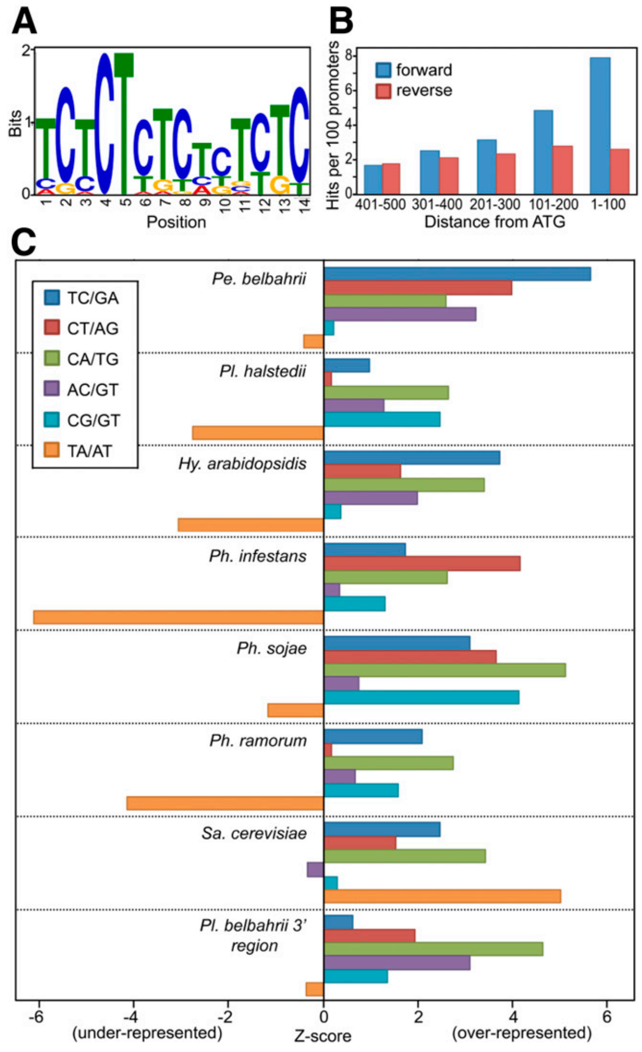

Fig. 2. Microsatellite motifs in promoters. **A,** TC-rich microsatellite-like motif detected in *Peronospora belbahrii* promoters. **B,** Orientation of the TC-rich motif in bins upstream of the start codon of *P. belbahrii* genes, based on matches with $P < 10^{-4}$. **C,** Representation of microsatellites with three or more perfect repeat units in the 100 bp upstream of the start codons of *P. belbahrii*, *Plasmopara halstedii*, *Hyaloperonospora arabidopsidis*, *Phytophthora infestans*, *Phytophthora sojae*, *Phytophthora ramorum*, and *Saccharomyces cerevisiae*, and within 100 bp 3′ of *P. belbahrii* coding sequences. Z-scores (positive for over-representation, negative for under-representation) were calculated using a twice-shuffled promoter dataset as a control.

not produce zoospores. To check whether its genome has retained genes normally required for forming flagella or remnants of such genes, sequences of 100 flagellum-associated proteins from *Phytophthora infestans* (Judelson et al. 2012) were used to search the *P. belbahrii* predicted proteome, using BLASTp, and its genome, using tBLASTn. No significant matches were detected, indicating a complete loss of its capacity to produce flagella.

Signaling.

Phospholipid modifying and signaling enzymes (PMSEs) may participate in both releasing carbon for metabolism and generating signaling molecules. Orthologs of nearly all genes encoding PMSE enzymes that have been identified in *Phytophthora* spp. and other oomycetes were found in *P. belbahrii* (Supplementary Table S10), but in comparison with *Phytophthora* spp., the total number is reduced (van den Hoogen et al. 2018). In *P. belbahrii*, as in *Hyaloperonospora arabidopsidis* and *Plasmopara halstedii*, there are fewer genes coding for phospholipase D (PLD) compared with *Phytophthora infestans*, largely due to a decrease in sPLDs, the subclass comprising PLDs that are likely secreted due to the presence of a signal peptide (Meijer et al. 2011). Also type H of the phosphatidylinositol kinase (PIK) genes are missing. From the largest PMSE family in oomycetes, phosphatidylinositol phosphate

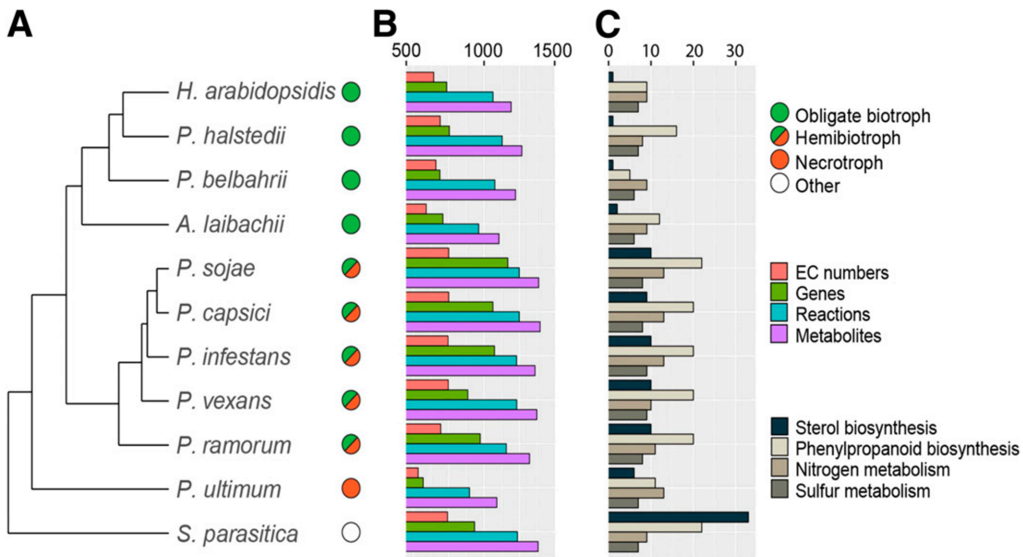

Fig. 3. Metabolic pathway analyses in 11 oomycete species with different lifestyles (depicted by circles). **A,** Hierarchical clustering of the species based on the number of reactions per KEGG pathway. **B,** The number of different EC numbers, genes, reactions, and metabolites included in the metabolic network and **C,** the number of reactions identified in four metabolic pathways for each of the species.

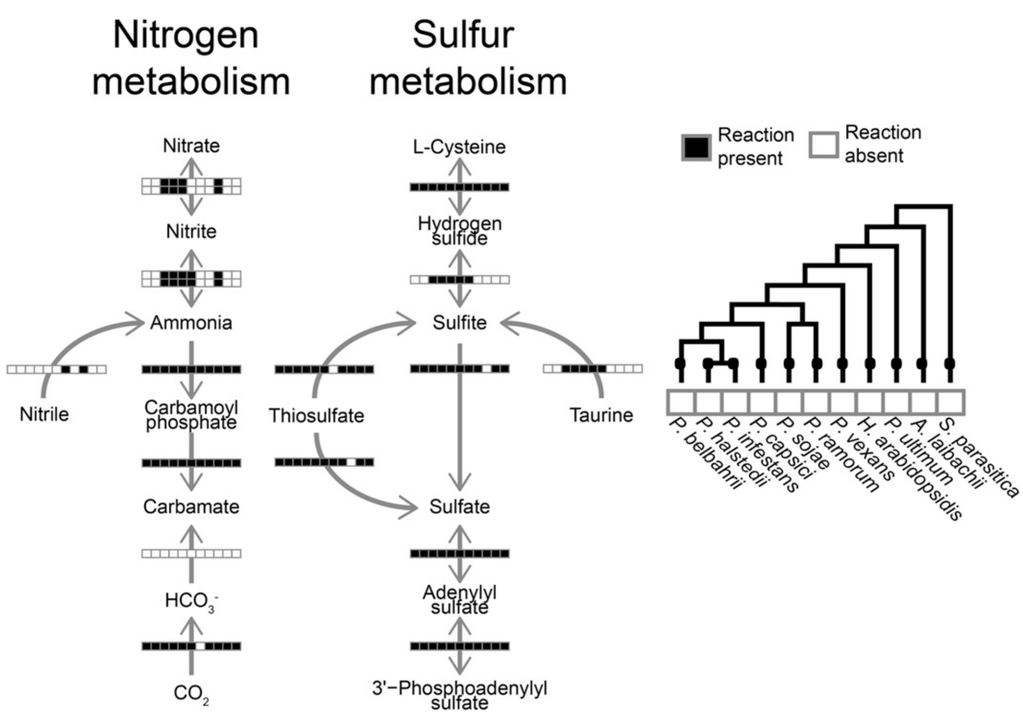

Fig. 4. Presence or absence analysis of enzymes in the nitrogen and sulfur metabolism pathways. Arrows indicate the direction of the reaction. The horizontal bars are composed of black and white squares representing the species in the order of the phylogenetic tree on the right and presence or absence of the enzyme catalyzing the reaction in the respective species.

kinases (PIPKs) with an N-terminal G-protein coupled receptor (GPCR) domain (GPCR-PIPKs), 11 members are detected in *P. belbahrii* (Supplementary Table S10). In common with other peronosporalean oomycetes, *P. belbahrii* lacks a canonical phospholipase C gene. The only light-sensing proteins detected in the genomes of *P. belbahrii* and other oomycetes belong to the CRY/PHR superfamily. Each species contained three such proteins, which formed three well-supported clades in phylogenetic analyses (Supplementary Fig. S5A). Each of the three *P. belbahrii* proteins (PEBEL_00732, PEBEL_04497, and PEBEL_06190) had the canonical (Chaves et al. 2011) photolyase and FAD chromophore binding domains expected for CRY/PHRs (Supplementary Fig. S5B). None had long C-terminal domains present on animal cryptochromes and are known to mediate circadian rhythm. *Peronospora belbahrii* also contained a fourth predicted protein (PEBEL_00742), which was 99.4% identical at the nucleotide level to PEBEL_06190, probably representing a duplicated gene.

The secretome.

Out of 9,049 protein-coding genes, 413 were predicted as encoding secreted protein–encoding genes. A set of 381 candidate genes encoding proteins with a leading signal peptide and no transmembrane domain was used as a starting point for further predictions and annotations. Additional curation resulted in 335 putative extracellular proteins (Table 1; Supplementary Table S11) organized as 76 tribes and 84 singletons. For roughly half of all proteins (171), it was possible to assign a particular functional category (i.e., not 'hypothetical'). Annotated proteins with functions previously assigned to the infection process (effectors, pathogen associated molecular patterns, SCP, cell wall–degrading enzymes [CWDE], proteases) comprise 31% of all secreted proteins (Table 1). In general, when compared with other obligate parasitic oomycetes, we found reduced repertoires of secreted proteins in the *P. belbahrii* secretome (Table 1). Strikingly, we identified no CRN (crinkling and necrosis) proteins with predicted signal peptides and only two NPP-like proteins. Proteins with similarity to CRNs but lacking a predicted signal peptide have been found in oomycete genomes previously, but the low number of NLPs is unprecedented.

Among predicted effectors (Supplementary Tables S11 and S12), the most abundant category encompasses RxLR-like effectors (22 proteins, five with a WY fold), three of which (PEBEL_01290, PEBEL_01294, PEBEL_02639) share sequence homology with known oomycete avirulence genes from *Phytophthora sojae* (Avh154, Avh160 and Avh152). Of 22 predicted RxLR effectors, 14 have the exact RxLR motif, while five harbor a RxLK motif and three singletons have other variations of the motif (Fig. 5).

Only two protease inhibitors were identified (PEBEL_30414, PEBEL_30415), both having sequence similarity to Kazal-like serine protease inhibitors (Supplementary Tables S11 and S12; Table 1).

Six *P. belbahrii* predicted secreted proteins contain potential nuclear localization sequence (NLS) signals (Supplementary Table S12). Presence of an NLS in combination with a signal peptide suggests a role of these proteins in the host nucleus, yet only one (PEBEL_07289) carries a recognizable RxLR motif implied in host translocation, while six other (PEBEL_06934.2, PEBEL_06683, PEBEL_07130, PEBEL_04801, PEBEL_08621.2, and PEBEL_05668) might remain extracellular or be translocated by an RxLR-independent mechanism.

The *P. belbahrii* secretome contains several hydrolytic enzymes (63, 19% of all extracellular proteins), which we classified into two main categories, CWDEs and proteases. CWDEs are mainly represented by glycosyl hydrolases (n = 24),

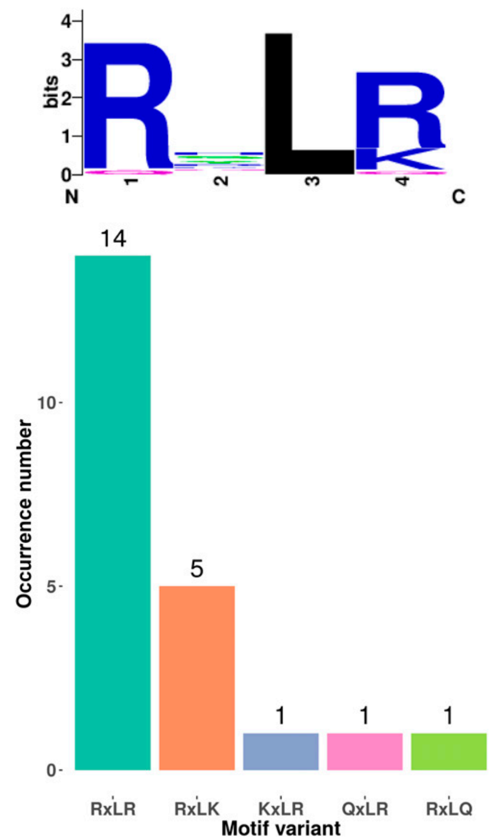

Fig. 5. RxLR motif conservation. Sequence logo for predicted RxLR-like effectors of *Peronospora belbahrii*. Only 14 canonical RxLR effectors and eight variants are present in the predicted secretome, which is, by far, the lowest number for any *Phytophthora* or downy mildew genome sequenced so far.

Table 1. Classes of secreted proteins in selected oomycete genomes

Genome/protein class	*Peronospora belbahrii*	*P. tabacina*[a]	*Plasmopara halstedii*[b]	*Hyaloperonospora arabidopsidis*[c]
NLP	2	13	19	10
RxLR effector	22	120	274	134
CRN	0	7	77	20
SCP	9	14	Not reported	8
Protease	26	23	131	18
CWDE	37	32	8	>65
Elicitin and elicitin-like	5	10	16	17

[a] Secreted proteins were predicted as described by Derevnina et al. (2015). The table contains averaged numbers for two sequenced strains (968-J2 and 968-S26).
[b] According to gene family annotation from Sharma et al. (2015).
[c] According to gene family annotation from Baxter et al. (2010).

while proteases are dominated by serine and cysteine proteases (Table 1). We detected only five proteins with similarity to known immune response elicitors, represented exclusively by elicitins.

To identify common and specific secreted protein families, OrthoMCL was used to group predicted secretome members of *P. belbahrii* with those from several plant-pathogenic oomycetes into protein families based on homology. Eleven families representing 23 proteins solely contain sequences from *P. belbahrii*. Nine of the 23 proteins (39%) are predicted as RxLR effectors, whereas the predicted RxLR effector proteins only make up 7.4% of the total predicted secretome. Furthermore, 91 *P. belbahrii* proteins did not cluster with any other protein in the set and are probably unique to the *P. belbahrii* secretome (Supplementary Tables S11 and S12).

Twenty-three ortholog families with a combined 68 genes were found to be exclusively present in one or more downy mildews. Of these, 12 families (35 genes) are not shared with *P. belbahrii*. A BLASTp search of the other families against the National Center for Biotechnology Information nonredundant database on the full proteins (including signal peptide) showed that the best non–downy mildew matches for the *P. belbahrii* proteins in six of these families do not have signal peptide (Supplementary Table S13). These proteins have potentially evolved to acquire a signal peptide and a new function in downy mildews.

DISCUSSION

We assembled and annotated the genome sequence for *P. belbahrii* with the aim to produce a useful resource to the community for follow-up studies. The size and quality of the resulting genome are comparable to genome assemblies of other members of the genus and other oomycetes. While the genome is similar to many other oomycetes in terms of size and architecture, a remarkable diversity in the occurrence of dinucleotide repeats in promoters was observed. The $(TC/GA)_n$ repeats were over-represented strongly not only in *P. belbahrii* but also in *Hyaloperonospora arabidopsidis* and three *Phytophthora* species but not in *Plasmopara halstedii*. Notably, $(TC/GA)_n$ elements were not over-represented in the regions downstream of *P. belbahrii* genes. Interestingly, (TA/AT)n showed a strong tendency toward under-representation in most oomycete promoters, even though it is over-represented in members of several other eukaryotic kingdoms including fungi (*Saccharomyces cerevisiae*) (Bunnik et al. 2014; Sawaya et al. 2013; Vinces et al. 2009). In other species, dinucleotide repeats are believed to affect the positioning of nucleosomes on DNA (Bunnik et al. 2014; Sawaya et al. 2013; Vinces et al. 2009). Tandemly repeated DNA sequences in promoters are of interest, since they display a propensity to mutate, which may enable the tuning of gene expression by affecting chromatin structure (Vinces et al. 2009). Under-representation of the AT-rich repeats in most oomycetes is, in itself, likely a signal of function and is suggestive of divergence in organization of chromatin in oomycetes. In many cases, SSR repeat number appears to be a key factor that determines gene expression and expression level. Some genes can only be expressed at a specific repeat number of SSRs. For example, the GAA array, which is the most frequent trinucleotide repeat (i.e., AAG/CTT) in the majority of analyzed oomycete genomes, including *P. belbahrii*, is reported to be the key factor influencing the promoter activity of the *Escherichia coli lacZ* gene (Liu et al. 2000). In addition to their potential importance in regulating gene expression, SSRs are also of considerable interest for population genetics studies. Therefore, a set of 110 microsatellite markers was developed in this study, which may prove to be a useful tool for future population genetic studies in *P. belbahrii* and closely related species on other culinary herbs and of which 19 were found to be variable in comparison with an unpublished genome assembly of *P. belbahrii*.

So far, relatively little is known about oomycete metabolism. Several metabolic pathways have been affected by the consequences of the close symbiosis of downy mildews and their hosts, which led to a reliance on the host metabolism to provide nutrients for the pathogen (Judelson 2017; Thines 2014). It is interesting to note that, based on hierarchical clustering of metabolic networks, the different oomycete species clustered together according to their lifestyle and that, therefore, *Albugo candida* clustered with the downy mildews, irrespective of its distant phylogenetic relationship to them (Thines 2014). Similarly, the hemibiotrophic *Phytophthora* species clustered together, forming a sister group to the obligate biotrophic taxa. This supports the theory that similar evolutionary forces due to the adaptation to similar lifestyles will lead to convergent patterns (Kemen et al. 2011; Sharma et al. 2015b). In line with previous studies, we found that the metabolic networks of obligate pathogens are generally smaller (Rodenburg et al. 2018; Spanu 2012) and lack enzyme-encoding genes for reactions in various metabolic pathways (Baxter et al. 2010; Judelson 2017; Kemen et al. 2011; Sharma et al. 2015a).

Many cellular pathways are linked to phospholipid signaling, including not only metabolic pathways, but also pathways regulating the expression of genes required for successful colonization. Despite the importance of phospholipid signaling demonstrated in many organisms, in oomycetes only a few PMSEs have been studied in detail. The domain composition of GPCR-PIPKs, shared by all oomycetes, points to a mechanism that directly links G-protein mediated signaling with phospholipid signaling (van den Hoogen et al. 2018). Silencing of a PIK-A and PIK-D in *Phytophthora sojae* showed that these two PIKs are required for full virulence (Lu et al. 2013), and GPCR-PIPKs were found to be involved in sporangia development, chemotaxis, zoospore development, and oospore development (Hua et al. 2013; Yang et al. 2013). Given this wide span of lifecycle stages affected, it seems possible that other developmental stages are also influenced by phospholipid signaling, such as formation of the hyphal network and the production of haustoria.

Extracellular effector proteins secreted from hyphae and haustoria play key roles in infection strategies of pathogenic oomycetes, helping to establish colonization and to modulate plant responses (Meijer et al. 2014). By investigating the protein domains of secreted proteins, their role in infection can be clarified. While a large part of the secretome of *P. belbahrii* could be classified through sequence similarity, 203 proteins remain hypothetical and could represent yet-to-be discovered effector proteins.

Besides canonical effector proteins such as RxLRs, two proteins that possess a carbonic anhydrase domain (Pfam PF00194) were identified in the secretome. In a study of the secretome of *Phytophthora infestans*, carbonic anhydrases were found to be expressed in planta during infection of potato and they were suggested to be potential new virulence factors (Raffaele et al. 2010). In addition, the secretome comparison revealed 23 families containing 68 proteins that are only found in downy mildews but not in more distantly related *Phytophthora*, *Pythium*, or *Saprolegnia* species. Of these, 10 families (28 proteins) include proteins derived from *P. belbahrii*. Remarkably, a similarity BLASTp search revealed that proteins in six of these families had an ortholog without signal peptide in other oomycetes, like in *Phytophthora* species. Possibly these genes are present in other oomycetes and have evolved to encode for secreted proteins within the lineage of the downy mildews. Some of these proteins have interesting domains, e.g., DUF953 (Seidl et al. 2011), E3 ubiquitin-protein ligase (Zeng et al. 2006), or bZIP_2 (Gamboa-Meléndez et al. 2013). For proteins in the last two downy mildew families that have a non–downy mildew ortholog without a predicted signal peptide, no known domain was identified. Interestingly, one of these families is the largest family only present in downy mildews, encompassing five proteins from three different species (*P. tabacina*, *P. belbahrii*, and *Plasmopara halstedii*). The other group has three proteins from three species (*P. belbahrii*, *Hyaloperonospora arabidopsidis*, and *Plasmopara halstedii*. The functions of these proteins are unknown, and since they seem to be present in the secretome of downy mildew species but not of other oomycetes included, they are interesting candidates for functional validation.

Strikingly, we found an underrepresentation of two common groups of effectors, the CRNs and NLPs compared with other secretomes. It is possible that CRNs are underrepresented because they might have not been predicted by SignalP, since several known CRNs have noncanonical signal peptides (Stam et al. 2013). However, the generally low number of classical effectors, such as the potentially necrosis-causing CRNs, NLPs, and RxLRs in *P. belbahrii*, might reflect the combined effects of adaption to biotrophy and host specificity.

Of the only 22 RxLR-like effectors found in *P. belbahrii*, seven did not have orthologs in other species. Possibly, these effectors can be considered species-specific or have become too dissimilar from their orthologs to be recognized. Among the RxLR effectors, PEBEL_07289 is a notable member, with C-terminal homology to Nudix hydrolases. The effector AVR3b *Phytophthora sojae*, another NUDIX domain containing an RxLR effector has been demonstrated to increase susceptibility to *Phytophthora capsici* and *Phytophthora parasitica* (Dong et al. 2011). It is noteworthy that the genome of *P. belbahrii* was predicted to contain the lowest number of RxLR-like effectors compared with all other sequenced Peronosporaceae. This might hint at a very high specialization of *Peronospora* species (Choi et al. 2009; Thines et al. 2009; Voglmayr 2003), necessitating only a few highly effective effectors, as targets need to be manipulated mostly in just one or very few closely related host species (Thines and Kamoun 2010). Alternatively, or in addition, this might reflect the evolution of a few highly efficient core effectors that also enable *P. belbahrii* to extend its range (Thines 2019) and infect other related species under optimal conditions (Naim et al. 2019).

In conclusion, analyses of the *P. belbahrii* genome revealed a rather small oomycete genome with little more than 300 secreted proteins. We found a remarkable signal of convergent evolution with respect to lifestyle in various oomycete genomes with metabolic pathway losses and patterns of diversification. Our analyses suggest a divergent organization of promoters, as suggested by the high frequency of TC repeats. In addition, we found the least amount of canonical effectors in any oomycete genome sequenced so far but found hints for the presence of virulence factors that have, so far, not been investigated and which are promising candidates for future functional studies.

MATERIALS AND METHODS

DNA isolation from spores and library preparation for genomic sequencing.

Several heavily infected basil plants with fresh sporulation (violet-brown down on the lower leaf surfaces) from a single lot were bought from a supermarket in Butzbach, Germany in summer 2015. From these, several leaves were rinsed with deionized water to collect spores. DNA was extracted from spores pelleted by centrifugation, as described previously (Sharma et al. 2015a). RNA was extracted from fresh leaves with freshly sporulating pathogen, as described previously (Sharma et al. 2015). DNA and RNA extracts were sent to a commercial sequencing provider (LGC Genomics) for preparation and sequencing of 300-bp, 800-bp, 3-kbp, and 8-kbp genomic libraries and a 300-bp library for RNA-Seq. Paired-end Illumina HiSeq sequencing was carried out with 100 bases sequenced from both ends of the fragments in the libraries.

Next-generation sequencing data processing and genome assembly.

Standard primers and adapter sequences were clipped from the raw genomic Illumina reads using Trimmomatic (Bolger et al. 2014). A window-based quality filtering using Trimmomatic was performed by keeping an average quality cut-off of 20 within a window of size 5 bp. Reads shorter than 65 bp were filtered out. The second round of filtering was performed using FastQFS (Sharma and Thines 2015). Read pairs were eliminated if either read had an ambiguous base or if an individual base had a Phred-scale quality score of 3 or lower or if the average Phred score was below 26. The filtered reads were assembled using Velvet v1.2.09 (Zerbino and Birney 2008). Several different assemblies were generated using different k-mer sizes and other parameters, including expected coverage and coverage cut-offs. Different assemblies were compared, and the best assembly was decided on the basis of N_{50}, L_{50}, number of scaffolds, and genome completeness, using Quast (Gurevich et al. 2013) and BUSCO3 (Simão et al. 2015). Assemblies were also generated using SPAdes version 3.12.0 and SOAPdenovo version 2.04, exploring a range of k-mer sizes; however, these yielded less-complete and less-contiguous assemblies than the Velvet assembly. The mitochondrial genome was assembled from the Illumina short reads, and a mitochondrial backbone sequence of *P. tabacina* (Derevnina et al. 2015) was assembled using the Mitomaker pipeline. The physical map of the genome was visualized using OGdraw (Lohse et al. 2007). An alignment of the mitochondrial genome of *P. tabacina* and *P. belbahrii* was made using Easyfig software (available online).

Repeat element and gene prediction.

Repeat elements were predicted and masked using the RepeatScout v1 (Price et al. 2005), RepeatModeler v1.0.4 (available online), and RepeatMasker pipelines (Tarailo-Graovac and Chen 2009), as described before (Sharma et al. 2015). The repeat-masked scaffolds were used for gene predictions. Both *ab initio* and transcript-based gene prediction tools were used to define gene boundaries as described previously (Sharma et al. 2015) (Supplementary Fig. S6). Translated protein sequences were searched for an extracellular secretion signal and transmembrane domain using SignalP (Petersen et al. 2011), TargetP (Emanuelsson et al. 2000), and TMHMM (Krogh et al. 2001) as described (Sharma et al. 2015). Due to low RNA-Seq support for some genes predicted to code for secreted proteins, gene boundaries of potential secreted effectors were manually curated by looking for potential alternate start codons up to 280 bp upstream of the start codon predicted by evidence modeling.

Estimation of heterozygosity.

The availability of shotgun sequencing reads at high coverage, drawn randomly from both chromosomes from each homologous pair, offers the opportunity to estimate patterns of heterozygosity across the genome of various oomycetes (Supplementary Table S4). First, poor-quality sequence read pairs were removed using TrimGalore with the –q 30 and –paired options, and then, the remaining read pairs from the 300-bp library were aligned against the *P. belbahrii* genome assembly using BWA-mem (Li and Durbin 2009). Multi-mapping reads were eliminated using samtools view (Li et al. 2009) with the –q 1 option. For each single-nucleotide position in this alignment of reads against the genome assembly, we counted the frequency of each of the four possible bases in the reads aligned at that site. Thereby, for each site in the genome, we were able to estimate whether it was likely to be homozygous (most common base close to 100% frequency and second-most common close to zero) or heterozygous (close to 50% most common base and for second-most common base). Using R (R Development Core Team 2013), we plotted a histogram describing the frequency distribution of relative abundance of the most abundant base and second-most abundant base at each position in the genome. If the genome is highly homozygous, then the histogram would be expected to display a single peak close to 100% abundance for the most common base and a single peak close to zero for the second-most abundant. However, heterozygous sites would contribute to a second peak close to 50% abundance of the most common base and for the second-most abundant base. We also took a sliding-window approach to identify regions of high and low heterozygosity relative to the average over the whole genome. Each non-overlapping window of 5 kbp was examined for heterozygous sites, i.e., sites where the abundance of the most common base was between 48 and 52%. The density of heterozygous sites was defined as the number of heterozygous sites per 5-kb window. Density of heterozygous sites was plotted along with depth of coverage against genomic position, using R. The R scripts used for these analyses are available from Github.

Metabolic networks.

Metabolic networks were reconstructed for *P. belbahrii*, *Plasmopara halstedii*, *Phytophthora infestans*, *Phytophthora capsic*i, *Phytophthora sojae*, *Phytophthora ramorum*, *Phytopythium vexans*, *Hyaloperonospora arabidopsidis*, *Pythium ultimum*, *Albugo laibachii*, and *Saprolegnia parasitica* using the RAVEN toolbox (Agren et al. 2013). Briefly, KEGG orthologous groups of enzymes (Chen et al. 2016; Kanehisa et al. 2012) were aligned using ClustalW2 v2.0.10 (Larkin et al. 2007), and, based on these multiple sequence alignments, hidden Markov models (HMMs) were trained using HMMER v2.3 (Eddy 2011). Subsequently, the proteomes were matched to the HMMs with an E-value threshold of 10^{-50}. The resulting enzyme orthologs were associated to KEGG reactions, which are linked to compounds and pathways. A metabolic reaction may occur if one or more catalyzing enzymes are encoded in the genome. Hence, for each KEGG metabolic pathway, the number of supported reactions per species was counted. For these numbers, the coefficient of variation (variance/mean) was used to select the most variant pathways between species. In addition, these numbers were used to perform hierarchical clustering (unweighted pair group method with arithmetic means) of the species, using 1 minus the Spearman correlation coefficient as a distance measure.

Promoter structure.

Core promoter motifs as defined for *Phytophthora infestans* (Roy et al. 2013) were used to search 200 bp of DNA 5′ of *P. belbahrii* coding sequences using FIMO (Bailey et al. 2009). Searches for new motifs were performed using MEME (Bailey et al. 2009). Microsatellites within promoters were scored using MICAS (Sreenu et al. 2003). Twice-shuffled promoter sequences were used as control datasets and Fisher's exact test was used for tests of significance.

Microsatellite motifs and prediction of SSR markers.

The overall statistics for the most frequent motif types, repeat numbers, and length of microsatellite sequences in *P. belbahrii* genome were calculated in Phobos (Mayer et al. 2010). SSRs with perfect repeat motifs ranging from mono- to hexanucleotides were considered, with the minimum number of ten (1 bp) or five (2 to 6 bp) repeat units. Searches of each motif were performed in separate runs (i.e., for the 3-bp array, the "Repeat unit size range" was set "from 3 to 3" and "Minimum length" set to "15 + 0 but not less than 15"), with results exported in "one-per-line" format, "Printing mode" set to "normalized – cyclic + rev. complement", and all other settings set as default. Subsequent data analyses were conducted in Microsoft Excel 2010. For comparison, the genomes of other oomycetes and selected fungi (Supplementary Table S14) were analyzed as for *P. belbahrii*.

To facilitate the design of *P. belbahrii* SSR markers, only perfect di- to hexanucleotide repeats, with minimum repeats set to 8, 7, 6, 6, and 6 respectively, were extracted using Phobos. Primer pairs were designed using Primer3 (Untergasser et al. 2012), yielding PCR products of 150 to 500 bp. Data resulting from SSR analyses refer to duplex DNA, even if we show only the sequence of the repeated motif on one strand for simplicity, i.e., notations like AC and $(AC)_n \bullet (GT)_n$ are equivalent. To see if the proportion of SSR arrays is correlated with the size of an assembled genome, linear regression in MS Excel was performed using the standard formula implemented in MS Excel 2010. Finally, the relative abundance (calculated as the ratio of the number of microsatellites per megabase of sequence analyzed) and relative density (the ratio of the microsatellite length in base pairs per megabase pairs of the sequence analyzed) of microsatellites were calculated separately.

Light sensing and phospholipid signaling.

Representative light-sensing proteins from other eukaryotes were used to query a database of predicted proteins from *P. belbahrii*, *Plasmopara halstedii*, *Hyaloperonospora arabidopsidis*, and *Phytophthora infestans*, using BLASTp. If no hits were found, the entire genome was searched using tBLASTn. Protein alignments were made using MUSCLE and trees generated using PhyML using the LG distance model.

The genome of *P. belbahrii* was screened for PMSE-encoding genes. A database of *Phytophthora infestans*, *Hyaloperonospora arabidopsidis*, and *Plasmopara halstedii* PMSE protein sequences was created, and BLASTp and tBLASTn and searches were performed with an E-value cut-off score of $1e^{-10}$. Hits were manually inspected and PMSE gene models were linked to gene IDs. All gene models were manually inspected. Multiple sequence alignments on catalytic domains were made, using MUSCLE (Edgar 2004), and iterated 100 times, and phylogenetic trees were built using neighbor-joining with the Jukes-Cantor genetic distance model in Geneious R9 (Kearse et al. 2012).

The secretome.

For annotating candidate secreted proteins, profiles were built using several approaches: BLASTp search against GenBank NR database with E-value cut-off of 10^{-6}; an InterProScan v5.16 (Mulder et al. 2005) search against databases of functional domains with default parameters; RxLR and EER motif prediction using regular expressions; WY motif prediction based on WY-fold HMM, using the hmmsearch function in the HMMR3 package (available online); LxLFLAK and HVLVVVP motif predictions based on an HMM model built using known CRN effectors from other oomycetes, according to Haas et al. (2009); and NLS motif prediction by NLStradamus version 1.8 (Nguyen Ba et al. 2009), posterior threshold = 0.6 and PredictNLS (Cokol et al. 2000) with default parameters. All obtained data are aggregated in Supplementary Table S6. Functional categories were assigned based on manual curation of the resulting table. The category "Hypothetical" was assigned to proteins either having similarity to only hypothetical proteins or when the best 20 hits of the BLASTp output did not show consistency in terms of distinct functional categories. Proteins having significant sequence similarity to ribosomal, transmembrane proteins, or proteins with either known intracellular localization (e.g., heat shock proteins), having respective domains identified by InterProScan, or both were marked as possible false predictions. The contamination category was assigned for proteins with significant sequence similarity (revealed by BLASTp) to amino acid sequences from phylogenetically distant taxa (e.g., plants or mammals). It should be noted that this category will also include highly conserved proteins. Entries marked as both "false prediction" or "contamination" were excluded from further consideration.

Nine oomycete genomes were selected for secretome comparison: three downy mildew species (*Hyaloperonospora arabidopsidis*, *P. tabacina* (two strains), and *Plasmopara halstedii*), three representative *Phytophthora* species (*Phytophthora infestans*, *Phytophthora sojae*. and *Phytophthora ramorum*), the more distantly related plant pathogen *Pythium ultimum*, and the fish-parasitizing oomycete *Saprolegnia diclina*. For each of these species the set of all predicted proteins was used to filter out the proteins that belong to the secretome as described for the *P. belbahrii* secretome. OrthoMCL 2.0 (Fischer et al. 2011) was used to create groups of orthologous proteins, with default settings.

AUTHOR-RECOMMENDED INTERNET RESOURCES

Easyfig software: https://www.ncbi.nlm.nih.gov/pubmed/21278367
Github: https://github.com/davidjstudholme/heterozygosity
HMMR3 package: http://hmmer.org
Mitomaker pipeline: http://sourceforge.net/projects/mitomaker
RepeatModeler v1.0.4: http://www.repeatmasker.org/RepeatModeler
TrimGalore: http://www.bioinformatics.babraham.ac.uk/projects/trim_galore

LITERATURE CITED

Agren, R., Liu, L., Shoaie, S., Vongsangnak, W., Nookaew, I., and Nielsen, J. 2013. The RAVEN toolbox and its use for generating a genome-scale metabolic model for *Penicillium chrysogenum*. PLOS Comput. Biol. 9: e1002980.

Ali, S. S., Shao, J., Lary, D. J., Kronmiller, B., Shen, D., Strem, M. D., Amoako-Attah, I., Akrofi, A. Y., Begoude, B. A., 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 *P. 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.

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W., and Noble, W. S. 2009. MEME SUITE: Tools for motif discovery and searching. Nucleic Acids Res. 37:W202-W208.

Baxter, L., Tripathy, S., Ishaque, N., Boot, N., Cabral, A., Kemen, E., Thines, M., Ah-Fong, A., Anderson, R., Badejoko, W., Bittner-Eddy, P., Boore, J. L., Chibucos, M. C., Coates, M., Dehal, P., Delehaunty, K., Dong, S., Downton, P., Dumas, B., Fabro, G., Fronick, C., Fuerstenberg, S. I., Fulton, L., Gaulin, E., Govers, F., Hughes, L., Humphray, S., Jiang, R. H. Y., Judelson, H., Kamoun, S., Kyung, K., Meijer, H., Minx, P., Morris, P., Nelson, J., Phuntumart, V., Qutob, D., Rehmany, A., Rougon-Cardoso, A., Ryden, P., Torto-Alalibo, T., Studholme, D., Wang, Y., Win, J., Wood, J., Clifton, S. W., Rogers, J., Van den Ackerveken, G., Jones, J. D. G., McDowell, J. M., Beynon, J., and Tyler, B. M. 2010. Signatures of adaptation to obligate biotrophy in the *Hyaloperonospora arabidopsidis* genome. Science 330:1549-1551.

Blackman, L. M., Cullerne, D. P., Torreña, P., Taylor, J., and Hardham, A. R. 2015. RNA-Seq analysis of the expression of genes encoding cell wall degrading enzymes during infection of lupin (*Lupinus angustifolius*) by *Phytophthora parasitica*. PLoS One 10:e0136899.

Bolger, A. M., Lohse, M., and Usadel, B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-2120.

Boutemy, L. S., King, S. R. F., Win, J., Hughes, R. K., Clarke, T. A., Blumenschein, T. M. A., Kamoun, S., and Banfield, M. J. 2011. Structures of *Phytophthora* RXLR effector proteins: A conserved but adaptable fold underpins functional diversity. J. Biol. Chem. 286: 35834-35842.

Bunnik, E. M., Polishko, A., Prudhomme, J., Ponts, N., Gill, S. S., Lonardi, S., and Le Roch, K. G. 2014. DNA-encoded nucleosome occupancy is associated with transcription levels in the human malaria parasite *Plasmodium falciparum*. BMC Genomics 15:347.

Chaves, I., Pokorny, R., Byrdin, M., Hoang, N., Ritz, T., Brettel, K., Essen, L. O., van der Horst, G. T., Batschauer, A., and Ahmad, M. 2011. The cryptochromes: Blue light photoreceptors in plants and animals. Annu. Rev. Plant Biol. 62:335-364.

Chen, L., Gong, Y., Cai, Y., Liu, W., Zhou, Y., Xiao, Y., Xu, Z., Liu, Y., Lei, X., Wang, G., Guo, M., Ma, X., and Bian, Y. 2016. Genome sequence of the edible cultivated mushroom *Lentinula edodes* (Shiitake) reveals insights into lignocellulose degradation. PLoS One 11:e0160336.

Choi, Y.-J., Kiss, L., Vajna, L., and Shin, H.-D. 2009. Characterization of a *Plasmopara* species on *Ambrosia artemisiifolia*, and notes on *P. halstedii*, based on morphology and multiple gene phylogenies. Mycol. Res. 113:1127-1136.

Cokol, M., Nair, R., and Rost, B. 2000. Finding nuclear localization signals. EMBO Rep. 1:411-415.

Dahlin, P., Srivastava, V., Ekengren, S., McKee, L. S., and Bulone, V. 2017. Comparative analysis of sterol acquisition in the oomycetes *Saprolegnia parasitica* and *Phytophthora infestans*. PLoS One 12:e0170873.

Derevnina, L., Chin-Wo-Reyes, S., Martin, F., Wood, K., Froenicke, L., Spring, O., and Michelmore, R. 2015. Genome sequence and architecture of the tobacco downy mildew pathogen *Peronospora tabacina*. Mol. Plant-Microbe Interact. 28:1198-1215.

Dong, S., Yin, W., Kong, G., Yang, X., Qutob, D., Chen, Q., Kale, S. D., Sui, Y., Zhang, Z., Dou, D., Zheng, X., Gijzen, M., Tyler, B. M., and Wang, Y. 2011. *Phytophthora sojae* avirulence effector Avr3b is a secreted NADH and ADP-ribose pyrophosphorylase that modulates plant immunity. PLoS Pathog. 7:e1002353.

Dou, D., Kale, S. D., Wang, X., Chen, Y., Wang, Q., Wang, X., Jiang, R. H., Arredondo, F. D., Anderson, R. G., Thakur, P. B., McDowell, J. M., Wang, Y., and Tyler, B. M. 2008. Conserved C-terminal motifs required for avirulence and suppression of cell death by *Phytophthora sojae* effector Avr1b. Plant Cell 20:1118-1133.

Eddy, S. R. 2011. Accelerated profile HMM searches. PLOS Comput. Biol. 7:e1002195.

Edgar, R. C. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792-1797.

Ellis, J. G., Dodds, P. N., and Lawrence, G. J. 2007. The role of secreted proteins in diseases of plants caused by rust, powdery mildew and smut fungi. Curr. Opin. Microbiol. 10:326-331.

Emanuelsson, O., Nielsen, H., Brunak, S., and von Heijne, G. 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J. Mol. Biol. 300:1005-1016.

Feng, B.-Z., Zhu, X.-P., Fu, L., Lv, R.-F., Storey, D., Tooley, P., and Zhang, X. G. 2014. Characterization of necrosis-inducing NLP proteins in *Phytophthora capsici*. BMC Plant Biol. 14:126.

Feng, C., Lamour, K. H., Bluhm, B. H., Sharma, S., Shrestha, S., Dhillon, B. D. S., and Correll, J. C. 2018. Genome sequences of three races of *Peronospora effusa*: A resource for studying the evolution of the spinach downy mildew pathogen. Mol. Plant-Microbe Interact. 31: 1230-1231.

Fischer, S., Brunk, B. P., Chen, F., Gao, X., Harb, O. S., Iodice, J. B., et al. 2011. Using OrthoMCL to Assign Proteins to OrthoMCL-DB Groups or to Cluster Proteomes Into New Ortholog Groups. Current Protocols in Bioinformatics. John Wiley & Sons, Inc., Hoboken, NJ, U.S.A.

Fletcher, K., Gil, J., Bertier, L. D., Kenefick, A., Wood, K. J., Zhang, L., Reyes-Chin-Wo, S., Cavanaugh, K., Tsuchida, C., Wong, J., and Michelmore, R. 2019. Genomic signatures of heterokaryosis in the oomycete pathogen *Bremia lactucae*. Nat. Commun. 10:2645.

Fletcher, K., Klosterman, S. J., Derevnina, L., Martin, F., Bertier, L. D., Koike, S., Reyes-Chin-Wo, S., Mou, B., and Michelmore, R. 2018. Comparative genomics of downy mildews reveals potential adaptations to biotrophy. BMC Genomics 19:851.

Gamboa-Meléndez, H., Huerta, A. I., and Judelson, H. S. 2013. bZIP transcription factors in the oomycete *phytophthora infestans* with novel DNA-binding domains are involved in defense against oxidative stress. Eukaryot. Cell 12:1403-1412.

Gaulin, E., Pel, M. J. C., Camborde, L., San-Clemente, H., Courbier, S., Dupouy, M. A., Lengellé, J., Veyssiere, M., Le Ru, A., Grandjean, F., Cordaux, R., Moumen, B., Gilbert, C., Cano, L. M., Aury, J. M., Guy, J., Wincker, P., Bouchez, O., Klopp, C., and Dumas, B. 2018. Genomics analysis of *Aphanomyces* spp. identifies a new class of oomycete effector associated with host adaptation. BMC Biol. 16:43.

Görg, M., Ploch, S., Kruse, J., Kummer, V., Runge, F., Choi, Y. J., and Thines, M. 2017. Revision of *Plasmopara* (Oomycota, Peronosporales) parasitic to *Impatiens*. Mycol. Prog. 16:791-799.

Grünwald, N. 2012. Genome sequences of *Phytophthora* enable translational plant disease management and accelerate research. Can. J. Plant Pathol. 34:13-19.

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. 2013. QUAST: Quality assessment tool for genome assemblies. Bioinformatics 29:1072-1075.

Haas, B. J. B. J., Kamoun, S., Zody, M. C., Jiang, R. H. Y. R. H. Y., Handsaker, R. E. R. E., Cano, L. M. L. M., Grabherr, M., Kodira, C. D., Raffaele, S., Torto-Alalibo, T., Bozkurt, T. O., Ah-Fong, A. M., Alvarado, L., Anderson, V. L., Armstrong, M. R., Avrova, A., Baxter, L., Beynon, J., Boevink, P. C., Bollmann, S. R., Bos, J. I., 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., Jones, J. D., 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., 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., 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., 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.

Hua, C., Meijer, H. J., de Keijzer, J., Zhao, W., Wang, Y., and Govers, F. 2013. GK4, a G-protein-coupled receptor with a phosphatidylinositol phosphate kinase domain in *Phytophthora infestans*, is involved in sporangia development and virulence. Mol. Microbiol. 88:352-370.

Jackman, S. D., Vandervalk, B. P., Mohamadi, H., Chu, J., Yeo, S., Hammond, S. A., Jahesh, G., Khan, H., Coombe, L., Warren, R. L., and Birol, I. 2017. ABySS 2.0: Resource-efficient assembly of large genomes using a Bloom filter. Genome Res. 27:768-777.

Jiang, R. H. Y., de Bruijn, I., Haas, B. J., Belmonte, R., Löbach, L., Christie, J., van den Ackerveken, G., Bottin, A., Bulone, V., Díaz-Moreno, S. M., Dumas, B., Fan, L., Gaulin, E., Govers, F., Grenville-Briggs, L. J., Horner, N. R., Levin, J. Z., Mammella, M., Meijer, H. J., Morris, P., Nusbaum, C., Oome, S., Phillips, A. J., van Rooyen, D., Rzeszutek, E., Saraiva, M., Secombes, C. J., Seidl, M. F., Snel, B., Stassen, J. H., Sykes, S., Tripathy, S., van den Berg, H., Vega-Arreguin, J. C., Wawra, S., Young, S. K., Zeng, Q., Dieguez-Uribeondo, J., Russ, C., Tyler, B. M., and van West, P. 2013. Distinctive expansion of potential virulence genes in the genome of the oomycete fish pathogen *Saprolegnia parasitica*. PLoS Genet. 9:e1003272.

Judelson, H. S. 2012. Dynamics and innovations within oomycete genomes: Insights into biology, pathology, and evolution. Eukaryot. Cell 11: 1304-1312.

Judelson, H. S. 2017. Metabolic diversity and novelties in the oomycetes. Annu. Rev. Microbiol. 71:21-39.

Judelson, H. S., Shrivastava, J., and Manson, J. 2012. Decay of genes encoding the oomycete flagellar proteome in the downy mildew *Hyaloperonospora arabidopsidis*. PLoS One 7:e47624.

Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. 2012. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res. 40:D109-D114.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., and Drummond, A. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647-1649.

Kemen, E., Gardiner, A., Schultz-Larsen, T., Kemen, A. C., Balmuth, A. L., Robert-Seilaniantz, A., Bailey, K., Holub, E., Studholme, D. J., Maclean, D., and Jones, J. D. 2011. Gene gain and loss during evolution of obligate parasitism in the white rust pathogen of *Arabidopsis thaliana*. PLoS Biol. 9:e1001094.

Kobayashi, M., Hiraka, Y., Abe, A., Yaegashi, H., Natsume, S., Kikuchi, H., Takagi, H., Saitoh, H., Win, J., Kamoun, S., and Terauchi, R. 2017. Genome analysis of the foxtail millet pathogen *Sclerospora graminicola* reveals the complex effector repertoire of graminicolous downy mildews. BMC Genomics 18:897.

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. 2001. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. J. Mol. Biol. 305:567-580.

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.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23:2947-2948.

Lévesque, C. A., Brouwer, H., Cano, L., Hamilton, J. P., Holt, C., Huitema, E., Raffaele, S., Robideau, G. P., Thines, M., Win, J., Zerillo, M. M., Beakes, G. W., Boore, J. L., Busam, D., Dumas, B., Ferriera, S., Fuerstenberg, S. I., Gachon, C. M., Gaulin, E., Govers, F., Grenville-Briggs, L., Horner, N., Hostetler, J., Jiang, R. H., Johnson, J., Krajaejun, T., Lin, H., Meijer, H. J., Moore, B., Morris, P., Phuntmart, V., Puiu, D., Shetty, J., Stajich, J. E., Tripathy, S., Wawra, S., van West, P., Whitty, B. R., Coutinho, P. M., Henrissat, B., Martin, F., Thomas, P. D., Tyler, B. M., De Vries, R. P., Kamoun, S., Yandell, M., Tisserat, N., and Buell, C. R. 2010. Genome sequence of the necrotrophic plant pathogen *Pythium ultimum* reveals original pathogenicity mechanisms and effector repertoire. Genome Biol. 11:R73.

Li, H., and Durbin, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R., and 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078-2079.

Links, M. G., Holub, E., Jiang, R. H. Y., Sharpe, A. G., Hegedus, D., Beynon, E., Sillito, D., Clarke, W. E., Uzuhashi, S., and Borhan, M. H. 2011. De novo sequence assembly of *Albugo candida* reveals a small genome relative to other biotrophic oomycetes. BMC Genomics 12:503.

Liu, X., Ng, C., and Ferenci, T. 2000. Global adaptations resulting from high population densities in *Escherichia coli* cultures. J. Bacteriol. 182: 4158-4164.

Lohse, M., Drechsel, O., and Bock, R. 2007. OrganellarGenomeDRAW (OGDRAW): A tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. Curr. Genet. 52: 267-274.

Lu, S., Chen, L., Tao, K., Sun, N., Wu, Y., Lu, X., Wang, Y., and Dou, D. 2013. Intracellular and extracellular phosphatidylinositol 3-phosphate produced by *Phytophthora* species is important for infection. Mol. Plant 6:1592-1604.

Mayer, C., Leese, F., and Tollrian, R. 2010. Genome-wide analysis of tandem repeats in *Daphnia pulex*—A comparative approach. BMC Genomics 11:277.

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

Meijer, H. J. G., Hassen, H. H., and Govers, F. 2011. *Phytophthora infestans* has a plethora of phospholipase D enzymes including a subclass that has extracellular activity. PLoS One 6:e17767.

Meijer, H. J. G., Mancuso, F. M., Espadas, G., Seidl, M. F., Chiva, C., Govers, F., and Sabidó, E. 2014. Profiling the secretome and extracellular proteome of the potato late blight pathogen *Phytophthora infestans*. Mol. Cell. Proteomics 13:2101-2113.

Mulder, N. J., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Binns, D., Bradley, P., Bork, P., Bucher, P., Cerutti, L., Copley, R., Courcelle, E., Das, U., Durbin, R., Fleischmann, W., Gough, J., Haft, D., Harte, N., Hulo, N., Kahn, D., Kanapin, A., Krestyaninova, M., Lonsdale, D., Lopez, R., Letunic, I., Madera, M., Maslen, J., McDowall, J., Mitchell, A., Nikolskaya, A. N., Orchard, S., Pagni, M., Ponting, C. P., Quevillon, E., Selengut, J., Sigrist, C. J., Silventoinen, V., Studholme, D. J., Vaughan, R., and Wu, C. H. 2005. InterPro, progress and status in 2005. Nucleic Acids Res. 33:D201-D205.

Naim, Y. B., Falach-Block, L., Ben-Daniel, B. H., and Cohen, Y. 2019. Host range of *Peronospora belbahrii*, causal agent of basil downy mildew, in Israel. Eur. J. Plant Pathol. Published online.

Nayaka, S. C., Shetty, H. S., Satyavathi, C. T., Yadav, R. S., Kishor, P. B. K., Nagaraju, M., Anoop, T. A., Kumar, M. M., Kuriakose, B., Chakravartty, N., Katta, A. V. S. K. M., Lachagari, V. B. R., Singh, O. V., Sahu, P. P., Puranik, S., Kaushal, P., and Srivastava, R. K. 2017. Draft genome sequence of *Sclerospora graminicola*, the pearl millet downy mildew pathogen. Biotechnol. Rep. (Amst.) 16:18-20.

Nguyen Ba, A. N., Pogoutse, A., Provart, N., and Moses, A. M. 2009. NLStradamus: A simple hidden Markov model for nuclear localization signal prediction. BMC Bioinformatics 10:202.

Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. 2011. SignalP 4.0: Discriminating signal peptides from transmembrane regions. Nat. Methods 8:785-786.

Price, A. L., Jones, N. C., and Pevzner, P. A. 2005. De novo identification of repeat families in large genomes. Bioinformatics 21 (Suppl 1):i351-i358.

R Development Core Team. 2013. R: A language and environment for statistical computing. R Found. Stat. Comput. 1:409.

Raffaele, S., Farrer, R. A., Cano, L. M., Studholme, D. J., MacLean, D., Thines, M., Jiang, R. H., Zody, M. C., Kunjeti, S. G., Donofrio, N. M., Meyers, B. C., Nusbaum, C., and Kamoun, S. 2010. Genome evolution following host jumps in the Irish potato famine pathogen lineage. Science 330:1540-1543.

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

Rodenburg, S. Y. A., Seidl, M. F., de Ridder, D., and Govers, F. 2018. Genome-wide characterization of Phytophthora infestans metabolism: A systems biology approach. Mol. Plant Pathol. 19:1403-1413.

Roy, S., Kagda, M., and Judelson, H. S. 2013. Genome-wide prediction and functional validation of promoter motifs regulating gene expression in spore and infection stages of *Phytophthora infestans*. PLoS Pathog. 9:e1003182.

Sawaya, S., Bagshaw, A., Buschiazzo, E., Kumar, P., Chowdhury, S., Black, M. A., and Gemmell, N. 2013. Microsatellite tandem repeats are abundant in human promoters and are associated with regulatory elements. PLoS One 8:e54710.

Schena, L., Cardle, L., and Cooke, D. E. L. 2008. Use of genome sequence data in the design and testing of SSR markers for Phytophthora species. BMC Genomics 9:620.

Schornack, S., Huitema, E., Cano, L. M., Bozkurt, T. O., Oliva, R., Van Damme, M., Schwizer, S., Raffaele, S., Chaparro-Garcia, A., Farrer, R., Segretin, M. E., Bos, J., Haas, B. J., Zody, M. C., Nusbaum, C., Win, J., Thines, M., and Kamoun, S. 2009. Ten things to know about oomycete effectors. Mol. Plant Pathol. 10:795-803.

Seidl, M. F., and Van den Ackerveken, G. 2019. Activity and phylogenetics of the broadly occurring family of microbial Nep1-like proteins. Annu. Rev. Phytopathol. 57:367-386.

Seidl, M. F., Van den Ackerveken, G., Govers, F., and Snel, B. 2011. A domain-centric analysis of oomycete plant pathogen genomes reveals unique protein organization. Plant Physiol. 155:628-644.

Sharma, R., and Thines, M. 2015. FastQFS–a tool for evaluating and filtering paired-end sequencing data generated from high throughput sequencing. Mycol. Prog. 14:60.

Sharma, R., Xia, X., Cano, L. M., Evangelisti, E., Kemen, E., Judelson, H., Oome, S., Sambles, C., van den Hoogen, D. J., Kitner, M., Klein, J., Meijer, H. J., Spring, O., Win, J., Zipper, R., Bode, H. B., Govers, F., Kamoun, S., Schornack, S., Studholme, D. J., Van den Ackerveken, G., and Thines, M. 2015a. Genome analyses of the sunflower pathogen *Plasmopara halstedii* provide insights into effector evolution in downy mildews and *Phytophthora*. BMC Genomics 16:741.

Sharma, R., Xia, X., Riess, K., Bauer, R., and Thines, M. 2015b. Comparative genomics including the early-diverging smut fungus *Ceraceosorus bombacis* reveals signatures of parallel evolution within plant and animal pathogens of fungi and oomycetes. Genome Biol. Evol. 7:2781-2798.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. 2015. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210-3212.

Soanes, D. M., Richards, T. A., and Talbot, N. J. 2007. Insights from sequencing fungal and oomycete genomes: What can we learn about plant disease and the evolution of pathogenicity? Plant Cell 19:3318-3326.

Soylu, E. M., and Soylu, S. 2003. Light and electron microscopy of the compatible interaction between *Arabidopsis* and the downy mildew pathogen *Peronospora parasitica*. J. Phytopathol. 151:300-306.

Spanu, P. D. 2012. The genomics of obligate (and nonobligate) biotrophs. Annu. Rev. Phytopathol. 50:91-109.

Sreenu, V. B., Ranjitkumar, G., Swaminathan, S., Priya, S., Bose, B., Pavan, M. N., Thanu, G., Nagaraju, J., and Nagarajaram, H. A. 2003. MICAS: A fully automated web server for microsatellite extraction and analysis from prokaryote and viral genomic sequences. Appl. Bioinformatics 2: 165-168.

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.

Tarailo-Graovac, M., and Chen, N. 2009. Using RepeatMasker to identify repetitive elements in genomic sequences. Chapter 4, Unit 4.10 in: Current Protocols in Bioinformatics. John Wiley & Sons, Inc., Hoboken, NJ, U.S.A.

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

Thines, M. 2019. An evolutionary framework for host shifts - jumping ships for survival. New Phytol. 224:605-617.

Thines, M., and Choi, Y. J. 2016. Evolution, diversity, and taxonomy of the Peronosporaceae, with focus on the genus *Peronospora*. Phytopathology 106:6-18.

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

Thines, M., Telle, S., Ploch, S., and Runge, F. 2009. Identity of the downy mildew pathogens of basil, coleus, and sage with implications for quarantine measures. Mycol. Res. 113:532-540.

Tian, M., Win, J., Savory, E., Burkhardt, A., Held, M., Brandizzi, F., and Day, B. 2011. 454 genome sequencing of *Pseudoperonospora cubensis* reveals effector proteins with a QXLR translocation motif. Mol. Plant-Microbe Interact. 24:543-553.

Tyler, B. M., Kale, S. D., Wang, Q., Tao, K., Clark, H. R., Drews, K., Antignani, V., Rumore, A., Hayes, T., Plett, J. M., Fudal, I., Gu, B., Chen, Q., Affeldt, K. J., Berthier, E., Fischer, G. J., Dou, D., Shan, W., Keller, N. P., Martin, F., Rouxel, T., and Lawrence, C. B. 2013. Microbe-independent entry of oomycete RxLR effectors and fungal RxLR-like effectors into plant and animal cells is specific and reproducible. Mol. Plant-Microbe Interact. 26:611-616.

Tyler, B. M., Tripathy, S., Zhang, X., Dehal, P., Jiang, R. H. Y., Aerts, A., Arredondo, F. D., Baxter, L., Bensasson, D., Beynon, J. L., Chapman, J., Damasceno, C. M., Dorrance, A. E., Dou, D., Dickerman, A. W., Dubchak, I. L., Garbelotto, M., Gijzen, M., Gordon, S. G., Govers, F., Grunwald, N. J., Huang, W., Ivors, K. L., Jones, R. W., Kamoun, S., Krampis, K., Lamour, K. H., Lee, M. K., McDonald, W. H., Medina, M., Meijer, H. J., Nordberg, E. K., Maclean, D. J., Ospina-Giraldo, M. D., Morris, P. F., Phuntumart, V., Putnam, N. H., Rash, S., Rose, J. K., Sakihama, Y., Salamov, A. A., Savidor, A., Scheuring, C. F., Smith, B. M., Sobral, B. W., Terry, A., Torto-Alalibo, T. A., Win, J., Xu, Z., Zhang, H., Grigoriev, I. V., Rokhsar, D. S., and Boore, J. L. 2006. *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. Science 313:1261-1266.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., and Rozen, S. G. 2012. Primer3—New capabilities and interfaces. Nucleic Acids Res. 40:e115.

van den Hoogen, D. J., Meijer, H. J. G., Seidl, M. F., and Govers, F. 2018. The ancient link between G-protein-coupled receptors and C-terminal phospholipid kinase domains. MBio 9:e02119-e17.

Vinces, M. D., Legendre, M., Caldara, M., Hagihara, M., and Verstrepen, K. J. 2009. Unstable tandem repeats in promoters confer transcriptional evolvability. Science 324:1213-1216.

Voglmayr, H. 2003. Phylogenetic relationships of *Peronospora* and related genera based on nuclear ribosomal ITS sequences. Mycol. Res. 107: 1132-1142.

Yang, X., Zhao, W., Hua, C., Zheng, X., Jing, M., Li, D., Govers, F., Meijer, H. J., and Wang, Y. 2013. Chemotaxis and oospore formation in *Phytophthora sojae* are controlled by G-protein-coupled receptors with a phosphatidylinositol phosphate kinase domain. Mol. Microbiol. 88: 382-394.

Ye, W., Wang, Y., Shen, D., Li, D., Pu, T., Jiang, Z., Zhang, Z., Zheng, X., Tyler, B. M., and Wang, Y. 2016. 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.

Zeng, L. R., Vega-Sánchez, M. E., Zhu, T., and Wang, G. L. 2006. Ubiquitination-mediated protein degradation and modification: An emerging theme in plant-microbe interactions. Cell Res. 16: 413-426.

Zerbino, D. R., and Birney, E. 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18:821-829.