Ancestral Chromosomes for Family Peronosporaceae Inferred from a Telomere-to-Telomere Genome Assembly of *Peronospora effusa*

Kyle Fletcher,¹ Oon-Ha Shin,² Kelley J. Clark,^{3,4} Chunda Feng,⁴ Alexander I. Putman,⁵ James C. Correll,⁴ Steven J. Klosterman,³ Allen Van Deynze,^{2,†} and Richard W. Michelmore^{1,6,†}

¹ The Genome Center, University of California, Davis, CA, U.S.A.

² Seed Biotechnology Center, Department of Plant Sciences, University of California, Davis, CA, U.S.A.

³ United States Department of Agriculture–Agricultural Research Station, 1636 East Alisal Street, Salinas, CA, U.S.A.

⁴ Department of Entomology & Plant Pathology, University of Arkansas, Fayetteville, AR, U.S.A.

⁵ Department of Microbiology and Plant Pathology, University of California, Riverside, CA, U.S.A.

⁶ Departments of Plant Sciences, Molecular & Cellular Biology, Medical Microbiology & Immunology, University of California, Davis, CA, U.S.A.

Accepted 24 February 2022.

Downy mildew disease of spinach, caused by the oomycete *Peronospora effusa*, causes major losses to spinach production. In this study, the 17 chromosomes of *P. effusa* were assembled telomere-to-telomere, using Pacific Biosciences high-fidelity reads. Of these, 16 chromosomes are complete and gapless; chromosome 15 contains one gap bridging the nucleolus organizer region. This is the first telomere-to-telomere genome assembly for an oomycete. Putative centromeric regions were identified on all chromosomes. This new assembly enables a reevaluation of the genomic composition of *Peronospora* spp.; the assembly was almost double the size and contained more repeat sequences than previously reported for any *Peronospora* species. Genome fragments consistently underrepresented in six previously reported assemblies of *P. effusa* typically encoded repeats. Some genes annotated as encoding effectors were organized into multigene clusters on several chromosomes. Putative effectors were annotated on 16 of the 17 chromosomes. The intergenic distances between annotated genes were consistent with compartmentalization of the genome into gene-dense and gene-sparse regions. Genes encoding putative effectors were enriched in gene-sparse regions. The near-gapless assembly revealed apparent horizontal gene transfer from Ascomycete fungi. Gene order was highly conserved between *P. effusa* and the genetically oriented assembly of the oomycete *Bremia lactucae*; high levels of synteny were also detected with *Phytophthora sojae*. Extensive synteny between phylogenetically distant species

suggests that many other oomycete species may have similar chromosome organization. Therefore, this assembly provides the foundation for genomic analyses of diverse oomycetes.

Keywords: effector clusters, genome organization, horizontal gene transfer, loss of heterozygosity, oomycete, synteny

The Peronosporaceae is a family in the Oomycota division with hundreds of described species, including *Phytophthora* and *Peronospora* spp. (Thines and Choi 2016). Many Peronosporaceae species are downy mildews, which are obligate biotrophs that are typically host-specific plant pathogens. Other members of the Peronosporaceae family are hemibiotrophic, paraphyletic *Phytophthora* spp. that typically have a wider host range than downy mildew species (Abad et al. 2019). Phylogenetic analyses indicated that adaptation to obligate biotrophy from hemibiotrophic ancestors has occurred at least twice within the family, such that downy mildews are polyphyletic (Bourret et al. 2018; Fletcher et al. 2018, 2019). *Peronospora* is the largest downy mildew genus and contains over 400 species (Thines and Choi 2016), including *Peronospora effusa*, the subject of the current analysis.

Downy mildew, caused by *P. effusa*, threatens spinach production because infection results in leaves unsuitable for sale and consumption. The disease is favored by cool temperatures and high humidity and manifests as yellow lesions with blue-gray sporulation, often on the abaxial surface of spinach leaves (Kandel et al. 2019; Klosterman 2016). In conventional spinach production, the disease can be managed with the use of fungicides and resistant cultivars; however, for organic production, deployment of resistant spinach cultivars is the only effective control available. Multiple resistance genes have been identified in the host but the pathogen can rapidly adapt and evade host detection (Correll et al. 2011; Kandel et al. 2019). To date, there are 19 named races of *P. effusa*; however, isolates with variation in virulence phenotype are continuously identified and the biology behind the emergence of new races is not clearly understood (International Working Group on Peronospora 2021). Previous genomic investigations have yielded fragmented and repeat-sparse genome assemblies (Feng et al. 2018a; Fletcher et al. 2018; Klein et al. 2020). The number of chromosomes the genome of *P. effusa* contained, the number and genomic

[†]Corresponding authors: R. W. Michelmore; rwmichelmore@ucdavis.edu, and A. Van Deynze; avandeynze@ucdavis.edu

The raw reads, genome assembly, and annotation are available under NCBI BioProject PRJNA745455.

Funding: The work was supported by a Novozymes Inc. Endowed Chair in Genomics grant to R. W. Michelmore and a grant received by A. Van Deynze from the California Seed Association Spinach Committee.

*The e-Xtra logo stands for “electronic extra” and indicates that supplementary figures and supplementary tables are published online.

The author(s) declare no conflict of interest.

Copyright © 2022 The Author(s). This is an open access article distributed under the CC BY-NC-ND 4.0 International license.

distribution of genes encoding virulence factors, and how much of the genome is conserved or variable between isolates remained unknown. Genome assemblies are important resources for determining the molecular basis for changes in virulence, resulting in the emergence of new pathotypes and providing informative loci for population studies.

The current study aimed to assemble a high-quality genome of *P. effusa*, using Pacific Biosciences high-fidelity (PacBio HiFi) reads, and examine the evolutionary relationship of this genome with those of other oomycetes to gain additional insights into its virulence. Seventeen telomere-to-telomere (T2T) chromosomes were assembled. Sixteen were assembled as single, gapless contigs. One was assembled into two contigs with a gap spanning the nucleolar organizer region (NOR). Comparison of this assembly with previous studies found the genome to be larger and more repetitive than anticipated. There was a high degree of synteny between *P. effusa*, *Phytophthora sojae*, and *Bremia lactucae*. Putative centromeric regions were identified through comparative genomics with *P. sojae*. This is a landmark genome assembly for oomycete genomics.

RESULTS

Peronospora effusa isolate UA202013 was collected in Thermal, CA, in 2020. The isolate was phenotyped for virulence on the standardized differential set of spinach lines to determine whether its phenotype was consistent with a defined race. This isolate was virulent on the universal susceptible line Viroflay and differential lines NIL3 and NIL6 and showed intermediate virulence on differential lines NIL5 and NIL4. The isolate was avirulent on the other six lines tested (Table 1). This virulence phenotype does not match any of those of the 19 denominated race isolates (International Working Group on *Peronospora* 2021).

The genome of isolate UA202013 of *P. effusa* was sequenced to 31× whole-genome coverage and assembled into 17 T2T chromosomes. Contigs for the final assembly came from several intermediate assemblies, as detailed below. The assembly was assessed by benchmarking universal single-copy orthologs (BUSCO) (Simão et al. 2015) as being 99.2% complete (232 of 234 models), of which 0.9% (2 models) were duplicated, using a protist-specific database. Sixteen of the chromosomes are complete and gapless; chromosome 15 contains a single gap bridging the NOR. Chromosome assembly length ranged from 1.78 Mb (chromosome 9) to 8.42 Mb (chromosome 1) (Fig. 1). Putative centromeric regions were identified on all 17 chromosomes (Fig. 1); putative centromeric regions of 15 chromosomes were identified as containing at

least one Copia-like transposon (CoLT) element (Supplementary Fig. S1). CoLT elements were previously hypothesized to be enriched in oomycete centromeres (Fang et al. 2020). The centromeric regions on chromosomes 1 and 15 lacked CoLT elements but were inferred by synteny with *Phytophthora sojae* centromeres (Supplementary Fig. S1). In total, 12 *Peronospora effusa* centromeric regions were syntenic with experimentally validated *Phytophthora sojae* centromeres (Fang et al. 2020), though some gene rearrangements were present around the centromeric regions of *Peronospora effusa* chromosomes 1, 6, 7, 8, and 15 (Supplementary Fig. S1). Additional analysis is required to determine the precise centromeric boundaries of each *P. effusa* chromosome. Telomeric satellite sequences (5'-CCCTAAA-3') were detected at both ends of every chromosome and ranged from 321 to 1,283 bp. The GC content of the *P. effusa* assembly was 48.6%, ranging from 49.2 to 48.1% per chromosome. Within 100-kb windows, the GC content ranged from 39.1 to 56.9% (Fig. 1C; Table 2).

The genome was 53.7% repetitive. The most common repeat elements were long-terminal-repeat retrotransposons (LTR-RTs), comprising 40.2% of the genome. Chromosomes ranged from 43.1 to 67.1% repetitive sequence. Repeat content did not correlate with chromosome length (Table 2) (Pearson's correlation [PC] = -0.38). The total genome size of 58.6 Mb was approximately twice the size of previous assemblies (23.9 to 32.1 Mb). Repeat content calculated for previous assemblies ranged from 14.4 to 39.0% (Table 3). Therefore, most of the size difference between the new assembly and previous assemblies is due to the repeat content assembled. Putative centromeres colocalized with repetitive regions (Fig. 1). Pairwise alignment of 66,858 clustered repeat elements demonstrated that many repeats shared high identity (≥98%) (Supplementary Fig. S2). High-identity alignments of repeats were found both within and between chromosomes (Supplementary Fig. S3). Nucleotide differences between LTR pairs of RTs in the new assembly were less diverged than in previous *P. effusa* assemblies (Supplementary Fig. S4). Therefore, similar LTR-RTs assembled in short read assemblies were likely chimeric, pairing LTRs from different elements. This suggests that short-read assemblies are inadequate to date LTR expansion events accurately in oomycetes. Interestingly, the LTR divergence in the new HiFi assembly was comparable with that reported for *Phytophthora ramorum* and *P. sojae*, both assemblies generated by Sanger sequencing (Tyler et al. 2006).

In total, 9,745 genes were annotated in the *Peronospora effusa* genome, with an average of 166 genes per megabase. The number of genes annotated per chromosome ranged from 237 to 1,482 (Table 2) and correlated with chromosome length (PC = 0.97). The density of genes ranged from 108 to 219 genes per megabase. Gene density did not correlate with chromosome length (PC = 0.29). On average, 27.6% of each 100-kb bin encoded genic sequence. Gene density was inversely related to repeat density (Fig. 1). Of these genes, 307 were annotated as effectors. These included 98 genes encoding crinkler effectors, of which 41 had evidence of expression in data from a prior study (Kandel et al. 2020), and 209 genes encoding RXLR effectors, of which 186 had evidence of expression (at least one read aligned from the previously reported RNA-sequencing [RNA-seq] study) (Supplementary Tables S1 and S2). Differential gene expression analysis was not feasible because early timepoints did not have enough reads aligned to *P. effusa*. Effectors were encoded on every chromosome except chromosome 11 (Table 2). The effector count per chromosome was weakly correlated with the chromosome length (PC = 0.82). Some effectors were organized as clusters of genes with high identity (Fig. 1; Supplementary Fig. S5). Three clusters of crinkler annotations on chromosome 1, spanning approximately 141, 34, and 84 kb, were detected. There was a fourth, larger crinkler cluster on chromosome 10, spanning 523 kb. Clusters of proteins annotated as having an RXLR-EER

Table 1. Disease incidence and reactions of spinach differential cultivars to isolate UA202013 of *Peronospora effusa*

Differential line	Disease incidence (%) ^a	Disease reactions ^b
Viroflay	100.0	+
NIL5	26.7 to 70.0	±
NIL3	100.0	+
NIL4	15.4 to 64.3	±
NIL6	100.0	+
NIL1	0.0	-
NIL2	0.0	-
Pigeon	0.0	-
Caladonia	0.0	-
Meerkat	0.0	-
Hydrus	0.0	-

^a Disease incidence is based on the percentage of plants (cotyledon or true leaves) evaluated showing either chlorosis, sporulation, or both.

^b Differential lines were considered susceptible (+) if disease incidence was greater than 85% or resistant (-) if disease incidence was less than 15%. The disease reaction was considered intermediate (±) if disease incidence was between 15 and 85%.

Fig. 1. Architecture of the 17 chromosomes of *Peronospora effusa*. **A**, Scaled chromosomes and ticks showing lengths. Chromosome number was designated based on syntenic linkage groups with *Bremia lactucae*. **B**, Dark gray boxes indicate putative positions of centromeric regions, some of which are syntenic with *Phytophthora sojae*. All subsequent tracks are plotted in 100-kb bins, with a 25-kb step. **C**, Heatmap of AT content, ranging from 0.4 to 0.6. **D**, Gene density depicted in red, repeat density depicted in black. The dark gray background shows 0 to 0.5, light gray 0.5 to 1. **E**, Effector annotations. The first row are WY effectors; red ticks indicate that the annotation contains a signal peptide prediction, RXLR and EER motif, and WY domain; blue ticks indicate annotation with a signal peptide and RXLR motif and WY domain; green ticks indicate annotation with signal peptide and WY domain; black ticks indicate a WY domain annotation with no signal peptide. In the second row are genes annotated as encoding RXLR-EER proteins with a signal peptide (all black). In the third row are crinkler annotations with (red) and without (black) predicted signal peptides. **F**, Clusters of high-identity crinkler (blue) and RXLR-EER/WY (red) annotation. **G**, Orthology detected with 33 other oomycetes. Blue indicates ortholog detected with other species, red indicates genes unique to *Peronospora effusa*, white indicates no annotation in a bin. **H**, Coverage of bin in six smaller *P. effusa* assemblies as determined by BLASTn. Colors indicate different assemblies; all are Illumina based except the bottom light blue, which was a hybrid PacBio and Illumina assembly. **I**, Normalized read depth of each bin. Long reads of UA202013 (31x) were plotted in red. Short reads of five other isolates were plotted in black. The background indicated 0 to 1x (dark gray) and 1 to 3x (light gray). **J**, Density of structural variant calls inferred by FreeBayes, using the long reads of UA202013, plotted with scale log base = 0.5. **K**, Single-nucleotide variant calls from alignments against UA202013 for alignments of UA202013 long reads and short reads of Pfs12, Pfs13, R13, Pfs14, and R14 plotted with scale log base = 0.5.

motif, WY domain, or both (RXLR-EER/WY) were identified on chromosomes 2 (310 kb), 4 (1.1 Mb), 5 (745 kb), 13 (53 kb), 14 (249 kb), 15 (245 kb), and 17 (168 kb) and three independent clusters on chromosome 16 (182, 360, and 378 kb) (Fig. 1; Supplementary Fig. S5). The RXLR-EER/WY clusters on chromosomes 4 and 17 spanned centromeres that shared synteny with *Phytophthora sojae* (Supplementary Fig. S1). Interchromosome, lower-identity alignments were detected for RXLR-EER/WY annotations but not crinkler annotations (Supplementary Fig. S5). These interchromosomal alignments could be due to gene duplication and transposition. Crinkler clusters were composed of protein sequences sharing higher identity to one another than RXLR-EER/WY clusters (Supplementary Fig. S6). Interchromosomal alignments of effector proteins rarely shared high identity compared with intra-chromosomal alignments (Supplementary Fig. S6).

To determine whether *Peronospora effusa* contained species-specific gene models, orthologous proteins were identified between 38 annotated genome assemblies representing 33 different oomycete species (Supplementary Table S3), including three genome assemblies of *P. effusa*. This analysis assigned 8,739 of the 9,745 annotations of *P. effusa* UA202013 to 6,425 multispecies orthogroups. An additional 725 UA202013 annotations assigned to 483 orthogroups were unique to *P. effusa* and were annotated in previous assemblies. The remaining 281 *P. effusa* UA202013 annotations had no orthology assigned with any oomycete assembly, including those of the assemblies of *P. effusa* isolates R13 and R14; however, short reads generated from five *P. effusa* isolates (R13, R14, Pfs12, Pfs13, and Pfs14) did align to these loci in UA202013, suggesting that these loci are present in the genomes

of these other isolates. Therefore, 1,006 genes were detected as unique to *P. effusa*. These genes were distributed across all chromosomes (Supplementary Fig. S7); 174 were annotated as effectors and a further 54 overlapped effector clusters.

The complete T2T genome assembly allowed for the investigation of genes unique to *P. effusa* as potential horizontal gene transfer (HGT) events. An obvious signal for HGT was two neighboring gene models in the subtelomeric region of chromosome 2 of *P. effusa*. The translated sequence of these two gene models was highly homologous to several fungal proteins. The fungal homologs also neighbored one another in their respective genome assemblies (Fig. 2A). One of these genes encoded a metallophosphatase domain (MPD); the second gene did not encode a detectable conserved domain. Both were annotated in previous *P. effusa* assemblies (Fletcher et al. 2018). The genes were embedded in a repeat-rich region on chromosome 2, approximately 22 and 64 kb away from flanking genes (Fig. 2A). Phylogenetically, the *P. effusa* genes nested within fungal protein sequences, closest to the sunflower pathogen *Diaporthe helianthi* and dieback-associated fungi *Valsa sordida* and *V. malicola* (Fig. 2B and C). The aligned amino acid identity between the fungal and oomycete homologs in the clade ranged from 61.3 to 75.7%. The flanking genes on *P. effusa* chromosome 2 were assigned orthology with other oomycete proteins and branched as expected phylogenetically (Fig. 2D and E). Flanking gene blocks were conserved between oomycete assemblies but rearrangements of the gene blocks were apparent (Fig. 2A). A paralogous MPD-encoding gene was present on *P. effusa* chromosome 15. There was no paralog for the second HGT candidate gene in this

Table 2. Summary statistics of chromosomes (Chr) assembled for *Peronospora effusa* isolate UA202013

Chr	Assembled length (bp)	Total gaps (bp)	Repeats (%)	AT content (%)	Genes annotated	Genes/Mb	Effectors annotated	Assembled length (bp)	
								5' telomere	3' telomere
1	8,417,896	0	51.5	51.4	1,482	176	70	706	685
2	3,956,849	0	53.5	51.7	596	151	23	597	958
3	3,727,724	0	49.5	50.9	666	179	10	723	616
4	4,163,955	0	51.4	51.3	751	180	28	803	695
5	3,019,954	0	57.3	51.5	393	130	16	680	1,039
6	4,409,438	0	54.9	51.5	692	157	14	1,283	649
7	2,262,584	0	67.1	51.5	261	115	6	756	703
8	2,258,044	0	56.3	51.6	334	148	5	502	656
9	1,779,483	0	49.7	51.8	315	177	3	682	523
10	6,355,493	0	48.6	51.3	1,185	186	26	669	652
11	2,566,028	0	43.1	51.8	563	219	0	707	867
12	2,778,632	0	54.8	50.8	577	208	4	785	505
13	2,196,953	0	63.0	51.2	237	108	17	603	321
14	2,153,540	0	61.1	51.6	310	144	12	801	619
15	2,369,557	100	62.3	51.8	281	119	24	525	654
16	2,944,237	0	54.1	51.3	479	163	27	795	784
17	3,218,066	0	52.8	51.9	623	194	22	708	553
Total	58,578,433	100	53.7	51.4	9,745	166	307	...	...

Table 3. Comparative assembly statistics of *Peronospora effusa* isolate UA202013 with previously published assemblies of *P. effusa*

Isolate	Scaffold N ₅₀	Scaffold count	Contig N ₅₀	Contig count	Assembly size (Mb)	Gaps (%)	Repeats (%) ^a	Sequencing technologies	Reported gene models	Study
UA202013	3.7 Mb	17	3.7 Mb	18	58.6	<0.001	53.7	PacBio HiFi	9,745	This study
R13	72 kb	784	48 kb	1,472	32.2	0.26	34.5	Illumina	8,607	Fletcher et al. 2018
R14	61 kb	880	52 kb	1,275	30.8	0.56	32.0	Illumina	8,571	Fletcher et al. 2018
Pfs1	33.7 kb	6,762	33 kb	6,777	32.1	0.02	39.0	Illumina + PacBio	13,277 ^b	Klein et al. 2020
Pfs12	21.3 kb	4,061	20.9 kb	4,505	25.2	0.2	17.5	Illumina	Not reported	Feng et al. 2018a
Pfs13	18.1 kb	4,387	17.8 kb	4,939	24	0.26	14.4	Illumina	Not reported	Feng et al. 2018a
Pfs14	22.1 kb	4,020	21.3 kb	4,488	24.9	0.22	16.6	Illumina	Not reported	Feng et al. 2018a

^a Repeat content calculated in this study, using a repeat library calculated from *P. effusa* UA202013.

^b Inflated gene model count due to some repeat elements annotated as genes.

region. The region on chromosome 15 was flanked by a block of oomycete genes and five copies of a repeated sequence with homology to the flanking repeat on chromosome 2 (Supplementary Fig. S8). Additional MPD-encoding homologous genes were annotated on *P. effusa* chromosomes 1 and 13. The intergenic

distance flanking these genes was less than 500 bp. These genes were assigned orthology with other oomycete MPD-encoding genes and branched as expected phylogenetically (Fig. 2). Therefore, MPD genes on chromosomes 1 and 13 were not acquired by HGT. Genes encoding an MPD domain on chromosomes 2 and

Fig. 2. Evidence of horizontal gene transfer in the genome of *Peronospora effusa*. **A**, Gene content of several fungal and oomycete contigs or scaffolds. The top four sequences are true fungi and show conserved gene order of two to three genes (linked by black lines). Two of those genes were found on chromosome 2 of *P. effusa*, colored blue and embedded in repeats. Red blocks indicate oomycete genes. Orthologs of oomycete genes were found on single contigs or scaffolds in the genomes of other oomycetes. The gene order of *Phytophthora sojae* and *Pythium oligandrum* was similar. Rearrangements were visible when the *Phytophthora sojae* and *Pythium oligandrum* sequences were compared with *Peronospora effusa* or *Plasmopara viticola*. **B to E**, Neighbor-joining trees of the horizontally acquired genes and the flanking gene. **B**, Only found in *Peronospora effusa*, the protein sequence was nested within fungal sequences. **C**, Several genes encoding proteins containing metallophosphatase domains were found in the *P. effusa* genome. Homologs on chromosomes 2 and 15 were nested in fungal sequences; no orthologs were annotated in oomycete assemblies. Orthologs in the genomes of other oomycetes were identified for the homologs on *P. effusa* chromosomes 1 and 13. **D and E**, Phylogenetic analysis of the glancing proteins produced expected topologies.

15 were likely acquired through a single HGT event from true fungi, followed by duplication. The read-depth of these genes in UA202013 and five other previously sequenced isolates was consistent with all four genes being present in all isolates sequenced (Supplementary Table S4). Evidence for expression was found for the three genes of fungal origin and the two MPD-encoding genes of oomycete ancestry (Supplementary Table S5).

Alignment of six previous *P. effusa* genome assemblies confirmed that most missing sequence from these previous assemblies was repeat-encoding and not genic. The percentage of each 100-kb window covered by BLASTn alignments generated with other *P. effusa* assemblies was calculated, revealing poor representation in multiple windows across all chromosomes (Fig. 1H; Supplementary Fig. S9). This result is consistent with the previous assemblies being smaller than the new draft assembly (Table 3). Between 1,140 and 1,655 genes annotated in *P. effusa* UA202013 were absent in the assemblies of other isolates based on BLASTn alignments. As previously noted, the inflated gene count of *P. effusa* isolate Pfs1 is likely due to repeats annotated as genes (Klein et al. 2020). In total, 3,620 genes annotated in *P. effusa* UA202013 were not covered by BLASTn alignments from at least one other assembly of *P. effusa*. Only 121 genes were inferred as absent from all other assemblies of *P. effusa* (Supplementary Fig. S10). Further investigation revealed that 98 of the 121 proteins lacked orthology with any oomycete protein and, therefore, may be unique to isolate UA202013 of *P. effusa* or annotation artifacts.

Approximately 1,000 more genes were annotated in UA202013 than had previously been annotated in R13 and R14 (Table 3). The results of the multispecies orthology analysis reported above were used to investigate whether there were genes unique to UA202013 compared with R13 and R14. Of the 6,425 orthogroups containing an annotated gene in UA202013, 6,263 also contained annotations from both R13 and R14 (Supplementary Fig. S11). However, 131 annotations from UA202013 were present in 97 multispecies orthogroups that did not contain R13 or R14 annotations. This suggests that most of these extra 1,000 genes are not due to novel genes annotated in the assembly of UA202013. Furthermore, short reads generated from other isolates aligned to these 131 gene-models, suggesting that the loci were also present in the other isolates.

Further analysis of orthologs showed that the increased gene count for UA202013 is mostly due to collapsed paralogs in the other assemblies. There were several instances of multiple annotations from UA202013 being assigned to orthogroups which contained fewer or no annotations from R13 or R14 (Supplementary Fig. S12). Analysis of read depth supported the correct number of paralogs in the UA202013 assembly. For the assemblies of R13 and R14, the genes assigned to these orthogroups had high coverage in their respective raw reads, indicating that the paralogs were collapsed in the short-read assemblies (Supplementary Fig. S13). Therefore, although short-read assemblies are still a valuable resource for obtaining the majority of the gene space of these organisms, they are not sufficient to resolve the highly repetitive architecture or sufficient to resolve paralogous genes in the genome of *P. effusa*.

Read-depth analysis showed that no copy number variation of whole chromosomes was present in isolates UA202013, Pfs12, Pfs13, Pfs14, R13, and R14 (Fig. 1I; Supplementary Fig. S14). The only significant discrepancy in coverage was observed at the gap on chromosome 15 when aligning reads of UA202013 back to the UA202013 assembly. This high coverage sequence encoded ribosomal subunits, consistent with it being the NOR. This increased coverage across the NOR was detected in the reads of other isolates (Fig. 1I; Supplementary Fig. S14). Additional significant peaks and troughs of coverage were recorded for the

five other isolates. Read depth differences correlated with the phenotypes, suggesting that the genomes of Pfs13 and R13 were similar, as were the genomes of Pfs12, Pfs14, and R14 (Supplementary Fig. S14). Interestingly, few reads for all five other isolates aligned to the effector cluster on chromosome 14 of UA202013 (Supplementary Figs. S14 and S15). This result suggests that the effectors encoded may be absent or highly diverged in UA202013 versus the other five isolates. The four effectors annotated in UA202013 contained a signal peptide prediction, an EER motif, two LWY domains, a partial GLY zipper domain, and a partial electron transport complex protein domain (Supplementary Fig. S16). Between all four, the N terminus was more conserved than the C terminus. The genomes of R13 and R14 each contained a single homologous annotation, though in both cases the partial electron transport complex protein domain was not assembled as in the UA202013 gene models (Supplementary Fig. S16). No significant depth variation was observed for other effector clusters in the UA202013 assembly (Supplementary Fig. S14). Therefore, the cluster of putative effectors on chromosome 14 contains candidates for proteins underlying the difference in virulence of isolate UA202013 compared with the other isolates.

A global analysis of homozygosity found that, in four of the six isolates, 100-kb windows which overlapped clusters of genes encoding candidate effectors were significantly more homozygous than windows which did not. Heterozygosity of single-nucleotide polymorphisms (SNPs) in overlapping 100-kb windows ranged from 0 to 1.24%. The average heterozygosity of windows for isolates UA202013, Pfs13, and R13 was higher than for Pfs12, Pfs14, and R14 (Fig. 3). For isolates UA202013, Pfs12, Pfs14, and R14, windows which overlapped clusters of effector genes were significantly more homozygous than windows which did not (Fig. 3). To account for the effect of coverage on

Fig. 3. Windows of overlapping clusters of effector-encoding genes are more homozygous for four of six isolates of *Peronospora effusa*. Across the genome assembly of UA202013, 100-kb windows were partitioned into those that overlapped with clusters of effector genes and those that did not. For UA202013, Pfs12, Pfs14, and R14, the windows overlapping clusters of effector-encoding genes were significantly more homozygous than for the other two isolates (Wilcoxon rank-sum test $P < 1.8 \times 10^{-6}$). A multivariate linear model was fitted to investigate the effect of coverage on heterozygosity, which indicated that clusters of genes encoding effectors are associated with reduced heterozygosity.

homozygosity, a multivariate linear model was fitted using heterozygosity as the dependent variable, and coverage and presence of clusters of effectors as independent variables. Models for each isolate suggested that, after accounting for coverage, the presence of effector clusters was significantly associated with reduced heterozygosity (Supplementary Table S6). Therefore, in four of the isolates, including UA202013, it appears that regions of the genome encoding clusters of effectors are more homozygous compared with other regions.

Loss of heterozygosity (LOH) on chromosome 6 may underlie the difference in virulence between races 12 and 14 of *P. effusa*. Races 12 and 14 have an identical virulence phenotype, except that race 14 is virulent on the differential cultivar Pigeon and race 12 is not (International Working Group on *Peronospora* 2021). Comparisons of polymorphisms called from all six isolates in this analysis revealed that the race 14 isolates R14 and Pfs14 and race 12 isolate Pfs12 were more similar to one another genome wide than to other isolates. Among these three isolates, clustering of SNPs consistently grouped Pfs14 with Pfs12, suggesting that they were genetically closer to each other than either was to R14 (Supplementary Fig. S17). However, the two independently collected race 14 isolates (Feng et al. 2018b; Fletcher et al. 2018) were both highly homozygous compared with Pfs12 over two regions on chromosomes 6 and 14 (Supplementary Figs. S18 and S19). Therefore, LOH in one or both of these regions may have resulted in the loss of an avirulence allele detectable by Pigeon. In the assembly of UA202013, seven genes were annotated as encoding effectors in the 750-kb region on chromosome 6. No effector-encoding genes were annotated in the chromosome 14 region (Supplementary Fig. S18 and S19). Therefore, one or more of these genes are good candidates for determining the avirulence phenotype.

To establish whether the genome of *P. effusa* was compartmentalized into gene-dense and gene-sparse regions, the distribution of intergenic distances between all genes was calculated. This

revealed a bimodal distribution for both the 5' and 3' flanking regions (Fig. 4; Supplementary Fig. S20). Most genes (80.5%) were flanked by 100 bp to 5 kb of intergenic sequence on both sides. Approximately 17% of annotations (1,675 of 9,745) had intergenic distances greater than 5 kb on one side, of which 221 had >5 kb of intergenic sequence flanking on both sides. This included 142 and 40 effector annotations, meaning 8.5 and 18% of the genes in these compartments, respectively, were effectors (Fig. 4). Genome wide, 3.1% of the genes are annotated as effectors; therefore, effectors were enriched in the gene-sparse regions of the assembly (Wilcoxon signed-rank test; 5' intergenic distance $P = 3.54 \times 10^{-24}$, 3' intergenic distance $P = 7.93 \times 10^{-26}$). The upper limit of this analysis was larger than a previously reported similar analysis of *P. effusa* Pfs1 (Klein et al. 2020), likely because the Pfs1 assembly was more fragmented than the T2T UA202013 assembly. Genes encoding effectors have been proposed as being enriched in gene-sparse regions in the genomes of some other phytopathogens, possibly facilitating more rapid diversification of these compartments compared with gene-dense compartments of the genome (Dong et al. 2015; Frantzeskakis et al. 2019). Genes with >50-kb intergenic regions were found on every chromosome, suggesting that every chromosome of *P. effusa* has compartmentalized, gene-sparse regions, potentially enriched for genes involved in pathogenicity.

Comparative genomics demonstrated that the chromosomes of *P. effusa* were highly syntenic with assemblies of other oomycetes (Figs. 4B and 5A). The genome of *B. lactucae* has been ordered into 19 linkage groups (LGs) (Fletcher et al. 2021), 15 of which were colinear with complete chromosomes of *P. effusa* (Fig. 5A). The other four *B. lactucae* LG scaffolds were syntenic with two chromosomes of *P. effusa*: LG 16 and LG 19 of *B. lactucae* were syntenic with chromosome 16 of *P. effusa* and LG 17 and LG 18 of *B. lactucae* were syntenic with chromosome 17 of *P. effusa*. Gene content was also conserved with *Phytophthora sojae* (Fig. 5B), though synteny of some scaffolds was found with

Fig. 4. Intergenic distances between annotated genes of *Peronospora effusa* isolate UA202013. For each gene, the distances 5' to the next gene are plotted on the x axis and the distances 3' to the next gene are plotted on the y axis. The top left plot shows the intergenic distance between all annotations, showing that the majority of genes cluster between 100 and 5,000 bp on either side (colored red). Some genes had larger intergenic distances on either or both sides. The other plots show individual chromosomes; a black x marks genes annotated as encoding effectors.

Fig. 5. The high synteny of *Peronospora effusa* chromosomes with other phylogenetically distinct oomycetes. Synteny is demonstrated by linking orthologous genes, colored by *P. effusa* (red) chromosome assignment against **A**, 19 genetically oriented scaffolds of the downy mildew *Bremia lactucae* (blue) and **B**, 15 scaffolds larger than 1 Mb of the hemibiotroph *Phytophthora sojae* (green). Plots are scaled to reflect the physical sizes of the chromosomal sequences. **C**, Maximum-likelihood phylogenetic tree calculated from the concatenated alignment of 18 BUSCO proteins surveyed in 31 assemblies of 27 Peronosporaceae spp. with *Globisporangium ultimum* (Pythiaceae) as an outgroup; multiple assemblies for *Peronospora effusa* and *Plasmopara viticola* were used. Branches colored red and blue indicate the two polyphyletic downy mildew clades. Species for which synteny was plotted are highlighted using the same color palette as in **A** and **B**. Parenthesized numbers indicate the *Phytophthora* clade a species has been assigned to. The most recent common ancestor of *Peronospora effusa* and *B. lactucae* is inferred as common to all downy mildews and *Phytophthora* clades 1 to 5 (arrowed CAi). The most recent common ancestor of *P. effusa* and *Phytophthora sojae* is inferred to be deeper in the phylogeny (arrowed CAii). Scale indicates the mean number of amino acid substitutions per site. Branch support was calculated from 1,000 bootstraps. This figure reused data from Figure 3 in Fletcher et al. (2019).

multiple chromosomes (*P. sojae* 116 = *Peronospora effusa* chromosomes 1 and 13; *Phytophthora sojae* 115 = *Peronospora effusa* chromosomes 6 and 10; and *Phytophthora sojae* 117 = *Peronospora effusa* chromosomes 9, 11, and 12). It remains unresolved whether these are true chromosomal fusions or assembly errors in *Phytophthora sojae*. Comparison with a more recent long-read assembly of *P. sojae* did not resolve this because the assembly was too fragmented (Supplementary Fig. S1).

Phylogenetic analysis of 18 conserved single-copy BUSCO genes supports polyphyly of the downy mildews. *Peronospora effusa* and *B. lactucae* are located in two distinct clades of biotrophic Peronosporaceae; *P. effusa* appears to be phylogenetically closer to *Phytophthora agathidicida*, which represents *Phytophthora* clade 5, while *B. lactucae* branched closer to clade 1 *Phytophthora* spp. such as *P. infestans* (Fig. 5C). This is consistent with previous evaluations (Bourret et al. 2018; Fletcher et al. 2018, 2019; McCarthy and Fitzpatrick 2017). Therefore, the most recent common ancestor of *Peronospora effusa* and *B. lactucae* is inferred as common to all downy mildews and *Phytophthora* spp. clades one to five (Fig. 5C). The most recent common ancestor of *Peronospora effusa* and *Phytophthora sojae* (*Phytophthora* clade 7) is more ancient than the common ancestor of *Peronospora effusa* and *B. lactucae*. Given the high levels of synteny between these three species (Fig. 5A and B), it is likely that the 17-chromosome architecture of *P. effusa* and *B. lactucae* will be ancestral to hundreds of other species in the Peronosporaceae family.

DISCUSSION

This report describes the first T2T assembly of an oomycete. In all, 16 of the 17 chromosomes of *P. effusa* are gapless; only chromosome 15 contains one gap, which likely encodes a highly repetitive ribosomal DNA (rDNA) array. This assembly will be a key resource for comparative genomics with all other oomycetes and will advance the study of many important pathogens.

Comparative genomics revealed a high level of synteny between *P. effusa* and the genetically orientated genome assembly of *B. lactucae* (Fletcher et al. 2021). The *P. effusa* chromosomes were numbered to match the corresponding chromosome-scale scaffolds of *B. lactucae*, reported previously (Fletcher et al. 2021), because they have highly conserved single-copy gene contents, similar gene orders along their chromosomes, and likely the same number of chromosomes (Fig. 5A). Therefore, the common ancestor of *P. effusa* and *B. lactucae* may have had a similar chromosomal architecture. Phylogenetic analysis indicated that downy mildews are polyphyletic (Fig. 5C); the most recent common ancestor of biotrophic *P. effusa* and *B. lactucae* is also common to every downy mildew, as well as *Phytophthora* clades 1 through 5 (Bourret et al. 2018; Fletcher et al. 2018, 2019). Given the high levels of synteny with *Phytophthora sojae*, which is a clade 7 species, the proposed ancestral architecture for many chromosomes is likely to be rooted deeper in the oomycete phylogeny than the last common ancestor of downy mildews (Fig. 5C). Consequently, many downy mildew and *Phytophthora* spp. may have the same ancestral chromosome configurations; therefore, this T2T assembly of *Peronospora effusa* will serve as a foundational reference for the study of Peronosporaceae species and other oomycetes.

Structural variation can be seen around the centromeres when comparing *P. effusa* to *Phytophthora sojae* (Supplementary Fig. S1) and in the subtelomeric region of chromosome 2 between *P. effusa*, *Plasmopara viticola*, and *Phytophthora sojae* (Fig. 2A). The interchromosomal structural variation observed between *Peronospora effusa* and *Phytophthora sojae* (Fig. 5B) may be due to misassemblies in *Phytophthora sojae* or unique structural rearrangements in the evolutionary history of *P. sojae*.

Interestingly, one misassembly in the *P. sojae* assembly that joins sequences syntenic to *Peronospora effusa* chromosomes 6 and 10 has previously been reported (Fang et al. 2020). The generation of additional chromosome-scale assemblies will be fundamental in determining how much structural variation exists within and between oomycete species, and this T2T assembly of *P. effusa* provides the foundation for such investigations.

This new genome assembly of *P. effusa* was used to analyze the accuracy of previous *P. effusa* draft assemblies. A striking difference is the size of the assemblies, with isolate UA202013 being nearly double the size of previous assemblies (Table 3). The increase in assembly size is attributed to the inability of short reads and the high error rate of long reads to assemble high-identity repeats and paralogs (Supplementary Figs. S12 and S13). High-identity repeats were collapsed, misassembled, or absent in previous short-read assemblies. Collapsed repeat sequences are reported to be present in genome assemblies of other oomycetes (Fletcher et al. 2019) and may result in unreliable analysis, including analyzing bursts of LTR activity by nucleotide variation. Calculated LTR divergence since insertion on the assembly of *P. effusa* UA202013 compared with R13 and R14 (generated from short reads) showed that the resulting profiles differed between long- and short-read assemblies (Supplementary Fig. S4). This result is consistent with short-read assemblies generating chimeric LTR-RTs with inflated divergence in their LTRs. Therefore, estimations of LTR divergence since insertion in oomycete assemblies generated from short reads should be treated with care (Fletcher et al. 2019).

Short-read assemblies of downy mildews are valuable in identifying most but not all of the gene space. The number of annotated genes for *P. effusa* isolate UA202013 is only slightly higher than isolates R13 and R14 (9,745, 8,607, and 8,571, respectively). More genes were annotated for isolate Pfs1 (Table 3), although many genes were reported to overlap repeats for Pfs1 (Klein et al. 2020); our identification of conserved domains characteristic of RTs in the Pfs1 annotations confirmed the inclusion of LTR-RTs in gene annotations. Orthology analysis demonstrated that the increased gene count for UA202013 was due to improved resolution of paralogs (Supplementary Fig. S12 and S13) and not due to the absence of single-copy genes from previous assemblies. These previously collapsed paralogs included putative crinkler proteins, explaining the increase of this class of effector. The generation of short-read oomycete assemblies is still informative for comparative analyses because most single-copy genes are recovered but paralogs may be underrepresented.

Short reads are also a valuable resource for comparative analysis between isolates to identify candidate regions containing molecular determinants of race phenotypes. Aligning reads of five other isolates showed that read depth, nucleotide variation, and heterozygosity correlated genome-wide with race phenotypes (Supplementary Figs. S14, S17, and S18). The two race 13 isolates sequenced independently across two studies are very similar, as are the two race 14 isolates. The SNP compositions of these isolates are consistent with somatic derivation from single founder genotypes. Race 14 was proposed to be derived from race 12 after an LOH event (Lyon et al. 2016). Our data support this and favor an LOH event in a region of chromosome 6, resulting in the loss of avirulence alleles (Supplementary Fig. S19). Interestingly, the two race 14 isolates have different lengths of homozygosity, suggesting that they were independent LOH events. Independent LOH events are consistent with Pfs12 (race 12 rather than race 14) and Pfs14 clustering closer to one another than Pfs14 to R14 (Supplementary Fig. S17).

Read-depth analysis also highlighted a cluster of high-identity LWY effectors of chromosome 14 that are unique to UA202013 compared with isolates representing races 12, 13, and 14. Therefore, these effectors may be recognized by spinach lines on which

UA202013 is avirulent, such as NIL2 (Table 1) (International Working Group on *Peronospora* 2021). Functional studies are required to test this hypothesis further. Other regions of the UA202013 assembly also had low read depth in the other isolates but did not overlap clusters of effectors (Supplementary Fig. S14).

The T2T assembly demonstrated that some candidate RXLR and crinkler (CRN) effectors were encoded by clusters of closely related genes (Fig. 1; Supplementary Fig. S5). Although clusters of RXLR and CRN effectors have been described on scaffolds of *Phytophthora* spp. (Schornack et al. 2009), they had not been reported in the previous fragmented assemblies of *Peronospora* spp. For four of the six isolates surveyed, these regions were more homozygous than other regions of the genome. Genomic organization of gene clusters encoding effectors may have consequences for the evolution of new virulence phenotypes. Recombination within tightly linked clusters, such as clusters encoding CRN effectors, will be less frequent than for unlinked genes. The larger clusters of RXLR effectors on the smaller chromosomes may have a higher frequency of recombination because there will be a higher density of crossovers on the smaller chromosomes. More frequent recombination within and between genes encoding effectors may result in the rapid evolution of novel virulence phenotypes. In contrast to larger clusters of genes encoding RXLR effectors, CRNs were tightly clustered on larger chromosomes. CRNs within a cluster were more similar than clustered RXLR effectors (Supplementary Fig. S6). Lower divergence implies that paralogous genes encoding CRNs may be diversifying more slowly than genes encoding RXLR effectors, reflecting different selection pressures. Interestingly, transient expression of 15 *Phytophthora sojae* CRNs in *Nicotiana benthamiana* showed that relatively few of the tested crinklers induced plant cell death (PCD); instead, most crinklers suppressed PCD, consistent with CRNs acting as PCD regulators, not inducers (Amaro et al. 2017; Shen et al. 2013).

The T2T genome assembly allowed an efficient search for HGT events by identifying an insertion of genes that are unique to *Peronospora effusa* that lacked orthology with annotations of other oomycete species. Several instances of HGT and subsequent duplication have been previously described for the oomycetes, identifying HGT from fungi or bacteria to oomycetes (Belbahri et al. 2008; McCarthy and Fitzpatrick 2016; Savory et al. 2015). These used comparative analysis of assemblies for multiple oomycete species or isolates and focused on *Phytophthora* and *Pythium* spp. Our comparison of a complete genome to assemblies of other oomycetes provided robust evidence for the recent HGT to an ancestor of *P. effusa*. Two genes have been acquired from phytopathogenic fungi on chromosome 2, one of which was duplicated on chromosome 15 (Fig. 2; Supplementary Fig. S7). The genes on chromosome 2 were also found as a pair in several fungal phytopathogens (Fig. 2). These genes were phylogenetically nested within phytopathogenic fungal sequences and were distinct from paralogs that had orthology to other oomycete species (Fig. 2); this provides strong support for HGT from a fungal species to an ancestor of *P. effusa*. The functions of these genes are unknown; the duplicated gene encoded an MPD. The acquisition and duplication of the MPD-encoding gene suggests a selective advantage (Savory et al. 2015). The large intergenic distances flanking the horizontally transferred genes means that they are in the gene-sparse regions of the genome. The MPD-encoding genes ancestral to oomycetes had intergenic distances less than 500 bp and, therefore, are in the gene-dense regions of the genome. Previous studies have proposed that genes encoded in gene-sparse regions of the genome evolve faster (Dong et al. 2015). Therefore, these horizontally acquired genes may be evolving differently from their homologs with oomycete ancestry.

In summary, the T2T assembly of *P. effusa* provides an important foundation for comparative genomics between oomycete species and for analysis of diversity within *P. effusa*. Future studies will determine how prevalent the 17-chromosome configuration is between diverse members of family Peronosporaceae and whether there are major structural rearrangements in some lineages. It will also be interesting to determine whether the size and chromosomal locations of effector clusters vary, especially between downy mildews, whose ancestors may have adjusted the number of effectors in their genome independently while adaptation to biotrophy occurred. Finally, this assembly provides sequences which can be tested through evolutionary, comparative, and functional analyses to determine what role they play in the success of *P. effusa*.

MATERIALS AND METHODS

Isolate propagation and virulence phenotyping.

Isolate UA202013 of *P. effusa* was collected from the spinach cultivar Sparrow in Thermal, CA on 23 March 2020 from a commercial production field. The isolate was propagated prior to DNA extraction using a method described previously (Feng et al. 2014). Sporangia were washed off infected leaves with cold filtered water (4°C) using a vortex mixer. The inoculum was diluted to 1×10^5 spores/ml and 50 ml was fine sprayed on a 2-week-old flat of Viroflay plants, a downy-mildew-susceptible spinach cultivar. The inoculated plants were put in a growth chamber where all programmable parameters were adjusted to simulate the average spinach-growing environment. The daily temperature slope was programmed to range between 13 and 18°C and relative humidity was maintained above 90% using manual misting with cold, filtered water. Light was on for 12 h per day with 150 $\mu\text{mol}/\text{m}^2/\text{s}$ intensity, except for during the first 24 h to induce sporangial germination in the dark. The plants were taken out of the growth chamber after 12 days to harvest sporangia.

Virulence phenotyping of this isolate was performed using differential lines from the International Working Group on *Peronospora* following a standardized protocol (Correll and du Toit 2018; Feng et al. 2014, 2018b), with minor modifications. Briefly, 20 to 30 seedlings of each cultivar, 2 weeks in age, were inoculated. Sporangia were filtered through four layers of cheesecloth, adjusted to a concentration of 10^5 sporangia/ml, and spray inoculated on the 11 differential cultivars using a Badger Basic Spray Gun, model 250. After inoculation, the plants were incubated in a dew chamber for 24 h, followed by a growth chamber for 5 days, and a final 24 h in the dew chamber to induce sporulation. The cotyledons and true leaves of each plant were evaluated for downy mildew disease symptoms 7 days postinoculation. Disease incidence (percentage) for each cultivar was assigned based on the number of plants showing chlorosis and sporulation out of the total number evaluated. Susceptible, resistant, and intermediate disease reaction evaluations were given based on disease incidence.

DNA extraction and library preparation.

Spore suspensions were centrifuged at $8,000 \times g$ for 5 min, the supernatant was removed, and pellets were frozen at -80°C . DNA was extracted from pellets (approximately 100 mg) using an OmniPrep kit (G-Biosciences, St. Louis, MO, U.S.A.) in 1 ml of lysis buffer (G-Biosciences) with 10 μl of proteinase K and digested for 1 h at 60°C with periodic mixing to keep spores suspended in the lysis buffer. Library construction and sequencing were done at the University of California, Davis Genome Center. DNA quantity and quality was verified with the Quantus fluorometer assay (QuantiFluor ONE dsDNA Dye; Promega Corp., Madison, WI, U.S.A.) and Femto Pulse Automated Pulsed Field (Agilent Technologies, Inc., Santa Clara, CA,

U.S.A.), respectively. Sheared genomic DNA (700 ng) was prepared with the Low Input DNA SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences, Menlo Park, CA, U.S.A.) and loaded into a single 8M SMRT cell on a Pacific Biosciences Sequel II.

Genome sequencing and assembly.

In total, 348,032 CCS reads were generated from the DNA extracted from the spore pellet of UA202013. PacBio HiFi reads were assembled with Hifiasm v0.14 (Cheng et al. 2021) and HiCanu v2.0 (Nurk et al. 2020), using default parameters. Hifiasm generated a 117-Mb assembly in 989 contigs, 11 of which were complete oomycete chromosomes presented in the final assembly. HiCanu generated a 155-Mb assembly in 2,040 contigs, 2 of which were complete chromosomes (chromosomes 5 and 9) also present in the Hifiasm assembly. These assemblies were filtered by identifying the top BLASTn (Altschul et al. 1990) hit in the NCBI nucleotide (nt) database (Fletcher and Micheltore 2018). Contigs with a top hit to an oomycete accession were compiled into filtered assemblies. The filtered Hifiasm assembly was 63.5 Mb in 99 contigs, and was 99.2% complete with 1.3% duplication, as calculated with BUSCO v2.0 (Simão et al. 2015) using the protist database. The filtered HiCanu assembly was 119 Mb in 1,083 contigs and was 99.6% completed with 74.8% duplication, as calculated with BUSCO v2.0 using the protist database. Reads were aligned to the HiCanu assembly with minimap2 v2.17-r954-dirty (Li 2018), using default parameters. The HiCanu assembly was used because it contained a higher level of duplication. This identified 239,965 reads as having their primary alignment to contigs which had a top BLASTn hit to an oomycete and, therefore, most likely originated from *P. effusa*. Assembling these reads with Hifiasm, again using default parameters, generated 57 contigs, including a 12th chromosome (chromosome 10) presented in the final assembly. Preliminary analysis was conducted by transferring annotations from *P. effusa* R13 (Fletcher et al. 2018) onto the new assembly using LiftOff v1.3.0 (Shumate and Salzberg 2020). Synteny of intermediate assemblies with the genetically oriented assembly of *B. lactucae* (GCA_004359215.2) (Fletcher et al. 2021) was determined using orthologs as calculated by OrthoFinder v2.2.1 (Emms and Kelly 2015). The initial 12 chromosomes were found to be colinear with *B. lactucae* LG-spanning scaffolds. Multiple *P. effusa* contigs were found to be colinear with other *B. lactucae* LG-spanning scaffolds. Long reads aligned to these *P. effusa* contigs with minimap2 v2.17-r954-dirty (Li 2018) were assembled independently again with Hifiasm v0.14 (Cheng et al. 2021) and complete chromosome contigs were assembled for two chromosomes (chromosomes 6 and 11). Chromosomes 4 and 16 were closed by aligning contigs generated by Hifiasm with contigs generated by HiCanu v2.0 (Nurk et al. 2020), extending the Hifiasm contigs by 8,619 and 12,694 bp, respectively, and adding telomeres to the sequence. Two telomeric contigs of 2,338,820 and 30,638 bp in length remained in the assembly, both terminating in repetitive rDNA sequences. They could not be aligned, nor could reads be found that bridged the sequences. Read alignment showed that the depth was high, indicating that a large repeat array was present, such that the gap was bridged with 100 unknown bases (N).

Genome annotation and comparative analysis.

Genome annotation was performed comparably with other *P. effusa* assemblies (Fletcher et al. 2018). A repeat library was defined with RepeatModeler v1.73 (Smit and Hubley 2008) and masked with RepeatMasker v4.0.9 (Smit et al. 2013). Repeats were annotated in six other *P. effusa* genome assemblies (Feng et al. 2018a; Fletcher et al. 2018; Klein et al. 2020) using the same repeat library. Repeat sequences of UA202013 identified by RepeatMasker were clustered using CD-Hit v4.8.1 (Li and Godzik 2006),

requiring 90% identity to the centroid sequences. Pairwise identity within each cluster was calculated using BLASTn v2.10.1 (Altschul et al. 1990). Inter- and intrachromosomal clusters were visualized using Circos v0.69-8 (Krzywinski et al. 2009). Nucleotide variation between LTRs assembled in UA202013 was calculated as previously described (Fletcher et al. 2019). Briefly, LTR-RTs were identified and annotated using LTRharvest v1.5.9 (Ellinghaus et al. 2008) and LTRdigest v1.5.9 (Steinbiss et al. 2009), respectively. Internal domains were clustered with VMatch v2.3.0 and within-cluster alignments of 5' and 3' LTRs were generated with ClustalO v1.2.0 (Sievers et al. 2011). The divergence between LTR pairs was calculated with BaseML v4.7b (Yang 2007). Previous LTR divergence results for *P. effusa*, *B. lactucae*, *Phytophthora infestans*, *P. ramorum*, and *P. sojae* were taken from a previous analysis (Fletcher et al. 2019). Results were plotted using R v4.0.1 base packages (R Development Core Team 2012).

Genes were annotated with MAKER v2.31.10 (Campbell et al. 2014) using a SNAP v2013-02-16 (Korf 2004) HMM model generated in a previous study (Fletcher et al. 2018). Additional effectors were predicted as previously described (Fletcher et al. 2018, 2019). Briefly, all open reading frames (ORFs) encoded in the genome were translated and surveyed for signal peptides using SignalP v3.0 and v4.1 (Bendtsen et al. 2004; Petersen et al. 2011). ORFs predicted as secreted but targeted to the mitochondria by TargetP v1.1 (Emanuelsson et al. 2000) were excluded from further analysis. ORFs with signal peptides were surveyed using regular-expression string searches to identify RXLR and EER motifs. All ORFs were surveyed using HMM models to identify CRN and (L)WY sequences. These were reconciled to identify proteins which encoded RXLR-EER motifs with (L)WY domains. Effector annotations from ORFs were reconciled with the MAKER annotations to ensure that no overlapping gene models were introduced. In the event of overlaps, the MAKER gene model was retained. Annotations were validated through orthology analysis with annotations of 37 other oomycete assemblies using OrthoFinder v2.2.1 (Emms and Kelly 2015). Intergenic distances between genes annotated for *Peronospora effusa* UA202013 were calculated from the resulting GFF file, excluding genes which overlapped on opposing strands. Enrichment of effectors was established by independent Wilcoxon signed-rank tests for both the 3' and 5' intergenic distances, classifying genes as encoding putative effectors or other genes.

Effector gene clusters were identified by clustering effector amino acid sequences using CD-Hit v4.8.1 (Li and Godzik 2006), requiring 40% identity to the centroid sequences for RXLR-EER/WY proteins and 70% for crinklers. Pairwise alignments within each cluster were calculated with BLASTp v2.10.1 (Altschul et al. 1990). Links were visualized with Circos v0.69-8 (Krzywinski et al. 2009). An effector gene cluster was manually called if the cluster contained three or more annotations on the same chromosome. Overlapping clusters were merged into a single cluster.

Expanded paralogs in UA202013 were identified by contrasting orthogroup gene counts, calculated with OrthoFinder v2.2.1 (Emms and Kelly 2015), for UA202013 with R13 and R14. These results were taken from the larger analysis containing annotations from 38 genome assemblies. Orthogroups in which UA202013 contained five more genes than either R13 or R14 were considered expanded. Read depth of the annotations in these orthogroups was calculated by aligning reads back to their respective assemblies using bwa v0.7.17 (Li 2013) or minimap2 v2.17-r954-dirty (Li 2018) and calculated with SAMtools v1.11 bedcov (Li et al. 2009). Stacked bar charts and dotplots were generated using R v4.0.1 (R Development Core Team 2012) and ggplot2 v3.3.3 (Wickham 2016). Completeness was calculated using BUSCO v2.0 with the protists_odb9 dataset (Simão et al. 2015). Translated single-copy orthologs were added to the 18-BUSCO-gene concatenated alignment previously reported

(Fletcher et al. 2019). Protein alignments for each gene were recalculated independent of one another using MAFFT v7.245 (Katoh and Standley 2013). Alignments were concatenated and a single tree was produced using RAXML v8.2.9 (Stamatakis 2014), with 1,000 bootstraps and the PROTGAMMAAUTO model. The tree was visualized in Geneious v8.0.5 (Kearse et al. 2012).

HGT was investigated in the genes unique to *P. effusa* isolate UA202013 using BLASTp v2.10.1 (Altschul et al. 1990) to search the NCBI nr database excluding oomycetes. Genes with a low e-value were tested phylogenetically. A ClustalW alignment was built containing the candidate HGT gene and top BLASTp hits from nonoomycete and oomycete species. This alignment was used to build a neighbor-joining consensus tree to determine whether the candidate gene was nested in nonoomycete sequences. A phylogenetic analysis was performed and visualized using Geneious v8.0.5 (Kearse et al. 2012). The same analysis was performed on flanking genes. Orthology of flanking genes was investigated and plotted for contiguous oomycete genome assemblies using R v4.0.1 (R Development Core Team 2012), ggplot2 v3.3.3 (Wickham 2016), and ggforce v0.3.3 (Pedersen 2019). Chromosomal fragments of *P. effusa* chromosomes 2 and 15 containing a duplicated gene were aligned against each other using BLASTn. The alignment results were visualized using R v4.0.1 (R Development Core Team 2012) and ggplot2 v3.3.3 (Wickham 2016).

Expression was investigated using RNA-seq reads from a previous study investigating the transcriptional response of spinach to infection with *P. effusa* (Kandel et al. 2020). All reads were downloaded from Sequence Read Archive BioProject PRJNA603356. Reads were aligned to a combined reference file containing the *P. effusa* assembly reported here and the NCBI RefSeq assembly of *Spinacia oleracea* (GCF_002007265.1) using STAR v2.7.9a (Dobin et al. 2013). Support for expression was inferred by the presence of at least one read aligning to the exon of the gene model. Differential gene expression analysis was not conducted because early timepoints in the compatible reaction had too few reads aligned to the genome assembly of *P. effusa*.

Synteny was plotted for the final *P. effusa* against *B. lactucae* and *Phytophthora sojae* by linking single-copy orthologs between the assemblies with Circos v0.69-8 (Krzywinski et al. 2009). Twelve putative *Peronospora effusa* centromeric regions were inferred through synteny with the more fragmented *Phytophthora sojae* assembly GCA_009848525.1, for which coordinates of 12 centromeres have been previously defined (Fang et al. 2020). An additional five putative centromeric regions (chromosomes 5, 7, 9, 12, and 13) were defined by identifying BLASTn hits (alignment length >500 bp) of a *P. sojae* CoLT previously hypothesized as diagnostic of oomycete centromeres (Fang et al. 2020). This alignment criterion was relaxed in comparison with similar surveys searching for *B. lactucae* and *P. citricola* centromeres (Fang et al. 2020). For all chromosomes in the assembly of *Peronospora effusa*, the coordinates of centromeric regions were expanded to the boundaries of flanking genes.

Previous *P. effusa* genome assemblies (Feng et al. 2018a; Fletcher et al. 2018; Klein et al. 2020) were aligned to the new genome assembly using BLASTn (Altschul et al. 1990), filtering for alignments with a minimum length of 2,500 bp and scoring $\geq 95\%$ identity. Genes for which 95% of the feature was within assembly-to-assembly alignments were reported as covered, regardless of the alignment length. Long reads of isolate UA202013 (this study) were aligned back to the assembly with Minimap2 v2.17-r954-dirty (Li 2018). Short reads of R13, R14 (Fletcher et al. 2018), pfs12, pfs13, and pfs14 (Feng et al. 2018a) were aligned with bwa v0.7.17 mem (Li 2013). Variants for long- and short-read alignments were called with FreeBayes v1.3.1-17-gaa2ace8 (Garrison and Marth 2012). Biallelic SNPs were used to cluster isolates genome wide and per chromosome. Homozygous references were

coded as 0, heterozygous as 0.5, and homozygous alternative as 1. Euclidean distances between isolates were calculated using R v4.0.1 base function dist() (R Development Core Team 2012). Results were visualized using Heatmap2 from the R package gplots v3.1.1 (Wames et al. 2020). LOH was calculated by comparing the difference of biallelic polymorphisms called as heterozygous by FreeBayes in 100-kb windows with 25-kb steps. Heterozygosity of windows which overlapped clusters of effector genes were compared with that of windows which did not using the nonparametric Wilcoxon signed-rank test with ggpubr v0.4.0. An analysis of variance model was used to explain the variance of heterozygosity using coverage and presence of cluster of effectors as independent variables, using lm() and anova() in R. The model can be expressed as heterozygosity $\approx$ coverage + presence of cluster effectors.

Boxplots were generated using ggplot2 (Wickham 2016) and ggpubr in R (R Development Core Team 2012). In particular, the search targeted a region which correlated with the previously hypothesized LOH between race 12 and race 14 isolates (Lyon et al. 2016).

For visualization, the genome was broken into 100-kb windows with a 25-kb step. The AT of each window was calculated using BEDTools2 v2.29.2 nuc (Quinlan 2014). The density of repeats, genes, structural variants (UA202013 only), and SNPs was calculated using BEDTools2 v2.29.2 (Quinlan 2014). The normalized read depth and coverage of previous draft genome assemblies per window were calculated using BEDTools2 v2.29.2 (Quinlan 2014). Per-window heterozygosity was calculated as the number of SNPs genotyped 0/1 by FreeBayes v1.3.1-17-gaa2ace8. All tracks were formatted as white-space separated files and plotted using Circos v0.69-8 (Krzywinski et al. 2009).

Data availability.

The raw reads, genome assembly, and annotation are available under NCBI BioProject PRJNA745455. All software used is described and cited in Materials and Methods.

ACKNOWLEDGMENTS

We thank A. Garcia-Llanos (University of California–Davis [UC Davis]) and A. Achieta (United States Department of Agriculture–Agricultural Research Station, Salinas, CA, U.S.A.) for DNA extractions, O. Nguyen (UC Davis) for library preparation and sequencing, H. Xu (UC Davis) for raw data submissions to NCBI, and E. Georgian (UC Davis) for editorial services. The sequencing was carried out by the DNA Technologies and Expression Analysis Cores at the UC Davis Genome Center, supported by NIH Shared Instrumentation Grant 1S10OD010786-01. The bioinformatic analysis was carried out using the UC Davis LSSC0 High Performance Computing cluster maintained by the UC Davis Bioinformatics Core.

LITERATURE CITED

- Abad, Z., Burgess, T., Bienapfl, J., Redford, A., Coffey, M., and Knight, L. 2019. IDphy: Molecular and morphological identification of *Phytophthora* based on the types. USDA APHIS PPQ Science and Technology Beltsville Laboratory, Beltsville, MD, U.S.A. <https://idtools.org/id/phytophthora/index.php>
- 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.
- Amaro, T. M. M., Thilliez, G. J. A., Motion, G. B., and Huitema, E. 2017. A perspective on CRN proteins in the genomics age: Evolution, classification, delivery and function revisited. *Front. Plant Sci.* 8:99.
- Belbahri, L., Calmin, G., Mauch, F., and Andersson, J. O. 2008. Evolution of the cutinase gene family: Evidence for lateral gene transfer of a candidate *Phytophthora* virulence factor. *Gene* 408:1-8.
- Bendtsen, J. D., Nielsen, H., von Heijne, G., and Brunak, S. 2004. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* 340:783-795.
- Bourret, T. B., Choudhury, R. A., Mehl, H. K., Blomquist, C. L., McRoberts, N., and Rizzo, D. M. 2018. Multiple origins of downy mildews and mito-nuclear discordance within the paraphyletic genus *Phytophthora*. *PLoS One* 13:e0192502.

- Campbell, M.S., Holt, C., Moore, B., and Yandell, M. 2014. Genome annotation and curation using MAKER and MAKER-P. *Curr. Protoc. Bioinf.* 48:4.11.11-14.11.39.
- Cheng, H., Concepcion, G. T., Feng, X., Zhang, H., and Li, H. 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* 18:170-175.
- Correll, J., and du Toit, L. 2018. Guidelines for spinach downy mildew: *Peronospora farinosa* f. sp. *spinaciae* (Pfs) (= *P. effusa*). Collaboration for Plant Pathogen Strain Identification, University of California, Davis, CA, U.S.A.
- Correll, J. C., Bluhm, B. H., Feng, C., Lamour, K., du Toit, L. J., and Koike, S. T. 2011. Spinach: Better management of downy mildew and white rust through genomics. *Eur. J. Plant Pathol.* 129:193-205.
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T. R. 2013. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15-21.
- Dong, S., Raffaele, S., and Kamoun, S. 2015. The two-speed genomes of filamentous pathogens: Waltz with plants. *Curr. Opin. Genet. Dev.* 35: 57-65.
- Ellinghaus, D., Kurtz, S., and Willhoeft, U. 2008. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinf.* 9:18.
- 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.
- Emms, D. M., and Kelly, S. 2015. OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16:157.
- Fang, Y., Coelho, M. A., Shu, H., Schotanus, K., Thimmappa, B. C., Yadav, V., Chen, H., Malc, E. P., Wang, J., Mieczkowski, P. A., Kronmiller, B., Tyler, B. M., Sanyal, K., Dong, S., Nowrousian, M., and Heitman, J. 2020. Long transposon-rich centromeres in an oomycete reveal divergence of centromere features in Stramenopila-Alveolata-Rhizaria lineages. *PLoS Genet.* 16:e1008646.
- Feng, C., Correll, J. C., Kammeijer, K. E., and Koike, S. T. 2014. Identification of new races and deviating strains of the spinach downy mildew pathogen *Peronospora farinosa* f. sp. *spinaciae*. *Plant Dis.* 98: 145-152.
- Feng, C., Lamour, K. H., Bluhm, B. H., Sharma, S., Shrestha, S., Dhillon, B. D. S., and Correll, J. C. 2018a. 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.
- Feng, C., Saito, K., Liu, B., Manley, A., Kammeijer, K., Mauzey, S. J., Koike, S., and Correll, J. C. 2018b. New races and novel strains of the spinach downy mildew pathogen *Peronospora effusa*. *Plant Dis.* 102: 613-618.
- 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 Micheltore, 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 Micheltore, R. 2018. Comparative genomics of downy mildews reveals potential adaptations to biotrophy. *BMC Genomics* 19:851.
- Fletcher, K., and Micheltore, R. 2018. From short reads to chromosome-scale genome assemblies. Pages 151-197 in: *Plant Pathogenic Fungi and Oomycetes. Methods in Molecular Biology*, vol. 1848. W. Ma and T. Wolpert, eds. Humana Press, New York, NY, U.S.A.
- Fletcher, K., Zhang, L., Gil, J., Han, R., Cavanaugh, K., and Micheltore, R. 2021. AFLAP: Assembly-free linkage analysis pipeline using k-mers from genome sequencing data. *Genome Biol.* 22:115.
- Frantzeskakis, L., Kusch, S., and Panstruga, R. 2019. The need for speed: Compartmentalized genome evolution in filamentous phytopathogens. *Mol. Plant Pathol.* 20:3-7.
- Garrison, E., and Marth, G. 2012. Haplotype-based variant detection from short-read sequencing. *arXiv* 1207.3907.
- International Working Group on Peronospora. 2021. Denomination of Pe: 18 and 19, two new races of downy mildew in spinach. *Plantum*, Gouda, The Netherlands. <https://ucanr.edu/blogs/blogcore/postdetail.cfm?postnum=46392#:~:text=Two%20new%20races%20of%20the,denominated%20as%20race%20Pe%3A%2018>
- Kandel, S. L., Hulse-Kemp, A. M., Stoffel, K., Koike, S. T., Shi, A., Mou, B., Van Deynze, A., and Klosterman, S. J. 2020. Transcriptional analyses of differential cultivars during resistant and susceptible interactions with *Peronospora effusa*, the causal agent of spinach downy mildew. *Sci. Rep.* 10:6719.
- Kandel, S. L., Mou, B., Shishkoff, N., Shi, A., Subbarao, K. V., and Klosterman, S. J. 2019. Spinach downy mildew: Advances in our understanding of the disease cycle and prospects for disease management. *Plant Dis.* 103:791-803.
- Katoh, K., and Standley, D. M. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* 30:772-780.
- 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.
- Klein, J., Neilen, M., van Verk, M., Dutilh, B. E., and Van den Ackerveken, G. 2020. Genome reconstruction of the non-culturable spinach downy mildew *Peronospora effusa* by metagenome filtering. *PLoS One* 15:e0225808.
- Klosterman, S. 2016. Spinach downy mildew—Threat, prevention and control. *Prog. Crop Consult* 1:12-15.
- Korf, I. 2004. Gene finding in novel genomes. *BMC Bioinf.* 5:59.
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J., and Marra, M. A. 2009. Circos: An information aesthetic for comparative genomics. *Genome Res.* 19:1639-1645.
- Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* 1303.3997.
- Li, H. 2018. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094-3100.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map (SAM) format and SAMtools. *Bioinformatics* 25:2078-2079.
- Li, W., and Godzik, A. 2006. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22:1658-1659.
- Lyon, R., Correll, J., Feng, C., Bluhm, B., Shrestha, S., Shi, A., and Lamour, K. 2016. Population structure of *Peronospora effusa* in the southwestern United States. *PLoS One* 11:e0148385.
- McCarthy, C. G. P., and Fitzpatrick, D. A. 2016. Systematic search for evidence of interdomain horizontal gene transfer from prokaryotes to oomycete lineages. *MSphere* 1:e00195-16.
- McCarthy, C. G. P., and Fitzpatrick, D. A. 2017. Phylogenomic reconstruction of the oomycete phylogeny derived from 37 genomes. *MSphere* 2:e00095-17.
- Nurk, S., Walenz, B. P., Rhie, A., Vollger, M. R., Logsdon, G. A., Grothe, R., Miga, K. H., Eichler, E. E., Phillippy, A. M., and Koren, S. 2020. HiCanu: Accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads. *Genome Res.* 30:1291-1305.
- Pedersen, T. L. 2019. ggforce: Accelerating 'ggplot2'. R package version 0.2.2. R Project for Statistical Computing, Vienna, Austria. <https://ggforce.data-imaginist.com/reference/ggforce-package.html>
- 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.
- Quinlan, A. R. 2014. BEDTools: The Swiss-army tool for genome feature analysis. *Curr. Protoc. Bioinf.* 47:11.12.11-11.12.34.
- R Development Core Team. 2012. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Savory, F., Leonard, G., and Richards, T. A. 2015. The role of horizontal gene transfer in the evolution of the oomycetes. *PLoS Pathog.* 11: e1004805.
- 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.
- Shen, D., Liu, T., Ye, W., Liu, L., Liu, P., Wu, Y., Wang, Y., and Dou, D. 2013. Gene duplication and fragment recombination drive functional diversification of a superfamily of cytoplasmic effectors in *Phytophthora sojae*. *PLoS One* 8:e70036.
- Shumate, A., and Salzberg, S. L. 2020. Liftoff: Accurate gene annotation mapping tool. *Bioinformatics* 37:1639-1643.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., and Higgins, D. G. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7:539.

- 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.
- Smit, A., and Hubley, R. 2008. RepeatModeler open-1.0. <http://www.repeatmasker.org>
- Smit, A., Hubley, R., and Green, P. 2013. RepeatMasker open-4.0. <http://www.repeatmasker.org>
- Stamatakis, A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312-1313.
- Steinbiss, S., Willhoeft, U., Gremme, G., and Kurtz, S. 2009. Fine-grained annotation and classification of de novo predicted LTR retrotransposons. *Nucleic Acids Res.* 37:7002-7013.
- Thines, M., and Choi, Y. J. 2016. Evolution, diversity, and taxonomy of the Peronosporaceae, with focus on the genus *Peronospora*. *Phytopathology* 106:6-18.
- 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. B., 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. G., Nordberg, E. K., Maclean, D. J., Ospina-Giraldo, M. D., Morris, P. F., Phuntumart, V., Putnam, N. H., Rash, S., Rose, J. K. C., Sakihama, Y., Salamov, A. A., Savidor, A., Scheuring, C. F., Smith, B. M., Sobral, B. W. S., 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.
- Warnes, G., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M., and Venables, B. 2020. gplots: Various R programming tools for plotting data. <https://rdrr.io/cran/gplots/>
- Wickham, H. 2016. ggplot2: Elegant graphics for data analysis. Springer, Cham, Switzerland.
- Yang, Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24:1586-1591.