S1 Text Nanopore sequencing and *de novo* assembly of the reference *P. sojae* genome.

Nanopore sequence statistics for *P. sojae* reads

To improve the genome assembly of the *P. sojae* reference strain (P6497), the long-read Oxford Nanopore Technology (ONT) was applied. In total, four flow cells were used (one for MinION, three for GridION), which produced high quality sequence data set: Approximately 251 x, 200 x, and 22 x coverage of the genome (95 Mb, ref. [1]) was obtained based on all reads, reads > 10 kb, and reads > 100 kb respectively (S1 Table). One flow cell (Flow cell 1) accompanying with MinION produced much fewer reads (S1 Table). This was probably because the flow cell used was an earlier version, and the genomic DNA (gDNA) quality for the second batch (Flow cell 2-4) that was associated with GridION may be better.

Whole genome assembly and quality estimation

As summarized in S5A Fig, to assemble the entire genome, ONT reads ≥10 kb from all four flow cells were used for an initial assembly employing SMARTdenovo (https://github.com/ruanjue/smartdenovo). The resulting contigs were corrected with two rounds of Racon [2] based on reads ≥10 kb derived from the ONT and the available PacBio sequencing (SRA: SRR10759964) data that were mapped to the assembly using GraphMap [3]. Afterwards, three rounds of Pilon (v1.22) [4] correction were performed employing "pseudo-Illumina" read pairs generated from the existing Sanger reads [1] (ftp://ftp-private.ncbi.nlm.nih.gov/pub/TraceDB/phytophthora_sojae/). The "pseudo-Illumina" read pairs were produced by a customized *Perl* script that generated a read pair (read length 150 bp, insert length 500 bp) from every 35th base along the Sanger reads, which resulted in approximately 20 million read pairs with a nominal coverage of ~70 x.

After getting the corrected assembly, mitochondrial sequences were removed based on the GC contents and BLAST search results. The assembly was finalized as Psojae2019.1.

To further increase the contiguity of the assembly, different scaffolding programs such as npScarf [5], SSPACE [6], and LINKS [7] were implemented. All scaffolding programs greatly increased the continuity, forming assemblies having 35-49 scaffolds, with N50 3.6-5 Mb (S3 Table).

Notably, we also attempted to link the contigs using the next generation optical mapping Bionano Saphyr system, which is based on a direct label stain (DLS). In total, 30 contigs were anchored, but most of them can only be partially covered by the Bionano molecules (S9A Fig and S6 File), except three contigs (contigs 48, 63 and 50) that were fully covered (S9E and S9F Fig). Eight contigs were suggested to join to form an assembly having 66 scaffolds (S9B-S9E Fig and S5 File). In addition, one contig (Contig 32) was identified as incorrectly assembled and the mis-assembly was resolved by breaking the contig at the divergent position (S9D Fig).

The overall coverage obtained in Bionano mapping is low (S6 File). This may in part be due to the gDNA extraction method (agarose embedding and releasing) adopted for Bionano, which was developed for mammalian systems and has not been well established for filamentous species. In addition, the DLE-1 enzyme used for labeling in the Bionano Saphyr system recommends a label density of 8 to 25 labels per 100 kb, while the density is about 9 in the *P. sojae* genome. This probably is a constraint of applying Bionano mapping to assemble “small” genomes or small contigs that do not have sufficient label density.

To evaluate the quality of the assemblies, we generated dot plot maps comparing the new genome assemblies with the existing Sanger assembly (S5B and S10 Figs). The contig-level assembly Psojae2019.1 is mostly colinear with the Sanger assembly, except

two major regions which are very repetitive (S5B Fig). In contrast, all *in silico* scaffolded assemblies demonstrated more conflicts when they were compared to the Sanger assembly (S10A-S10C Fig). As Bionano mapping only scaffolded limited numbers of contigs, it generated an assembly showing similar collinearity as that of the P2019.1 assembly (S10D Fig). Remarkably, contigs that were joined by Bionano were also combined by the Sanger assembly.

Further examination of the scaffolded assemblies revealed that the scaffolding program SSPACE engulfed 9 (out of 13) telomeres in the assembly (S3 Table). Later centromere mapping experiments showed that npScarf linked contigs that had centromeres. These indicate that SSPACE and npScarf generated substantial errors in the scaffolded assemblies. One aspect that we observed from the Bionano mapping is that several regions were duplicated in the new genome (S9B, S9C and S9E Fig). This could be caused by higher heterozygosity of those regions, given that *P. sojae* is diploid and the genome is not 100% homozygous [8]. Taken together, despite various scaffolders stitched contigs and generated assemblies with higher contiguity, they also generated gaps or other structural errors. Without coverage support from long-reads for the joints, we have opted to retain contig-level assembly in our study.

Metric of the Psojae2019.1 assembly

Statistically, the resulting assembly of the nuclear genome (Psojae2019.1) has a size of 86 Mb contained in 70 gap-free contigs, with a contig N50 of 2 Mb (S5C Fig). Many contigs proved to contain long tandem tRNA repeats (Fig 3 and S6 Fig), implying that tRNA repeats are one of the main obstacles challenging the genome assembly. Interestingly, most of the tRNA repeats are homogenous. Based on searching for the motif (TTTAGGG) that was proposed for oomycete telomere repeats [9], telomeric sequences were identified at single ends of 13 contigs (versus 6 ends in Sanger, S2 File). Analysis of

the new assembly revealed ~31 % of repeat sequences (versus ~27% of Sanger), and 24,415 protein-coding genes were predicted for the repeat-masked assembly (S5C Fig).

Supplementary references

1. Tyler BM, Tripathy S, Zhang X, Dehal P, Jiang RH, Aerts A, et al. *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. Science. 2006;313(5791):1261-6. Epub 2006/09/02. doi: 10.1126/science.1128796. PubMed PMID: 16946064.
2. Vaser R, Sović I, Nagarajan N, Šikić M. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. 2017;27:737-46.
3. Sović I, Šikić M, Wilm A, Fenlon SN, Chen S, Nagarajan N. Fast and sensitive mapping of nanopore sequencing reads with GraphMap. Nat Commun. 2016;7:11307.
4. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLOS ONE. 2014;9:e112963.
5. Cao MD, Nguyen SH, Ganesamoorthy D, Elliott AG, Cooper MA, Coin LJ. Scaffolding and completing genome assemblies in real-time with nanopore sequencing. Nat Commun. 2017;8:14515. Epub 2017/02/22. doi: 10.1038/ncomms14515. PubMed PMID: 28218240; PubMed Central PMCID: PMCPMC5321748.
6. Boetzer M, Pirovano W. SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. BMC Bioinformatics. 2014;15(1):211. doi: 10.1186/1471-2105-15-211. PubMed PMID: WOS:000338264200003.
7. Warren RL, Yang C, Vandervalk BP, Behsaz B, Lagman A, Jones SJ, et al. LINKS: Scalable, alignment-free scaffolding of draft genomes with long reads. Gigascience. 2015;4(1):35. Epub 2015/08/06. doi: 10.1186/s13742-015-0076-3. PubMed PMID: 26244089; PubMed Central PMCID: PMCPMC4524009.
8. Fletcher K, Gil J, Bertier LD, Kenefick A, Wood KJ, Zhang L, et al. Genomic signatures of heterokaryosis in the oomycete pathogen *Bremia lactucae*. Nat Commun. 2019;10(1):2645. Epub 2019/06/16. doi: 10.1038/s41467-019-10550-0. PubMed PMID: 31201315; PubMed Central PMCID: PMCPMC6570648.
9. Fulneckova J, Sevcikova T, Fajkus J, Lukesova A, Lukes M, Vlcek C, et al. A broad phylogenetic survey unveils the diversity and evolution of telomeres in eukaryotes. Genome Biol Evol. 2013;5(3):468-83. Epub 2013/02/12. doi: 10.1093/gbe/evt019. PubMed PMID: 23395982; PubMed Central PMCID: PMCPMC3622300.