Whole-Genome Sequence of the Phlox Powdery Mildew Pathogen *Golovinomyces magnicellulatus* Strain FPH2017-1

Coralie Farinas,[a] Emile Gluck-Thaler,[a] Jason C. Slot,[a] Francesca Peduto Hand[a]

[a]Department of Plant Pathology, The Ohio State University, Columbus, Ohio, USA

ABSTRACT Powdery mildew (PM) fungi are obligate biotrophs capable of infecting diverse plant hosts, ranging from monocotyledonous agricultural crops to dicotyledonous ornamental crops. The PM lifestyle poses significant challenges for studying these pathogens in isolation from their host. We present a draft genome of *Golovinomyces magnicellulatus*, a host-specific PM on *Phlox* species.

*G*olovinomyces magnicellulatus* (Leotiomycetes, Ascomycota) is an obligate host-specific fungal biotroph that causes powdery mildew (PM) disease on ornamental plants in the *Phlox* genus (1). Due to difficulties in growing PM fungi under axenic conditions, little is known regarding the genetic and evolutionary bases of their lifestyles, presenting an opportunity to gain insight through a genome-focused approach.

G. magnicellulatus strain FPH2017-1 was isolated from *Phlox paniculata* in Leipsic, Ohio. A single spore was isolated on a detached leaf bioassay (2) and grown on *P. paniculata* 'Starfire' plants in a growth chamber. Spores were harvested periodically over 1 month by rinsing infected leaves with 0.1% Tween solution and then filtering with Miracloth and stabilizing using 10 mM Tris buffer (pH 7). The solution was centrifuged, and the resulting pellet was immersed in liquid nitrogen and kept at −80°C. DNA was extracted from the pellet using the DNeasy plant minikit (Qiagen).

DNA libraries were prepared using the NEBNext Ultra II DNA library prep kit and sequenced using the Illumina MiSeq PE300 platform. Unsheared DNA extracts were prepared using a ligation sequencing kit (SQK-LSK109) and sequenced using MinION (Oxford Nanopore Technologies).

Illumina sequencing generated 17,742,739 reads (35 to 300 bp long) at 46× coverage (Table 1). Reads were trimmed using Trimmomatic v.0.36 (3) with the options ILLUMINACLIP, TruSeq3-PE.fa:2:30:10, CROP:290, SLIDINGWINDOW:10:25, HEADCROP: 10, and MINLEN:100 (4). Nanopore sequencing generated 427,831 reads (46 to 42,472 bp long) at 4× coverage (Table 1). The reads were quality filtered using Albacore v.2.3.1 (5). Iterative BLASTn searches against the NCBI nucleotide database (last accessed 11 February 2019) in conjunction with BBSplit v.37.93 (6) were used to identify and remove contaminant reads that had >75% identity and >50% query coverage to database entries originating from nonfungal organisms. We then performed *de novo* hybrid genome assembly using SPAdes v.3.12.0 (7) and identified known and *de novo* repeat elements using RepeatModeler v.1.0.11 (8).

We annotated the assembly using three iterations of a MAKER v.2.31.9 (9) pipeline. In the first iteration, we provided MAKER with transcriptome sequencing (RNA-seq) data of *Golovinomyces cichoracearum* (http://genome.jgi.doe.gov/Golci1) and 10 protein data sets from other Leotiomycetes species (*Blumeria graminis* f. sp. hordei DH14 [http://mycocosm.jgi.doe.gov/Blugr1], *B. graminis* f. sp. hordei Race1 [https://mycocosm.jgi.doe.gov/BlugrR1_1], *B. graminis* f. sp. tritici 96224 [https://mycocosm.jgi.doe.gov/Blugra1], *Erysiphe necator* [https://mycocosm.jgi.doe.gov/Erynec1], *G. cichoracearum*

TABLE 1 Summary statistics of assembly and annotation of *G. magnicellulatus*

Parameter	Value
Assembly	
Genome size (Mb)	129.9
Avg coverage ($\times$) (no. of Illumina reads)[a]	46 (97)
Avg coverage ($\times$) (no. of Nanopore reads)[a]	4 (60)
No. of contigs	84,604
N_{50} (bp)	4,118
Longest scaffold (kbp)	197
GC content (%)	44
BUSCO[b] (% recovered)	88.2
Annotation	
Total no. of protein-coding genes	8,172
Avg gene length (bp)	1,764
No. of coding sequences[c]	8
No. of repeat sequences[c]	40
No. of proteins with at least one Pfam domain[d]	6,396
No. of secreted proteins[e]	304

[a] Percentage of reference bases covered, estimated using BBMap v.37.93 (6).
[b] Sordariomyceta data set.
[c] Percent assembly size.
[d] Identified using InterProScan v.5.25-64 (15).
[e] Identified using SignalP-5.0 (16) and TMHMM v.2.0 (http://www.cbs.dtu.dk/services/TMHMM/) to exclude transmembrane proteins.

[https://mycocosm.jgi.doe.gov/Golci1], *Amorphotheca resinae* [https://mycocosm.jgi.doe.gov/Amore1], *Meliniomyces variabilis* [https://mycocosm.jgi.doe.gov/Melva1], *Sclerotinia sclerotiorum* [https://mycocosm.jgi.doe.gov/Sclsc1], *Rhizoscyphus ericae* [https://mycocosm.jgi.doe.gov/Rhier1/], and *Botrytis cinerea* [https://mycocosm.jgi.doe.gov/Botci1]). For the second iteration, we provided MAKER with the *ab initio* gene predictors SNAP v.2013-02-16 (10) (trained using high-quality predictions from round 1) and AUGUSTUS v.3.3 (11) (trained using BUSCO v.3.0.1 [12]). For the final iteration, we provided MAKER with updated evidence from SNAP and AUGUSTUS (both retrained using high-quality predictions from round 2) and set the option keep_preds to 1.

Many PM genomes are estimated to range in size from 120 to 220 Mb (13), due in part to high repeat content. Conversely, PMs generally possess fewer protein-coding genes (6,000 to 7,000) than other fungi (13, 14). Our genome falls within the reported PM genome size, 129.9 Mb, while our annotation process recovered 8,172 protein-coding genes, more than generally reported (Table 1), which we attribute to the multiple lines of *ab initio* evidence used in the annotation process.

Data availability. This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession numbers VCMJ00000000 and PRJNA540711 (SRA database).

REFERENCES

1. Matsuda S, Takamatsu S. 2003. Evolution of host–parasite relationships of Golovinomyces (Ascomycete: Erysiphaceae) inferred from nuclear rDNA sequences. Mol Phylogenet Evol 27:314–327. https://doi.org/10.1016/S1055-7903(02)00401-3.

2. Farinas C, Jourdan P, Paul PA, Peduto Hand F. 2019. Development and evaluation of two laboratory bioassays to study powdery mildew pathogens of Phlox in vitro. Plant Dis 103:1536–1543. https://doi.org/10.1094/PDIS-01-19-0031-RE.

3. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

4. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

5. Pomerantz A, Peñafiel N, Arteaga A, Bustamante L, Pichardo F, Coloma LA, Barrio-Amoros CL, Salazar-Valenzuela D, Prost S. 2018. Real-time DNA barcoding in a rainforest using nanopore sequencing: opportunities for rapid biodiversity assessments and local capacity building. GigaScience 7:giy033. https://doi.org/10.1093/gigascience/giy033.

6. Bushnell B. 2014. BBMap: a fast, accurate, splice-aware aligner (no. LBNL-7065E). Lawrence Berkeley National Lab (LBNL), Berkeley, CA.

7. Antipov D, Korobeynikov A, McLean JS, Pevzner PA. 2016. hybridSPAdes: an algorithm for hybrid assembly of short and long reads. Bioinformatics 32:1009–1015. https://doi.org/10.1093/bioinformatics/btv688.

8. Bao W, Kojima KK, Kohany O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA 6:11. https://doi.org/10.1186/s13100-015-0041-9.

9. Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics 12:491. https://doi.org/10.1186/1471-2105-12-491.

10. Korf I. 2004. Gene finding in novel genomes. BMC Bioinformatics 5:59. https://doi.org/10.1186/1471-2105-5-59.

11. Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics 24:637–644. https://doi.org/10.1093/bioinformatics/btn013.

12. Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva V, Zdobnov EM. 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol Biol Evol 35:543–548. https://doi.org/10.1093/molbev/msx319.

13. Wu Y, Ma X, Pan Z, Kale YS, Song Y, King H, Zhang Q, Presley C, Deng X, Wei C-I, Xia S. 2018. Comparative genome analyses reveal sequence features reflecting distinct modes of host-adaptation between dicot and monocot powdery mildew. BMC Genomics 19:705. https://doi.org/10.1186/s12864-018-5069-z.

14. Sonah H, Deshmukh RK, Belanger RR. 2016. Computational prediction of effector proteins in fungi: opportunities and challenges. Front Plant Sci 7:126. https://doi.org/10.3389/fpls.2016.00126.

15. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S-Y, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236–1240. https://doi.org/10.1093/bioinformatics/btu031.

16. Armenteros JJA, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol 37:420–423. https://doi.org/10.1038/s41587-019-0036-z.