Draft Genome Resource for the Ex-types of *Phytophthora ramorum*, *P. kernoviae*, and *P. melonis*, Species of Regulatory Concern, Using Ultra-Long Read MinION Nanopore Sequencing

Subodh K. Srivastava,^{1,2,†} Z. Gloria Abad,^{1,†} Leandra M. Knight,^{1,2} Kurt Zeller,¹ Vessela Mavrodieva,¹ and Mark Nakhla¹

¹ USDA-APHIS-PPQ, Science and Technology Beltsville Laboratory, Beltsville, MD 20705, U.S.A.

² Department of Entomology and Plant Pathology, North Carolina State University, Raleigh, NC 27695, U.S.A.

Abstract

Phytophthora ramorum, *P. kernoviae*, and *P. melonis* are each species of current regulatory concern in the United States, the United Kingdom, and other areas of the world. Ex-type material are cultures and duplicates of the type that was used to describe each species and that are deposited in additional culture collections. Using these type specimens as references is essential to designing correct molecular identification and diagnostic systems. Here, we report a whole genome sequence for the Ex-type material of *P. ramorum*, *P. kernoviae*, and *P. melonis* generated using high-throughput sequencing via the MinION third generation platform from Oxford Nanopore Technology. We assembled the quality filtered reads into contigs for each species. We assembled the continuous contigs of *P. ramorum*, *P. kernoviae*, and *P. melonis* (1,322, 545, and 2,091 contigs, respectively). The *ab initio* prediction of genes from these species reveals that there are 16,838, 12,793, and 34,580 genes in *P. ramorum*, *P. kernoviae*, and *P. melonis*, respectively. Of the 34,580 *P. melonis* genes, 10,164 genes were conserved among all three of these *Phytophthora* species which may include pathogenicity genes. We compared the ex-type of *P. ramorum* EU1 lineage assembly with another selected isolate of EU1 available at the National Center for Biotechnology Information and found 251,859 single nucleotide polymorphisms (SNPs) genome-wide; the comparison with the EU2 lineage genome isolate revealed 441,859 SNPs genome-wide. This genome resource of the ex-types of *P. ramorum*, and *P. kernoviae* is a significant contribution as these species are among the most important pathogens of regulatory concern in different regions of the world.

Genome Announcement

The genus *Phytophthora* now has 188 described species and includes numerous devastating plant pathogens that represent severe threats to trade, food security, agricultural systems, and natural ecosystems worldwide. In this work, we describe whole

Funding

This research was supported by U.S. Department of Agriculture Farm Bill grant FB18 3.0630.01 provided to Z. Gloria Abad. This research was supported in part by the intramural research programs of the U.S. Department of Agriculture/Animal and Plant Health Inspection Service (APHIS)/Plant Protection and Quarantine (PPQ).

Keywords

genome, ex-type, high throughput sequencing, MinION ONT ultra-long read sequencing, *Phytophthora*, third generation high throughput sequencing, 3G HTS

The findings and conclusions in this preliminary publication have not been formally disseminated by U.S. Department of Agriculture and should not be construed to represent any agency determination of policy.

[†]Corresponding authors: Z. G. Abad: Gloria.Abad@aphis.usda.gov and S. K. Srivastava: subodh.k.srivastava@usda.gov

The author(s) declare no conflict of interest.

Accepted for publication 26 February 2020.

Table 1. Assembly statistics of *Phytophthora ramorum*, *P. kernoviae*, and *P. melonis* MinION reads using Canu assembler

Attribute	<i>P. ramorum</i>	<i>P. kernoviae</i>	<i>P. melonis</i>
WPC/S&T BL codes ^a	P10103/BL 55G (ET)	P19827/BL 91 (ET)	P6870/BL 23 (ET)
Quality total reads (>1,000 nt)	625,764	601,746	1,621,287
Read mean length	2,673	2750	3164
Times Coverage	42x	40x	50x
Total assembled contigs(>1,000)	1,322	545	2,091
L ₅₀	147	65	276
Min	1,006	1,004	1,005
N ₅₀	97,267	183,615	107,202
Largest contigs	588,354	1,023,884	720,402
Sum	52 Mb	42 Mb	112 Mb
Contigs >50 kb	291	223	605
Contigs >100 kb	139	131	313
GC%	54.7	50.4	54.73
Assemble completeness (CEGMA)	88%	95%	94%
Assemble completeness (BUSCO)	54%	64%	77%
Predicted genes	16,838	12,793	34,580
NCBI accession	RYEP00000000.1	VKKV000000000	VXDT000000000

^a WPC = World Phytophthora Collection, Riverside California. BL = USDA-PPQ Science & Technology Beltsville Lab. ET = Ex-type.

genome sequences (WGS) of three ex-type isolates of *Phytophthora* species, generated using an ultra-long sequence read third generation high throughput sequencing (3G HTS) technology. We have generated these WGS from the type isolates from each of *Phytophthora ramorum*, *P. kernoviae*, and *P. melonis*. The *P. melonis* genome is the first representative of that species to have a whole-genome sequence. While there are other genome sequences from *P. kernoviae* and *P. ramorum* available in databases such as the National Center for Biotechnology Information (NCBI) database, however none of those other genome sequences are from the isolate used to describe those species (the type isolate). The reference sequences derived from type species (which act as references for species) served to label the correct species name (O'Leary et al. 2016). *Phytophthora ramorum* is a well-studied pathogen that causes the epidemic disease known as sudden oak death (SOD). SOD has resulted in an estimated property loss of \$135 million just in northwestern California (Kovacs et al. 2011) and is one of the most damaging *Phytophthora* species in nursery and forest settings. *Phytophthora kernoviae* was described in 2005 (Brasier et al. 2005) and causes aerial dieback symptoms on a range of woodland, heathland, and ornamental plants and extensive bleeding stem lesions and tree mortality on trees and shrubs in parts of Great Britain (Brasier et al. 2005). Currently the species is present in The United Kingdom, New Zealand, and South America (Chile). *Phytophthora kernoviae* is one of the important species of the present time, as the wide host range is present in many countries of the world (Widmer 2010). *Phytophthora melonis*, described in 1976 (Katsura 1976), is a significant pathogen in cucurbits in Iran and south China and its importance has not been much investigated in recent trade (Wu et al. 2011).

During the last ten years, HTS technologies have enhanced the production of whole genome sequences in biological areas. The 3G HTS MinION portable device produced by Oxford Nanopore Technologies (ONT) has the capacity to produce ultra-long reads for up to 882 kb and, indeed, higher with a high sequencing accuracy, reaching more than 95% for 1D or "one-directional" sequencing and 2D or "two directional" sequencing experiments. (Lu et al. 2016; Tyler et al. 2018). The ultra-long single DNA molecule, sequence reads produced by the MinION device should ensure a better resolution of structural genomic variants and of repeat content when doing genome assemblies (Kchouk et al. 2017) than can be achieved with other HTS platforms like Illumina NGS. The MinION 3G HTS platform, with its longer DNA sequence reads, has been used successfully in many recent genome assemblies from organisms with larger, generally complex genomes (Istace et al. 2017; Mondal et al. 2017).

We prepared genomic libraries for these *Phytophthora* species following ONT standard protocols, with slight modifications and standardization required for the organisms in this study. We isolated 400 ng of DNA from pure culture using the Qiagen DNeasy Mericon kit for the extraction of total DNA from samples of raw or processed food material that we modified for DNA extraction of *Phytophthora* species. We prepared the sequencing libraries with ONT ligation sequencing kit SQK-LSK109 following their standard protocol. All runs were

performed using flow-cells version R9.4.1 on the MinION device for 48 h, using the ONT sequencing protocol. We converted the *P. ramorum* FAST5 raw sequences to fastq files by base-calling with Albacore v.2.02 using the FLOMIN106 ONT pipeline. We base-called MinION output files from *P. kernoviae* and *P. melonis* genomes with Guppy Basecalling software from ONT version 2.3.1. The initial analysis reports were generated by nanoplot (De Coster et al. 2018) followed by a Canu genome assembly. We filtered raw data to include only data with a minimum quality read score of above 7 and with a DNA sequence read length of above 1,000 bases to improve the assembly.

The remaining data were used to generate draft genomes for these species. The quality filtered reads were as follows: from *P. ramorum*, 625,764 (2,673 mean read length and 42× coverage), *P. kernoviae*, 601,746 (2,750 mean read length and 40× coverage), and *P. melonis* 1,621,287 (3,164 mean read length and 50× coverage) reads of *Phytophthora* species and were assembled using the high-noise single-molecule sequencing assembler Canu version 1.8 (Koren et al. 2017) (Table 1). The assembled contigs were polished with ONT reads for any mis-assembly and error bases using Pilon (Walker et al. 2014). Finally, we evaluated the assembled contigs with quality assessment tools for genome assemblies known as QUAST (Gurevich et al. 2013) and core eukaryotic genes (CEG) mapping approach (CEGMA) model sets (Nishimura et al. 2017; Simão et al. 2015). We used taxonomic category custom (max intron length 5,000 and CEGMA gene flanks 2,000 nt) with a core CEG reference gene set.

We trained the gene prediction model for AUGUSTUS software (Cui et al. 2019; Stanke and Waack 2003; Stanke et al. 2008) using the previously sequenced species *Phytophthora capsici* as a model organism. This model predicted 16,838, 12,793, and 34,580 proteins in *P. ramorum*, *P. kernoviae*, and *P. melonis*, respectively. The predicted proteins with a significant cut-off level, $E \leq 10^{-9}$, was used to identify 10,164 common proteins with BLASTP (Altschul et al. 1997), using all three *Phytophthora* proteins common between these *Phytophthora* species (Srivastava et al. 2014).

Phytophthora ramorum has been separated into four clonal lineages: EU1, EU2, NA1, and NA2 (Grünwald et al. 2009; Kamoun et al. 2015; LeBoldus et al. 2017) that have different geographic distributions and that display differences in pathogenicity. We performed comparative genomics among the three *P. ramorum* lineages EU1, EU2, and NA1, to visualize genome differences among the lineages, and found 88% covered between our MinION-derived EU1 lineage genome assembly and an Illumina-derived EU1 lineage genome assembly, with 12% more bases covered in EU1 genome sequenced with MinION technology. We believe that this additional 12% is likely attributable to a sequenced accessory genome, which reveals the advantage of the 3G HTS ultra-long sequencing technology in our genome assembly (Table 1). Our genome-wide analyses also revealed that there is a significant lineage-specific genome content (1.5% of the EU1 genome) that could not have a match to the other published genomes of EU2 and NA1 lineage isolates (Turner et al. 2017). We compared our Ex-type duplicate (isolate CPHST-BL 55G) genome EU1 lineage assembly with the genome assembly from *P. ramorum* isolate CC1011 (also EU1 lineage) that is available at the NCBI, and we found 251,859 quality single nucleotide polymorphisms (SNPs) genome-wide. When we compared our *P. ramorum* EU1 lineage genome assembly with an available genome EU2 lineage isolate (SOD69/12), we found 441,859 SNPs genome-wide. This information may be indicate targets that can be used for race-specific identification and development of diagnostic methods to differentiate strains and lineages of *P. ramorum*.

These whole-genome shotgun projects have been deposited into GenBank project PRJNA510798, *P. ramorum* accession number RYEP00000000 with raw data submission SRR8360551, *P. kernoviae* accession number VKKV00000000 with submission SRR9703450 under project PRJNA554119, and *P. melonis* accession number VXDT00000000 with submission SRR9951418 under project PRJNA556690.

Acknowledgments

We are very thankful to Y. Rivera, J. Bienapfl, K. Jennings, and S. Nunziata at the USDA-APHIS-PPQ Science and Technology Beltsville Laboratory Beltsville Laboratory for technical discussion and software support, respectively. The authors would also like to thank M. Coffey at the World Phytophthora Collection at University of California-Riverside for his collaboration providing specimens used in this study. The findings and conclusions in this presentation are

Literature Cited

- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* 25:3389-3402.
- Brasier, C. M., Beales, P. A., Kirk, S. A., Denman, S., and Rose, J. 2005. *Phytophthora kernoviae* sp. nov., an invasive pathogen causing bleeding stem lesions on forest trees and foliar necrosis of ornamentals in the UK. *Mycol. Res.* 109:853-859.
- Cui, C., Herlihy, J. H., Bombarely, A., McDowell, J. M., and Haak, D. C. 2019. Draft assembly of *Phytophthora capsici* from long-read sequencing uncovers complexity. *Mol. Plant-Microbe Interact.* 32:1559-1563.
- De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M., and Van Broeckhoven, C. 2018. NanoPack: Visualizing and processing long-read sequencing data. *Bioinformatics* 34:2666-2669.
- Grünwald, N. J., Goss, E. M., Ivors, K., Garbelotto, M., Martin, F. N., Prospero, S., Hansen, E., Bonants, P. J., Hamelin, R. C., Chastagner, G., Werres, S., Rizzo, D. M., Abad, G., Beales, P., Bilodeau, G. J., Blomquist, C. L., Brasier, C., Brière, S. C., Chandelier, A., Davidson, J. M., Denman, S., Elliott, M., Frankel, S. J., Goheen, E. M., de Gruyter, H., Heungens, K., James, D., Kanaskie, A., McWilliams, M. G., Man in 't Veld, W., Moralejo, E., Osterbauer, N. K., Palm, M. E., Parke, J. L., Sierra, A. M., Shamoun, S. F., Shishkoff, N., Tooley, P. W., Vetrano, A. M., Webber, J., and Widmer, T. L. 2009. Standardizing the nomenclature for clonal lineages of the sudden oak death pathogen, *Phytophthora ramorum*. *Phytopathology* 99:792-795.
- Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. 2013. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* 29:1072-1075.
- Istace, B., Friedrich, A., d'Agata, L., Faye, S., Payen, E., Beluche, O., Caradec, C., Davidas, S., Cruaud, C., Liti, G., Lemaître, A., Engelen, S., Wincker, P., Schacherer, J., and Aury, J. M. 2017. De novo assembly and population genomic survey of natural yeast isolates with the Oxford Nanopore MinION sequencer. *Gigascience* 6:1-13.
- Kamoun, S., Furzer, O., Jones, J. D., Judelson, H. S., Ali, G. S., Dalio, R. J., Roy, S. G., Schena, L., Zambounis, A., Panabieres, F., Cahill, D., Ruocco, M., Figueiredo, A., Chen, X. R., Hulvey, J., Stam, R., Lamour, K., Gijzen, M., Tyler, B. M., Grünwald, N. J., Mukhtar, M. S., Tomé, D. F., Tör, M., Van Den Ackerveken, G., McDowell, J., Daayf, F., Fry, W. E., Lindqvist-Kreuz, H., Meijer, H. J., Petre, B., Ristaino, J., Yoshida, K., Birch, P. R., and Govers, F. 2015. The Top 10 oomycete pathogens in molecular plant pathology. *Mol. Plant Pathol.* 16:413-434.
- Katsura, K. 1976. Two new species of *Phytophthora* causing damping-off of cucumber and trunk rot of chestnut. *Trans. Mycol. Soc. Jpn.* 17:238-242.
- Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. 2017. Canu: Scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res.* 27:722-736.
- Kovacs, K., Václavík, T., Haight, R. G., Pang, A., Cunneiffe, N. J., Gilligan, C. A., and Meentemeyer, R. K. 2011. Predicting the economic costs and property value losses attributed to sudden oak death damage in California (2010-2020). *J. Environ. Manage.* 92:1292-1302.
- LeBoldus, J. M., Sondreli, K. L., Sutton, W., et al. 2017. First report of *Phytophthora ramorum* lineage EU1 infecting Douglas fir and grand fir in Oregon. *Plant Dis.* 102:455.
- Lu, H., Giordano, F., and Ning, Z. 2016. Oxford Nanopore MinION sequencing and genome assembly. *Genomics Proteomics Bioinformatics* 14:265-279.
- Kchouk, M., Gibrat, J.-F., and Eloum, M. 2017. Generations of sequencing technologies: From first to next generation. *Biol. Med. (Aligarh)* 9:3.
- Mondal, T. K., Rawal, H. C., Gaikwad, K., Sharma, T. R., and Singh, N. K. 2017. First *de novo* draft genome sequence of *Oryza coarctata*, the only halophytic species in the genus *Oryza*. *F1000 Res.* 6:1750.
- Nishimura, O., Hara, Y., and Kuraku, S. 2017. gVolante for standardizing completeness assessment of genome and transcriptome assemblies. *Bioinformatics* 33:3635-3637.
- O'Leary, N. A., Wright, M. W., Brister, J. R., Ciuffo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., Farrell, C. M., Goldfarb, T., Gupta, T., Haft, D., Hatcher, E., Hlavina, W., Joardar, V. S., Kodali, V. K., Li, W., Maglott, D., Masterson, P., McGarvey, K. M., Murphy, M. R., O'Neill, K., Pujar, S., Rangwala, S. H., Rausch, D., Riddick, L. D., Schoch, C., Shkeda, A., Storz, S. S., Sun, H., Thibaud-Nissen, F., Tolstoy, I., Tully, R. E., Vatsan, A. R., Wallin, C., Webb, D., Wu, W., Landrum, M. J., Kimchi, A., Tatusova, T., DiCuccio, M., Kitts, P., Murphy, T. D., and Pruitt, K. D. 2016. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44 (D1):D733-D745.
- 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.
- Srivastava, S. K., Huang, X., Brar, H. K., Fakhoury, A. M., Bluhm, B. H., and Bhattacharyya, M. K. 2014. The genome sequence of the fungal pathogen *Fusarium virguliforme* that causes sudden death syndrome in soybean. *PLoS One* 9:e81832.
- Stanke, M., Diekhans, M., Baertsch, R., and Haussler, D. 2008. Using native and syntetically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics* 24:637-644.
- Stanke, M., and Waack, S. 2003. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* 19 (Suppl 2):ii215-ii225.
- Turner, J., O'Neill, P., Grant, M., Mumford, R. A., Thwaites, R., and Studholme, D. J. 2017. Genome sequences of 12 isolates of the EU1 lineage of *Phytophthora ramorum*, a fungus-like pathogen that causes extensive damage and mortality to a wide range of trees and other plants. *Genom. Data* 12:17-21.
- Tyler, A. D., Mataseje, L., Urfano, C. J., Schmidt, L., Antonation, K. S., Mulvey, M. R., and Corbett, C. R. 2018. Evaluation of Oxford Nanopore's MinION sequencing device for microbial whole genome sequencing applications. *Sci. Rep.* 8:10931.
- Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., and Earl, A. M. 2014. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963.
- Widmer, T. L. 2010. *Phytophthora kernoviae* oospore maturity, germination, and infection. *Fungal Biol.* 114:661-668.
- Wu, Y., Lu, S., Huang, S., Fu, G., Chen, L., Xie, D., Li, Q., and Cen, Z. 2011. [Field resistance of *Phytophthora melonis* to metalaxyl in South China]. *Wei Sheng Wu Xue Bao* 51:1078-1086.