Updated Assembly of *Phytophthora ramorum* pr102 Isolate Incorporating Long Reads from PacBio Sequencing

Mathu Malar C,^{1,2} Jennifer David Yuzon,^{3,4} Arijit Panda,^{1,2} Takao Kasuga,^{3,4,†} and Sucheta Tripathy^{1,2,†}

¹ Computational Genomics Lab, Structural Biology and Bioinformatics Division, CSIR Indian Institute of Chemical Biology, Kolkata 700032, India

² Academy of Scientific and Innovative Research (AcSIR), New Delhi, India

³ Department of Plant Pathology, University of California, Davis, CA, U.S.A.

⁴ USDA-ARS, Davis, CA, U.S.A.

Abstract

The NA1 clonal lineage of *Phytophthora ramorum* is responsible for sudden oak death, an epidemic that has devastated California coastal forest ecosystems. An NA1 isolate, Pr102, derived from coast live oak in California, was previously sequenced and reported with a 65-Mb assembly containing 12 Mb of gaps in 2,576 scaffolds. Here, we report an improved 70-Mb genome in 1,512 scaffolds with 6,752 bp of gaps after incorporating PacBio P5-C3 long reads. This assembly contains 19,494 gene models (average gene length of 2,515 bp) compared with 16,134 genes (average gene length of 1,673 bp) in the previous version. We predicted 29 new RXLR genes and 76 new paralogs of a total 392 RXLR genes from this assembly. We predicted 35 CRN genes compared with 19 in an earlier version with six paralogs. Our long non-coding RNA prediction identified 255 candidates. This new resource will be invaluable for future evolution studies on the invasive plant pathogen.

Genome Announcement

The oomycete genus *Phytophthora* includes several of the most destructive plant pathogens that threaten agriculture as well as natural ecosystems (Kamoun et al. 2015). The sudden oak death pathogen *P. ramorum* affects not only oak trees but also causes disease on California bay *Umbellularia californica*, Tanoak, *Notholithocarpus densiflorus*, rhododendrons, camellias, and other forest species and common horticultural plants. The origin of *P. ramorum* is still not clear. So far, four lineages of *P. ramorum* are identified in North America and Europe. The first draft genome sequence of *P. ramorum* was reported by Tyler and coworkers (2006), containing 12 Mb of gaps with an assembly size of 65 Mb. *P. ramorum* Pr102 was isolated in 2001 from coast live oak in Marin County, California. Pr102 represents non-wild type isolates (Brasier et al. 2006), which typically originate from oaks, are genomically unstable, and show reduced aggressiveness in comparison with wild-type isolates (Elliott et al. 2018; Kasuga et al. 2012, 2016).

Mathu Malar C and Jennifer David Yuzon were equal contributors.

†Corresponding authors: T. Kasuga; tkasuga@ucdavis.edu and S. Tripathy; tsucheta@iicb.res.in

The author(s) declare no conflict of interest.

Accepted for publication 10 July 2019.

Funding

This work was funded through a DBT Government of India Ramalingaswamy fellowship to S. Tripathy. This work was also supported by the United States Department of Agriculture (USDA) Agriculture Research Service, CRIS Project number 2032-22000-016-00D to T. Kasuga, the USDA Farm Bill sections 10201 and 10007, the USDA Animal and Plant Health Inspection Service number 12-8130-0191-CA.

Keywords

gaps, long read assembly, *Phytophthora ramorum*, repeats, sudden oak death

Table 1. Genome assembly statistics comparison of *Phytophthora ramorum* older and newer version assemblies

Parameters	<i>P. ramorum</i> v1	<i>P. ramorum</i> v2
Total no. of genes	16,134	19,494
Average gene length	1,673	2,515.59
Total no. of gaps	12 Mb	6,752 bp
Total no. of scaffolds	2,576	1,512
Genome assembly size	65 Mb	70 Mb
% Repeat content	29.08%	54.4%
Stramenopile BUSCOs missing	4	1
Largest scaffold	1.2 Mb	2.5 Mb
No. of RXLR genes	370	392
No. of CRN genes	19	35

The genomic DNA of Pr102 was extracted from 7-day-old colonies grown on polycarbonate membrane filters, using a nuclear isolation method (Maler C et al. 2019; Zhang et al. 2012). A total of 80x coverage (435,399 reads) PacBio P5-C3 reads were generated at the University of California Davis Genome Center, with an average read length of 14 kb. In addition, 17,196,940 Illumina genomic reads with an average length of 100 bp paired-end reads at 25x coverage was generated.

PacBio reads, Illumina reads, and Sanger reads from the National Center for Biotechnology Information (NCBI) Sequence Read Archive were used to assemble the genome using MaSuRCA (Zimin et al. 2013) assembler. To reduce indel errors from the MaSuRCA assembled genome, the assembly was polished, initially using Sanger reads, then, Sanger polished contigs were polished using Illumina reads with Pilon (Walker et al. 2014). Total assembled genome length was approximately 70 Mb with an N₅₀ value of 186,608 bp. The largest scaffold length was 2.5 Mb. The genome assembly completeness was computed using Stramenopile BUSCO (Simão et al. 2015), estimated to be 99.2%. The gene calling was done using Augustus (Stanke and Waack, 2003) with Pr102 V1 gene models as training datasets, resulting in 19,494 predicted genes. Pfam domain searches using InterProScan (Jones et al. 2014) identified 3,255 Pfam domains from our predicted protein sequences. Our effector prediction method (Malar C et al. 2019) resulted in 392 RXLR and 35 CRN effectors. Repeat analysis using Repeat Masker (A. Smit, R. Hubley, and P. Green,) identified nearly twice as many repeats in the genome, in comparison to the previous version (54.4% versus 29.1%). Our assembly statistics comparison with the older version Pr102 V1 (Tyler et al. 2006) assembly indicates that the present assembly version is a substantial improvement over the earlier version, with the number of scaffolds, largest scaffold, assembly size, and number of genes and repeats better than the earlier version. Comparison of assembly statistics are provided in Table 1.

The *P. ramorum* Pr102 sequence will provide a valuable resource for population genomics and future comparative genomic studies. The genome sequence has been submitted to NCBI with accession number RHLB00000000. Annotation and gene model files can be downloaded from the Eumicrobe database (Panda et al. 2018).

Acknowledgments

The authors acknowledge Computational Biology and Genomics Lab members at CSIR-IICB for valuable discussion. We acknowledge D. Rizzo from the University of California Davis for providing the isolates Pr102 and associated information.

Literature Cited

- Brasier, C., Kirk, S., and Rose, J. 2006. Differences in phenotypic stability and adaptive variation between the main European and American lineages of *Phytophthora ramorum*. Pages 166-173 in: Progress in Research on Phytophthora Diseases of Forest Trees. Forest Research, Farnham, U.K.
- Elliott, M., Yuzon, J., Tripathy, S., Bui, M., Chastagner, G. A., Coats, K., Rizzo, D. M., Garbelotto, M., and Kasuga, T. 2018. Characterization of phenotypic variation and genome aberrations observed among *Phytophthora ramorum* isolates from diverse hosts. BMC Genomics 19:320.
- Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas,

- A., Scheremetjew, M., Yong, S. Y., Lopez, R., and Hunter, S. 2014. InterProScan 5: Genome-scale protein function classification. *Bioinformatics* 30:1236-1240.
- 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-Kreuzer, 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.
- Kasuga, T., Bui, M., Bernhardt, E., Swiecki, T., Aram, K., Cano, L. M., Webber, J., Brasier, C., Press, C., Grünwald, N. J., Rizzo, D. M., and Garbelotto, M. 2016. Host-induced aneuploidy and phenotypic diversification in the sudden oak death pathogen *Phytophthora ramorum*. *BMC Genomics* 17:385.
- Kasuga, T., Kozanitis, M., Bui, M., Hüberli, D., Rizzo, D. M., and Garbelotto, M. 2012. Phenotypic diversification is associated with host-induced transposon de-repression in the sudden oak death pathogen *Phytophthora ramorum*. *PLoS One* 7:e34728.
- Malar, C. M., Yuzon, J. D., Das, S., Das, A., Panda, A., Ghosh, S., Tyler, B. M., Kasuga, T., and Tripathy, S. 2019. Haplotype-phased genome assembly of virulent *Phytophthora ramorum* isolate ND886 facilitated by long-read sequencing reveals effector polymorphisms and copy number variation. *Mol. Plant-Microbe Interact.* 32:1047-1060.
- Panda, A., Sen, D., Ghosh, A., Gupta, A., C. M. M., Prakash Mishra, G., Singh, D., Ye, W., Tyler, B. M., and Tripathy, S. 2018. EumicrobeDBLite: A lightweight genomic resource and analytic platform for draft oomycete genomes. *Mol. Plant Pathol.* 19:227-237.
- 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.
- Stanke, M., and Waack, S. 2003. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* 19 (Suppl 2):ii215-ii225.
- Tyler, B. M., Tripathy, S., Zhang, X., Dehal, P., Jiang, R. H., Aerts, A., Arredondo, F. D., Baxter, L., Bensasson, D., Beynon, J. L., Chapman, J., Damasceno, C. M., Dorrance, A. E., Dou, D., Dickerman, A. W., Dubchak, I. L., Garbelotto, M., Gijzen, M., Gordon, S. G., Govers, F., Grünwald, N. J., Huang, W., Ivors, K. L., Jones, R. W., Kamoun, S., Kramps, K., Lamour, K. H., Lee, M. K., McDonald, W. H., Medina, M., Meijer, H. J., Nordberg, E. K., Maclean, D. J., Ospina-Giraldo, M. D., Morris, P. F., Phuntumart, V., Putnam, N. H., Rash, S., Rose, J. K., Sakihama, Y., Salamov, A. A., Savidor, A., Scheuring, C. F., Smith, B. M., Sobral, B. W., Terry, A., Torto-Alalibo, T. A., Win, J., Xu, Z., Zhang, H., Grigoriev, I. V., Rokhsar, D. S., and Boore, J. L. 2006. *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science* 313:1261-1266.
- 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.
- Zhang, M. P., Zhang, Y., Scheuring, C. F., Wu, C. C., Dong, J. J., and Zhang, H. B. 2012. Preparation of megabase-sized DNA from a variety of organisms using the nuclei method for advanced genomics research. *Nat. Protoc.* 7:467-478.
- Zimin, A. V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S. L., and Yorke, J. A. 2013. The MaSuRCA genome assembler. *Bioinformatics* 29:2669-2677.