Whole Genome Sequences of the Raspberry and Strawberry Pathogens *Phytophthora rubi* and *P. fragariae*

Javier F. Tabima,1 Brent A. Kronmiller,2 Caroline M. Press,3 Brett M. Tyler,1,2 Inga A. Zasada,3 and Niklaus J. Grünwald†

1 Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR, 97331, U.S.A.
2 Center for Genome Biology and Biocomputing (CGRB), Oregon State University
3 Horticultural Crops Research Laboratory, USDA-ARS, Corvallis, OR, 97330, U.S.A.

Abstract

*Phytophthora rubi* and *P. fragariae* are two closely related oomycete plant pathogens that exhibit strong morphological and physiological similarities but are specialized to infect different hosts of economic importance, namely, raspberry and strawberry. Here, we report the draft genome sequences of these two *Phytophthora* species as a first step toward understanding the genomic processes underlying plant host adaptation in these pathogens.

Genome Announcement

More than 150 species of oomycete plant pathogens are harbored in the genus *Phytophthora* (Erwin and Ribeiro 1996; Kamoun et al. 2015). These organisms are highly diverse in lifestyle, host preference, and economic importance (Kamoun et al. 2015). Whole-genome sequences of various species in the genus *Phytophthora* have recently been published, yielding novel insights into the molecular basis of plant disease, such as the identification of effector proteins that can disrupt the physiology of the host (Haas et al. 2009; Tyler et al. 2006). Here, we present the whole-genome sequences of two soil-borne *Phytophthora* sister species, *P. fragariae* and *P. rubi*. These are highly similar in morphology and physiology but infect different hosts. Sequencing the genomes of these species will advance our understanding of the genomic mechanisms underlying host adaptation and knowledge of molecular mechanisms of plant pathogenicity (Stewart et al. 2014).

Genomic DNA was extracted from *P. rubi* ‘pd0101050015038’ (isolated in The Netherlands by K. Rosendhal from red raspberry) and *P. fragariae* CBS 209.46 (isolated in England by C. J. Hickman from strawberry). Genomes were sequenced by the Beijing Genomics Institute (BGI, Beijing), using the Illumina HiSeq2000 platform (Illumina, San Diego, CA, U.S.A.), using TruSeq libraries (paired end reads, insert size of 500 bp, average read length of 90 bp). Assemblies were performed using SOAPdenovo2 (kmer size of 36) (Luo et al. 2012). Transcriptomes of both species were also sequenced by BGI, using the Illumina HiSeq2000 technology (paired end reads, insert size of 500 bp, average read length of 90 bp). To obtain high-quality gene calls, the transcriptomes were assembled using Trinity (using nonstrand specific parameters, 100 Gb of RAM, and 10 CPU cores) (Grabherr et al. 2011) and were translated into amino-acid sequences using TransDecoder (Haas et al. 2013). To generate a reference database, gene prediction and annotation for the nuclear genomes of both species were performed using MAKER (Cantarel et al. 2008) trained with gene calls from other *Phytophthora* genomes, including *P. ramorum*, *P. sojae*, and *P. infestans* (Haas et al.

†Corresponding author: N. J. Grünwald; E-mail: nik.grunwald@ars.usda.gov

Accepted for publication 4 July 2017.

Funding

This work was supported by funds from the United States Department of Agriculture (USDA) Agricultural Research Service Current Research Information System projects 2072-12220-004-00-D and 2072-22000-039-00-D and the USDA National Institute of Food and Agriculture Risk Avoidance and Mitigation Program project 2010-511001-21649.

Keywords
genome, *Phytophthora*, raspberry, strawberry

This article is in the public domain and not copyrightable. It may be freely reprinted with customary crediting of the source. The American Phytopathological Society, 2017.
Table 1. Summary statistics of the genome assemblies of *Phytophthora fragariae* and *P. rubi*

<table>
<thead>
<tr>
<th>Species</th>
<th>No. supercontigs</th>
<th>No. contigs</th>
<th>Total length (bp)</th>
<th>Genome coverage</th>
<th>N50 (bp)</th>
<th>N90 (bp)</th>
<th>Max (bp)</th>
<th>Min (bp)</th>
<th>GC content (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>P. fragariae</em></td>
<td>8,511</td>
<td>10,440</td>
<td>76,752,825</td>
<td>76×</td>
<td>18,978</td>
<td>3,110</td>
<td>258,010</td>
<td>200</td>
<td>53.35</td>
</tr>
<tr>
<td><em>P. rubi</em></td>
<td>9,434</td>
<td>11,232</td>
<td>74,656,193</td>
<td>92×</td>
<td>16,735</td>
<td>2,697</td>
<td>139,498</td>
<td>200</td>
<td>53.42</td>
</tr>
</tbody>
</table>

a N50 = median length of more than 50% of the supercontigs in the entire assembly; N90 = median length of more than 90% of the supercontigs in the entire assembly; Max and Min = maximum and minimum lengths, respectively, of supercontigs found in the entire assembly; GC content = percentage of guanine-cytosine bases in the entire assembly.

2009; Tyler et al. 2006). The assembled transcripts were used as RNA evidence in MAKER to improve the quality of the called gene models. The predicted gene models were annotated using InterproScan 5 (Jones et al. 2014).

A total of 6,733 Mbp and 8,382 Mbp raw sequence data were obtained for *P. fragariae* and *P. rubi*, respectively. A low percentage of reads included adapter sequences (0.02% for *P. fragariae*, 0.01% for *P. rubi*) or duplicate reads (2% for *P. fragariae*, 6% for *P. rubi*). Up to 12 and 16% of the reads were filtered on the per position quality reported in the FASTQ files for *P. fragariae* and *P. rubi*, respectively, resulting in a total of 5,880 Mbp of clean nucleotides for *P. fragariae* and 6,960 Mbp of clean nucleotides for *P. rubi*.The *P. fragariae* assembly was estimated to be approximately 76 Mbp in size, assembled into 8,511 scaffolds with an N50 of 18,987 bp and a genome sequence depth of 76× (Table 1). The *P. rubi* genome was assembled into 9,434 scaffolds spanning a total length of 74 Mbp and a N50 of 16,735 bp and a genome sequence depth of 92× (Table 1). The *P. rubi* genome was assembled into 9,434 scaffolds spanning a total length of approximately 74 Mbp and a N50 of 16,735 bp and a genome coverage of 92× (Table 1). A total of 20,448 and 23,476 gene models were found for *P. fragariae* and *P. rubi*, respectively. Both species show a high abundance of Gene Ontology terms for DNA integration, nucleic acid binding, protein binding, peptidase/helicase and hydrolase activities, integrases, endonucleases, repeats, and transporter domains.

The genomes of *P. fragariae* and *P. rubi* will provide a unique resource for these closely related soil-borne oomycete plant pathogens and the molecular mechanisms associated with plant pathogenicity.

This Whole Genome Shotgun project has been deposited at GenBank under the accession numbers MWJK00000000 (*P. fragariae*) and MWJL00000000 (*P. rubi*). The complete data set used in this report, including the Illumina raw reads and whole genome sequence assembly have been deposited as a National Center for Biotechnology Information BioProject under the accession number PRJNA375089.

Acknowledgments

We thank D. Shen, B. J. Knaus, and J. Stewart for their support with advice and preliminary data. We also thank the Beijing Genomics Institute (BGI) for conducting the sequencing and assembly. Mention of trade names or commercial products in this manuscript are solely for the purpose of providing specific information and do not imply recommendation or endorsement.

Literature Cited


Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J.,
Couger, M. B., Eccles, D., Li, B., Lieber, M., Macmanes, M. D., Ott, M., Orvis, J.,
Pochet, N., Strocci, F., Weeks, N., Westerman, R., William, T., Dewey, C. N.,
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.,
Kamoun, S., Furzer, O., Jones, J. D. G., Judelson, H. S., Ali, G. S., Dallo, R. J. D.,
Roy, S. G., Schena, L., Zambounis, A., Panabieres, F., Cahill, D., Ruocco, M.,
Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y.,
Stewart, J. E., Kroese, D., Tabima, J. F., Larsen, M. M., Fieland, V. J., Press, 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., Ivons, K. L.,