Supplementary File S1

Discovery and assembly of two novel fungal genomes in the coast redwood genome data

For the Illumina sequencing effort in coast redwood, we used megagametophytes extracted from two different coast redwood seeds (see Methods). Within the Illumina read data for each seed, we unexpectedly found novel fungal species. The two fungi represent two distinct genera, *Pestalotiopsis* and *Diaporthe*, as described below. In each case, we collected a large number of reads that were sufficient to produce an assembly that appears to cover nearly all of the new genome.

Assembly of a new species of *Pestalotiopsis*: From one of the haploid coast redwood megagametophytes (identified here as Meg1), we generated 10.8 billion pairs of 150-bp reads in six runs and 31 lanes of an Illumina HiSeq instrument. Each run contained a large number of fungal reads, so many that we needed only a small subset of the reads to reconstruct the fungal genome. We chose one-half of one lane of data with 347 million reads (173 million pairs), and assembled it using SOAPdenovo2 (Luo *et al.* 2012), yielding 5.9 million scaffolds. (Note that increasing the amount of data did not improve the assembly.) From this dataset, we used depth of coverage, GC content, and BLAST alignments to identify scaffolds likely to represent the fungal genome.

The coast redwood genome has a GC content of 35.9%, while fungi in the *Pestalotiopsis* genus have a GC content above 48% (e.g., *Pestalotiopsis* sp. JCM 9685 has 51.7% GC.) The fungus was also present at considerably deeper coverage than the genome, indicating that the sample

contained many more fungal cells than coast redwood cells. We used both of these features to identify scaffolds in the assembly that were likely fungal in origin, as follows.

We identified 3,710 scaffolds containing 43.7 Mbp that had a GC content between 47.8% and 57.8%, and that also had a depth of coverage around 26X. We then considered any additional scaffolds that aligned via BLAST to a *Pestalotiopsis* species, which gave us 5,985 scaffolds containing 46.9 Mbp. We then removed duplicate scaffolds and also removed scaffolds that were shorter than 2 Kbp. The histogram of the number of scaffolds as a function of depth of coverage has a clear peak at 26X caused by the fungal scaffolds (**Figure 1**). We also ran a separate assembly using MEGAHIT (Li *et al.* 2015), which yielded a better assembly of the fungal mitochondrial genome. We replaced the fungal contigs from the SOAPdenovo2 assembly with the larger ones from this second assembly. We also replaced the contigs containing ITS (internal transcribed spacer) sequences, which were larger in the MEGAHIT assembly.

Figure 1: Histogram showing the number of scaffolds (y-axis) plotted against coverage. The peak at 4X is the host (coast redwood) genome, and the peak at 26X is the genome of the apparently novel fungus species in the genus *Pestalotiopsis*. The mitochondrial genomes form smaller peaks at much deeper coverage, not visible as peaks at this scale.
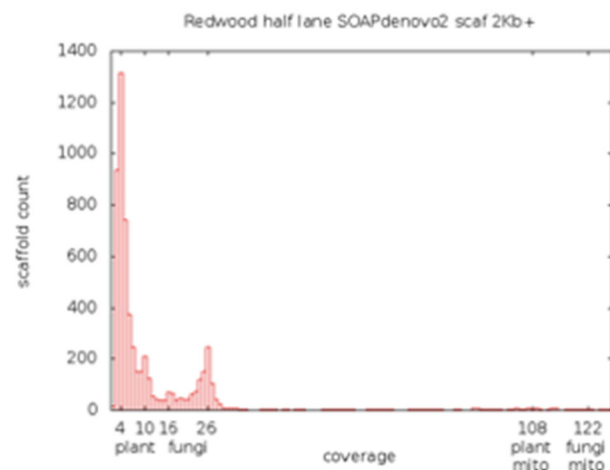
This assembly and filtering process yielded 1,118 fungal scaffolds containing 44,918,601 bases, with an average GC content of 52% and an N50 scaffold size of 286,884 bp.

All of the assembled scaffolds have a best BLAST match (averaging 92% identical) to *Pestalotiopsis* sp. JCM 9685 (GenBank accession BCGF00000000), a species that was obtained originally from a Pacific yew (*Taxus brevifolia*) tree in Bozeman, Montana (Pulici *et al.* 1997). The longest scaffold, at 2,314,818 bp in length, is 94% identical to the JCM 9685 genome. Note that the length of the *Pestalotiopsis* sp. JCM 9685 genome is 48.2 Mbp, about 7% larger than the 44.9 Mbp in our assembly. This could be due to missing sequence, or it could be that the novel fungus has a slightly smaller genome.

To more accurately identify the species of our assembled fungus, we aligned the entire assembly to the NCBI database of fungal ITS sequences (Schoch *et al.* 2014), which includes the UNITE database (Nilsson *et al.* 2018) and other sources, and which contains ITSs from over 450,000 species (9,552 fungi). The best match is to *Pestalotiopsis monochaeta*, which is 100% identical (across 602 bp) to our new fungus. Based on this analysis, we suggest our assembly represents a novel fungal species in genus *Pestalotiopsis*.

Assembly of an apparent new species of *Diaporthe*: From a second coast redwood megagametophyte, referred to here as Meg2, we collected 11.23 billion pairs of 150-bp reads in 6 runs and 30 lanes of an Illumina HiSeq instrument. (Note that Meg2 was used only in a pilot run, and the reads collected from it were not used for the final coast redwood assembly.) The amount of fungal DNA in this sample was much lower than in Meg1; consequently, we used the assembly of the full data set to identify fungal sequence.

We used BLASTN to identify the largest contig matching a fungus, a contig of 384,423 bp that was 87% identical to *Diaporthe ampelina* DA912, a pathogen of grapevine trunks (Morales-Cruz *et al.* 2015), which has a draft genome of 47.3 Mbp. The identified contig had a GC content of 53.6% and a depth of coverage of 27.7X, so we used these parameters to search further. We extracted all contigs longer than 2000 bp with a GC content within 10% of 53.6% and with coverage in the range 22.7 to 32.7. (Note again that the coast redwood genome is 35.9% GC.) This yielded 29,833 contigs covering 142 Mbp. From this set, we extracted all contigs >500 bp with 10% of their length matching any fungal genome and at least one *Diaporthe* match. The result after these filtering steps was a set of 1,229 contigs containing 55,328,661 bp, with a mean GC content of 53.1%.

We then identified scaffolds from the overall assembly that contained any of the above contigs. The scaffolds were re-aligned to the fungi database, and the ones with over 10% of their length aligned were extracted. The result was a set of 486 scaffolds containing 57,904,009 bp, with a scaffold N50 length of 259,065 bp.

The best match for the large majority of these scaffolds was *Diaporthe ampelina*. We also aligned the scaffolds to the ITS database as described above, and found a scaffold matching *Diaporthe lonicerae* MFLUCC 17-0963 (100% identical over 288 bp) and another scaffold matching *Penicillium spinuloramigenum* CBS 344.59 (100% identical over 500 bp). Because many more scaffolds matched *Diaporthe* species than *Penicillium* species, we concluded that the best genus assignment for this novel fungus species is *Diaporthe*.

DATA AVAILABILITY

The assemblies of these apparently novel fungal species were deposited at NCBI as GenBank accession JACVEI000000000 for *Pestalotiopsis* sp. DP-2020a and accession JACVEP000000000 for *Diaporthe* sp. DP-2020a.

LITERATURE CITED

Li, D., Liu, C. M., Luo, R., Sadakane, K., Lam, T. W., 2015 MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31: 1674-1676.

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen Y., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C., Cheung, D. W., Yiu, S.-M., Peng S., Xiaoqian, Z., Liu, G., Liao, X., Li, Y., Yang, H., Wang, J., Lam, T.-W., Wang, J., 2012 SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* 1: 18.

Morales-Cruz, A., Amrine, K. C. H., Blanco-Ulate, B., Lawrence, D. P., Travadon, R., Rolshausen, P. E., Baumgartner, K., Cantu, D., 2015 Distinctive expansion of gene families associated with plant cell wall degradation, secondary metabolism, and nutrient uptake in the genomes of grapevine trunk pathogens. *BMC Genomics* 16: 469.

Nilsson, R. H., Larsson, K. H., Taylor, A. F. S., Bengtsson-Palme, J., Jeppesen, T. S., Schigel, D., Kennedy, P., Picard, K., Glöckner, F. O., Tedersoo, L., Saar, I., Koljalg, U., Abarenkov, K., 2018 The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Res.* 47: D259-D264.

Pulici, M., Sugawara, F., Koshino, H., Okada, G., Esumi, Y, Uzawa, J., Yoshida, S., 1997

   Metabolites of *Pestalotiopsis* spp., endophytic fungi of *Taxus brevifolia*. *Phytochemistry*

   46: 313-319.

Schoch, C. L., Robbertse, B., Robert, V., Vu, D., Cardinali, G., *et al.*, 2014 Finding needles in

   haystacks: linking scientific names, reference specimens and molecular data for Fungi.

   Database 2014: bau061, doi: 10.1093/database/bau061.