**Draft Genome Assembly of Passalora sequoiae a Needle Blight Pathogen on Leyland Cypress**

Methodology

A single spore isolate of *P. sequoiae* 9LC2 from an infected Christmas tree near Hattiesburg, MS, USA, was grown on potato dextrose agar (Benton, Dickinson and Co., Sparks, MD, USA) for 6 months, then grown in corn meal broth for 3 months. The fungus mycelium was removed from the broth, freeze dried using a lyophilizer (Labconco Corporation, Kansas City, MO, USA), and ground into powder using liquid nitrogen. DNA was extracted using a modified CTAB protocol [21]. Genomic DNA was quantified using Qubit dsDNA High Sensitivity assay kit, the A260/A280 was determined by using a Nanodrop (Thermo Fisher Scientific, Waltham, MA), and the quantity and integrity of the DNA analyzed using the Agilent Genomic DNA ScreenTape (Agilent Technologies, CA, USA). DNA was sheared to approximately 20 kb fragments using a g-TUBE (Covaris, USA). SMRTbell library was prepared using the SMRTbell Template Preparation Kit 1.0 and the Damage Repair Kit (Pacific Biosciences, CA, USA) according to manufacturer’s instructions and the sequencing guide. The library was sequenced on a PacBio Sequel sequencer using one SMRT cell with P6-C4 sequencing chemistry at USDA-ARS Genomics and Bioinformatics Research Unit, Stoneville, MS, USA. Bam files were converted to subreads and then processed using the Finishing Module 20.0 of CLC\_Bio Workbench v.12 (Qiagen LLC, Hilden, Germany). A total of 519,499 subreads with a total of 6,612,712,889 nt, average length 14,247 nt, N50 21720, were generated. Subreads were corrected and *de novo* assembled, resulting in 19 contigs. These contigs were manually examined for conflicts observed as sudden change in coverage and were split when necessary, rendering a total of 44 contigs of 722,016 nt average and 44 x coverage and an overall GC content of 48.3 %. The total number of reads obtained from Illumina sequencing was 244,368,646 with an average length of 148 nt after trimming. These reads were mapped to each of the PacBio contigs assembled by the Finishing Module and resulted in an average of 1011 x coverage. A small percentage of gaps, 2-4 nt in length, approximately 2-3 gaps every 150,000 nt were observed in the assembly of Illumina reads compared to the PacBio assembly, and they corresponded to microsatellites being present; thus, in all cases, the PacBio assembly was chosen.

Basic Local Alignment Search Tool (BLAST) [22] was used to analyze fragments of the contigs. The 9360 nt contig uploaded to NCBI as “P.sequoiae\_contig\_43” contained the 18S rDNA gene and internal transcribed spacers of *P. sequoiae* isolate 9LC2. Alignment of this contig to the 5476 nt NCBI entry GU214667.1 from *Passalora sequoiae* had a 99.65 % match, 5457/5476 nt, confirming identity of isolate 9LC2. The 5476 bp sequence entry in GenBank GU214667.1 from *Passalora* *sequoiae*, containing 18S ribosomal RNA gene, internal transcribed spacer (ITS) 1, 5.8S ribosomal RNA (rRNA) gene, ITS2, and 28S rRNA partial sequence [10], was used to locate the same region within the isolate of *P. sequoiae* 9LC2. Once found, the sequence was used in BLAST at NCBI to retrieve 20 closely related sequences with 100 % coverage of that region. Sequences were aligned and trimmed to a 5465 bp that matched 5’ and 3’ ends. A Neighbor Joining [23] maximum likelihood phylogenetic radial tree was constructed using the nucleotide substitution model of Jukes Cantor [24] for DNA substitution rate parameters and Bootstrap with 100 resampling. Data were processed in CLC\_Genomics Workbench 20.0 (Qiagen, Aarhus, Denmark).

Structural annotation of the genome assembly was determined using MAKER v.2.31.8 [25]. The MAKER pipeline, including programs 1) RepeatMasker v.4.0.6 [26] to mask interspersed repeats and low complexity DNA sequences; 2) three gene predictors: GeneMark-ES [27]; SNAP [28], trained with Sordariomycetidae proteins from the Uniprot database; and Augustus [29]; and 3) tRNAscan [30] to identify tRNA genes in the genomic sequence. Predicted transcripts (≥150 bp) and proteins (≥50 amino acids) were analyzed using two databases for analysis of putative genes: dbCAN2 for putative carbohydrate-active enzymes [31], and PHI-base for putative virulence-associated proteins [32]. Blast hits with 80% minimum identity and 70% minimum coverage were kept as significant.