

**Complete chloroplast genome sequence of a white
spruce (*Picea glauca*) genotype from eastern
Canada**

TBD

January 24, 2019

Genome Announcement

The *P. glauca* isolate WS77111 sample was collected from Otonabee-South Monaghan (44°19' 48" N 78°9' 0" W; elevation of 250 m) in southern Ontario. The tissue sample collected was from the needles of the white spruce tree, supplied by John MacKay of Université Laval. These tissue samples were then sequenced at the British Columbia Cancer's Genome Sciences Centre as an objective of the SpruceUp and SMarTForests project.

Genomic DNA libraries were constructed according to British Columbia Cancer's Genome Sciences Centre plate-based and paired-end library protocols on a Microlab NIMBUS liquid handling robot (Hamilton, USA). Briefly, 1 µg of high molecular weight genomic DNA was sonicated (Covaris LE220) in 62.5 µL volume to 400 bp. Sonicated DNA was purified with PCRClean DX magnetic beads (Aline Biosciences). The DNA fragments were end-repaired, phosphorylated and bead purified in preparation for A-tailing using a custom NEB Paired-End Sample Prep Premix Kit (New England Biolabs). Illumina sequencing adapters were ligated overnight at 16°C and adapter ligated products bead purified and enriched with 6 cycles of PCR using primers containing a hexamer index that enables library pooling. Pooled libraries were sequenced with paired-end 250 bp reads on an Illumina HiSeq2500 instrument in rapid mode.

To assemble the complete chloroplast genome, these reads were subsampled in the sizes of read pairs (in millions). Each subset was then assembled using ABySS v2.1.0. Then, each assembly was filtered for the chloroplast genome by alignment to the reference chloroplast genome, the *Picea glauca* isolate PG29 complete chloroplast genome (Genbank KT634228), where only contigs greater or equal to 500 bp that aligned were kept, using BWA v0.7.17. The resulting contigs were pieces of WS77111 chloroplast, assessed using QUAST v5.0.2 and the reference chloroplast genome. Of all the subsets, the 1.5M subset had the least number of misassemblies, with its chloroplast genome in 14 pieces, while the 6M subset had the least number of pieces at 6, but contained 1 misassembly. For this reason, the 1.5M, 6M and the 3M subset (subset in between) were chosen to advance for an ABySS parameter sweep. Of these consequent assemblies, two were chosen to advance to the next step: 1.5M and 6M assembly. LINKS v1.8.5 revealed that the 1.5M and 6M subset contained the fewest pieces and misassemblies. After LINKS, both assemblies had combined the chloroplast genome into one single piece (the largest resulting contig). QUAST analysis showed that the 1.5M subset contained 12 gaps, and the 6M one contained 5 gaps. After using Sealer (part of ABySS) to close the gaps, 1.5M subset had the fewest remaining gaps and thereby determined to be the best assembly. This assembly was then modified to have the same start and end as the reference chloroplast genome using BLAST v.2.7.1. Finally, Pilon v1.22 was run to polish the final assembly.

The WS77111 chloroplast genome is 123,421bp in length, with a GC content of 38.74%.

It has a total of 114 genes: 74 protein-coding genes, 36 tRNA-coding genes, 4 rRNA-coding genes. This chloroplast genome was annotated using GeSeq, with all available *Picea* NCBI RefSeq chloroplast genomes as reference genomes. Five genes required manual annotation: *rps12*, *petB*, *petD*, *rpl16*, and *psbZ*. OGDRAW v1.2 was draw to visualize the circular chloroplast genome map. Adding the chloroplast genome of WS77111 to the resource pool of available *Picea* chloroplasts provides valuable insight to the phylogeny and evolution of these spruce trees.

Accession number(s). The complete chloroplast genome sequence of *Picea glauca*, isolate WS77111 can be found in NCBI Genbank under the accession number MK174379. The tissue samples used can be found under BioSample: SAMN02736787 and BioProject: PRJNA242552. The NCBI accession numbers of the other *Picea* chloroplasts used in annotation are as follows: *Picea abies* (NC_021456), *Picea asperata* (NC_032367), *Picea glauca* (NC_028594), *Picea morrisonicola* (NC_016069), and *Picea sitchensis* (NC_011152).