

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

TBD

February 11, 2019

Abstract

Picea glauca, is a white spruce tree native to eastern Canada. Here we present the complete chloroplast genome sequence of a *Picea glauca*, isolate WS77111. This sequence contributes data to the study of the evolutionary phylogeny of the existing *Picea* organisms, and consequently, facilitating the improvement of Canada's forestry industry.

Genome Announcement

As a part of the SpruceUp (1) and SMarTForests (2) projects, we sequenced, assembled and annotated the chloroplast genome of *Picea glauca*, isolate WS77111. This work contributes to improving the genomic selection in spruce breeding programs, a vital part of Canada’s forestry industry.

The *Picea glauca* isolate WS77111 tissue sample of the needles was collected in southern Ontario (44°19’48”N, 78°9’0”W; elevation: 250m), from Université Laval. Subsequently, the sample was sequenced at the British Columbia Genome Sciences Centre (BCGSC).

To sequence the sample, genomic DNA libraries were constructed according to BCGSC plate-based and paired-end library protocols on a Microlab NIMBUS liquid handling robot (Hamilton, USA). Briefly, 1µg of high molecular weight genomic DNA was sonicated (Covaris LE220) in 62.5µL 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 while adapter ligated products were 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.

For the chloroplast genome assembly, we subsampled the reads into a few million read pairs. ABySS v2.1.0 (3) assembled each subset. Then, BWA v0.7.17 (4) filtered for contigs greater than 500bp that aligned with the reference chloroplast, *Picea glauca* isolate PG29. QUAST v5.0.0 (5) revealed that the 1.5M subset contained zero misassemblies (n=14; N50=18371), while the 6M subset contained the least contigs (n=6; N50=43759). Therefore, we selected the 1.5M, 6M and 3M subset (middle ground) for a parameter sweep, where the 3M subset was eliminated. After LINKS v1.8.5 (6), both assemblies combined the contigs into one piece. QUAST alluded that the 1.5M subset contained 12 gaps, whereas the 6M contained five. We then employed ABySS-sealer to close the gaps in both subsets, discovering that 1.5M subset contained the fewest gaps. We maintained our product consistency by modifying our assembly to match the reference using BLAST v.2.7.1 (7) and ultimately polished the final assembly with a runthrough from Pilon v1.22 (8).

The WS77111 chloroplast genome is 123,421-bp in length, with a GC content of 38.74%. GeSeq (9) annotated 114 genes: 74 protein-coding, 36 tRNA-coding, and four rRNA-coding genes, using all available *Picea* chloroplast genomes as reference. Only five genes required manual annotation: *rps12*, *petB*, *petD*, *rpl16*, and *psbZ*. OGDRAW v1.2 (10) drew the

circular chloroplast genome map. By adding this new chloroplast genome to the available resource pool, further analysis of phylogeny and evolution of these spruce trees can be conducted to achieve the project’s objective.

Accession number(s). The complete chloroplast genome sequence of *Picea glauca*, isolate WS77111 can be found in Genbank under MK174379. The tissue samples used can be found with BioSample: SAMN02736787 and BioProject: PRJNA242552. The annotation NCBI references are as follows: *Picea abies* (NC_021456), *Picea asperata* (NC_032367), *Picea glauca* isolate PG29 (NC_028594), *Picea morrisonicola* (NC_016069), and *Picea sitchensis* (NC_011152).

Acknowledgements

This work was conducted as a part of the SpruceUp and SMarTForests projects.

References

3. Jackman SD, Vandervalk BP, Mohamadi H, Chu J, Yeo S, Hammond SA, Jahesh G, Khan H, Coombe L, Warren RL, Birol I. 2017. ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter. *Genome research* 27:768-777.
4. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* (Oxford, England) 25:1754-1760.
5. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* (Oxford, England) 29:1072-1075.
6. Warren RL, Yang C, Vandervalk BP, Behsaz B, Lagman A, Jones SJM, Birol I. 2015. LINKS: Scalable, alignment-free scaffolding of draft genomes with long reads. *GigaScience* 4:35-35.
7. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215:403-410.
8. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PloS one* 9:e112963-e112963.
9. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. 2017. GeSeq - versatile and accurate annotation of organelle genomes. *Nucleic acids research* 45:W6-W11.

10. Lohse M, Drechsel O, Kahlau S, Bock R. 2013. OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic acids research* 41:W575-W581.