

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

Diana Lin^a, Lauren Coombe^a, Shaun D. Jackman^a, Kristina K. Gagalova^a, René L. Warren^a, S. Austin Hammond^{a*}, Heather Kirk^a, Pawan Pandoh^a, Yongjun Zhao^a, Richard A. Moore^a, Andrew J. Mungall^a, Carol Ritland^{b,f}, Barry Jaquish^c, Nathalie Isabel^d, Jean Bousquet^e, Steven J.M. Jones^a, Joerg Bohlmann^{b,f}, and Inanc Birol^a

^aCanada's Michael Smith Genome Sciences Centre, BC Cancer, Vancouver, BC, Canada

^bDepartment of Forest and Conservation Sciences, University of British Columbia, Vancouver, BC, Canada

^cBritish Columbia Ministry of Forests, Lands and Natural Resource Operations, Tree Improvement Branch, Kalamalka Forestry Centre, Vernon, BC, Canada

^dNatural Resources Canada, Laurentian Forestry Centre, Quebec City, QC, Canada

^eCanada Research Chair in Forest Genomics, Université Laval, Quebec City, QC, Canada

^fMichael Smith Laboratories, University of British Columbia, Vancouver, BC Canada

Running Head: Complete chloroplast genome of a *Picea glauca* isolate

#Address correspondence to Inanc Birol, ibirol@bcgsc.ca

*Current address: S. Austin Hammond, Next-Generation Sequencing Facility, University of Saskatchewan, Saskatoon, SK, Canada.

Abstract

Here we present the complete chloroplast genome sequence of white spruce (*Picea glauca*, isolate WS77111), a coniferous tree widespread in the boreal forests of North America. This sequence contributes to genomic and phylogenetic analyses of the *Picea* genus, part of ongoing research to understand their adaptation to environmental stress.

Genome Announcement

We sequenced, assembled, and annotated the chloroplast genome of white spruce (*Picea glauca* genotype WS77111), a dominant species in the Canadian boreal forest (1). Over tens of millions of years, conifers such as *P. glauca* have evolved to cope with adverse environmental conditions (2). However, climate change poses new challenges to the capacity of these trees to adapt to conditions such as prolonged drought and increased pressure from forest insect pests (3). *P. glauca* has three different genomes, a nuclear genome, a mitochondrial genome, and a plastid (i.e. chloroplast) genome. In general, chloroplast genomes are derived from the ancestral genomes of the microbial endosymbiont from which these organelles originated (4). The nuclear genome of *P. glauca* (WS77111) was published in 2015 (5).

A *P. glauca* (genotype WS77111) needle tissue sample was collected in southeastern Ontario (44°19'48"N, 78°9'0"W; elevation: 250m). The sample was sequenced at Canada's Michael Smith Genome Sciences Centre (GSC).

To sequence the sample, genomic DNA libraries were constructed according to GSC plate-based and paired-end library protocols on a Microlab NIMBUS liquid handling robot (Hamilton, USA), and sonicated into 400 bp fragments, as previously described (6-7). Pooled libraries were sequenced with paired-end 250 bp reads on an Illumina HiSeq2500 instrument in rapid mode.

To assemble the chloroplast genome, we generated various random subsamples of the full read set (0.75, 1.5, 3, 6, 12, 25, 50, 200 million read pairs), and assembled each subset with ABySS v2.1.0 (8) ($k = 128$, $kc = 3$). The 1.5M, 3M, and 6M read subsets produced the best ABySS chloroplast genome assemblies ($N_{50} = 3692, 1313, 949$, respectively), as determined by comparing these assemblies to the white spruce admix (PG29) chloroplast genome (NCBI accession NC_028594.1; (9)) using QUAST v5.0.0 (10). We then performed additional ABySS assemblies with varying k and kc parameters using these three subsets ($k = 96, 112, 128, 144, 160$, $kc = 3, 4$). The assembly with the fewest aligning contigs ($n = 14$) and fewest misassemblies (1.5M read pairs, $k = 96$, $kc = 3$) was chosen for further scaffolding. Scaffolding the assembly using LINKS v1.8.5 (11) and the PG29 chloroplast genome joined the contigs into one piece. We then used Sealer (12) to close the scaffold gaps. We modified the start position of our assembly to match the PG29 reference using BLAST v.2.7.1 (13), and polished the final assembly with Pilon v1.22 (14).

The complete WS77111 chloroplast genome is 123,421 bp long, with 38.74% GC content. Using GeSeq (15) with several other *Picea* chloroplast genomes as reference (9, 16), we annotated 114 genes: 74 protein-coding, 36 tRNA-coding, and four rRNA-coding genes. Only *rps12*, *petB*, *petD*, *rpl16*, and *psbZ* required manual annotation. The genome map in

Figure 1 was generated using OGDRAW v1.2 (17).

The assembly of this new chloroplast genome will enable further analysis of *Picea* phylogeny and genetics.

Accession number(s). The complete chloroplast genome sequence of *Picea glauca*, genotype WS77111 is available from Genbank under accession MK174379, and the raw reads are in the SRA under SRX525336. The annotations used as references were from *Picea abies* (NC_021456), *Picea asperata* (NC_032367), *Picea glauca* genotype PG29 (NC_028594), *Picea morrisonicola* (NC_016069), and *Picea sitchensis* (NC_011152, KU215903).

Figures and Data

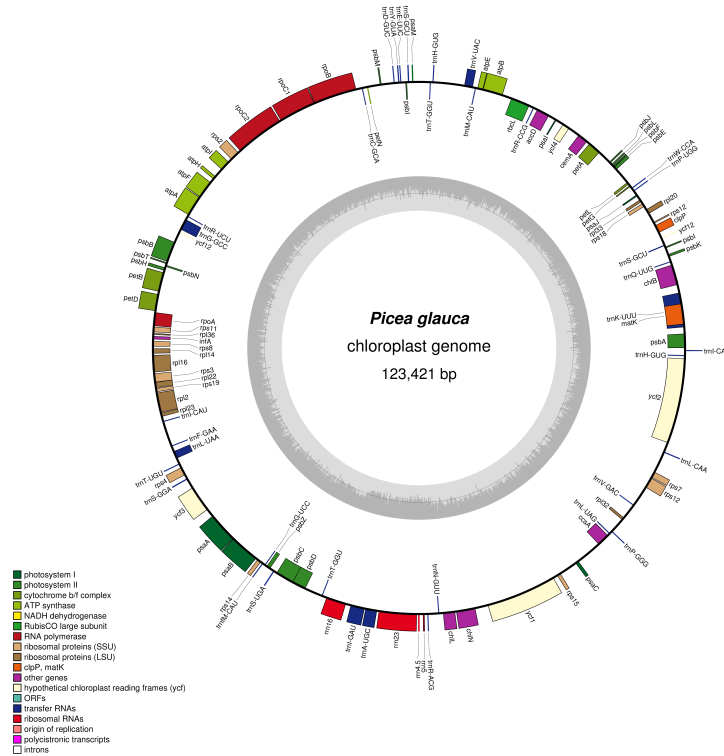


Figure 1: The complete chloroplast genome of *Picea glauca*, isolate WS77111. The *Picea glauca* chloroplast genome was annotated using GeSeq (15), and plotted using OGDRAW (17). The inner grey circle illustrates the GC content of the genome.

Acknowledgements

This work was supported by funds from Genome Canada, Genome BC, and Genome Quebec as part of the Spruce-Up (www.spruce-up.ca) [243FOR] and AnnoVis [281ANV] projects.

References

1. Li P, Beaulieu J, Bousquet J. 1997. Genetic structure and patterns of genetic variation among populations in eastern white spruce (*Picea glauca*). *Can J For Res* 27:189-198.
2. Bouillé M, Bousquet J. 2005. Trans-species shared polymorphisms at orthologous nuclear gene loci among distant species in the conifer *Picea* (*Pinaceae*): implications for the long-term maintenance of genetic diversity in trees. *Am J Bot* 92:63-73
3. Kiss GK, Yanchuk AD. 1991. Preliminary evaluation of genetic variation of weevil resistance in interior spruce in British Columbia. *Can J For Res* 21:230-234.
4. Ku C, Nelson-Sathi S, Roettger M, Sousa FL, Lockhart PJ, Bryant D, Hazkani-Covo E, McInerney JO, Landan G, Martin WF. 2015. Endosymbiotic origin and differential loss of eukaryotic genes. *Nature* 524:427-432.
5. Warren RL, Keeling CI, Yuen MMS, Raymond A, Taylor GA, Vandervalk BP, Mohamadi H, Paulino D, Chiu R, Jackman SD, Robertson G, Yang C, Boyle B, Hoffmann M, Weigel D, Nelson DR, Ritland C, Isabel N, Jaquish B, Yanchuk A, Bousquet J, Jones SJM, Mackay J, Birol I, Bohlmann J. 2015. Improved white spruce (*Picea glauca*) genome assemblies and annotation of large gene families of conifer terpenoid and phenolic defense metabolism. *Plant J* 83:189-212
6. Jones MR, Schrader KA, Shen Y, Pleasance E, Chng C, Dar N, Yip S, Renouf DJ, Schein JE, Mungall AJ, Zhao Y, Moore R, Ma Y, Sheffield BS, Ng T, Jones SJM, Marra MA, Laskin J, Lim HJ. 2016. Response to angiotensin blockade with irbesartan in a patient with metastatic colorectal cancer. *Ann Oncol* 27:801-806.
7. Tsang ES, Shen Y, Chooback N, Ho C, Jones M, Renouf DJ, Lim HJ, Sun S, Yip S, Pleasance E, Ma Y, Zhao Y, Mungall AJ, Moore R, Jones S, Marra M, Laskin JJ. 2017. Clinical outcomes after whole genome sequencing in patients with metastatic non-small cell lung cancer. *J Clin Oncol* 35.
8. 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 Res* 27:768-777.

9. Jackman SD, Warren RL, Gibb EA, Vandervalk BP, Mohamadi H, Chu J, Raymond A, Pleasance S, Coope R, Wildung MR, Ritland CE, Bousquet J, Jones SJM, Bohlmann J, Birol I. 2016. Organellar genomes of white spruce (*Picea glauca*): Assembly and annotation. *Genome Biol Evol* 8:29–41.
10. Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. 2018. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* 34:i142–i150.
11. 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.
12. Paulino D, Warren RL, Vandervalk BP, Raymond A, Jackman SD, Birol I. 2015. Sealer: a scalable gap-closing application for finishing draft genomes. *BMC Bioinformatics* 16.
13. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410.
14. 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.
15. 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 Res* 45.
16. Coombe L, Warren RL, Jackman SD, Yang C, Vandervalk BP, Moore RA, Pleasance S, Coope RJ, Bohlmann J, Holt RA, Jones SJM, Birol I. 2016. Assembly of the complete Sitka spruce chloroplast genome using 10X Genomics’ GemCode sequencing data. *PLoS One* 11.
17. 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 Res* 41:W575–W581.