



Complete Chloroplast Genome Sequence of a White Spruce (*Picea glauca*, Genotype WS77111) 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} 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

^dLaurentian Forestry Centre, Natural Resources Canada, 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

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

Over tens of millions of years, conifers such as the white spruce (*Picea glauca*) have evolved to cope with adverse environmental conditions (1, 2), such as prolonged drought and increased pressure from forest insect pests (3). Plants have three different genomes, namely, a nuclear, a mitochondrial, 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* (genotype 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, 250 m). Genomic DNA was extracted from 60 g of tissue by Bio S&T using an organelle exclusion method yielding 300 µg of high-quality purified nuclear DNA, as previously described (6). The sample was sequenced at Canada's Michael Smith Genome Sciences Centre (GSC). Here, we report on the assembled and annotated chloroplast genome sequence of this genotype.

To sequence the sample, genomic DNA libraries were constructed according to the plate-based and paired-end library protocols at the GSC on a Microlab Nimbus liquid-handling robot (Hamilton, USA). Briefly, 1 µg of genomic DNA was sonicated (Covaris LE220) in 62.5 µl to 400 bp and purified with PCRclean DX magnetic beads (Aline Biosciences). Illumina sequencing adapters were ligated overnight at 16°C. Pooled libraries were sequenced with paired-end 250-bp reads on an Illumina HiSeq 2500 instrument in rapid mode. Using this protocol, four libraries were generated, sequencing approximately 400 million reads from each one.

To assemble this genome sequence, we generated various random subsamples of read pairs from one lane of one library (i.e., 42,881,319 read pairs), producing subsets with 21×, 43×, 88×, 172×, 345×, 711×, 1,219×, and 5,619× coverage of the chloroplast genome. Each subset was assembled with ABySS v2.1.0 (7) (using the parameters $k = 128$ and $kc = 3$). Due to the large number of chloroplasts per cell, the chloroplast genome would be sequenced at a very high coverage over a full lane of data. Therefore, by subsampling the full data set, the coverages of the nuclear and

Citation Lin D, Coombe L, Jackman SD, Gagalova KK, Warren RL, Hammond SA, Kirk H, Pandoh P, Zhao Y, Moore RA, Mungall AJ, Ritland C, Jaquish B, Isabel N, Bousquet J, Jones SJM, Bohlmann J, Birol I. 2019. Complete chloroplast genome sequence of a white spruce (*Picea glauca*, genotype WS77111) from eastern Canada. Microbiol Resour Announc 8:e00381-19. <https://doi.org/10.1128/MRA.00381-19>.

Editor Antonis Rokas, Vanderbilt University

Copyright © 2019 Lin et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

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

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

Received 9 April 2019

Accepted 13 May 2019

Published 6 June 2019

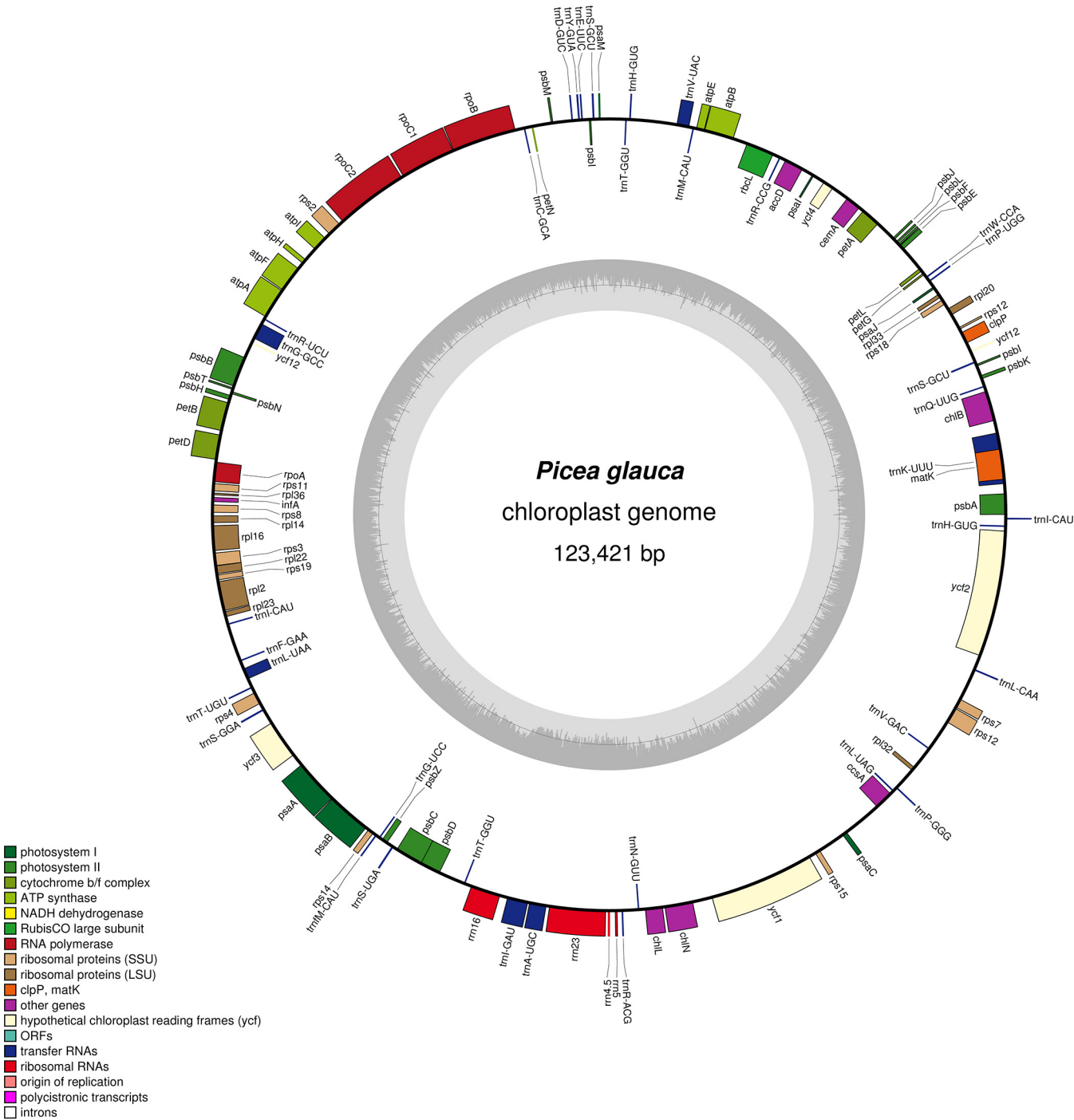


FIG 1 Complete chloroplast genome of *Picea glauca* genotype WS77111. The *Picea glauca* chloroplast genome was annotated using GeSeq v1.65 (16) and plotted using OrganellarGenomeDRAW v1.2 (17). The inner gray circle illustrates the G+C content of the genome.

mitochondrial genomes were lowered to a level where these sequences do not assemble well, while the coverage of the chloroplast genome was still sufficient enough for a high-quality assembly. The 43×, 88×, and 172× subsets produced the best ABYSS assemblies (N_{50} lengths, 3,692, 1,313, and 949 bp, respectively), as determined by a QUAST analysis (v5.0.0) (8). For comparison, we used the white spruce admix (PG29) chloroplast genome (NCBI GenBank accession number [NC_028594](https://www.ncbi.nlm.nih.gov/nuccore/NC_028594)) (9), the published chloroplast genome that is most closely related to the WS77111 genotype. The use of this admix as a reference was established previously (10), as it is a

naturally occurring ingress of *Picea glauca*, *Picea engelmannii*, and *Picea sitchensis* (5). We then performed additional ABySS assemblies with various *k* and *kc* parameters using these three subsets (*k* = 96, 112, 128, 144, and 160; *kc* = 3 and 4). The assembly with the fewest aligning contigs (*n* = 14) and fewest misassemblies (43×; *k* = 96; *kc* = 3) was chosen for further scaffolding with the PG29 chloroplast genome, using LINKS v1.8.5 (11), thereby joining the contigs into one piece. We then used Sealer v2.1.0 (12) to close the scaffold gaps. To be consistent with previously published chloroplast genomes when reporting gene annotations, we adjusted the start position of our assembly using BLAST v.2.7.1 (13) and polished the final assembly with Pilon v1.22 (14), using BWA v0.1.7 (15) for read alignment.

The complete genotype WS77111 chloroplast genome is 123,421 bp long, with a G+C content of 38.74%. Using GeSeq v1.65 (16) with several *Picea* sp. chloroplast genomes as references (9, 10), we annotated 114 genes, namely, 74 protein-coding, 36 tRNA-coding, and 4 rRNA-coding genes. Five genes (*rps12*, *petB*, *petD*, *rpl16*, and *psbZ*) required manual annotation. The genome map in Fig. 1 was generated using OrganellarGenomeDRAW v1.2 (17).

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

Data availability. The complete chloroplast genome sequence of *Picea glauca*, genotype WS77111, is available in GenBank under accession number [MK174379](#), and the raw reads are in the SRA under accession numbers [SRX525336](#) and [SRR1259605](#). The annotations used as references were from *Picea abies* (GenBank accession number [NC_021456](#)), *Picea asperata* (GenBank accession number [NC_032367](#)), *Picea glauca* genotype PG29 (GenBank accession number [NC_028594](#)), *Picea morrisonicola* (GenBank accession number [NC_016069](#)), and *Picea sitchensis* (GenBank accession numbers [NC_011152](#) and [KU215903](#)).

ACKNOWLEDGMENTS

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

- 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. [https://doi.org/10.1139/x96-159](#).
- Bouill   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. [https://doi.org/10.3732/ajb.92.1.63](#).
- 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. [https://doi.org/10.1139/x91-028](#).
- 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. [https://doi.org/10.1038/nature14963](#).
- 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. [https://doi.org/10.1111/tpj.12886](#).
- Birol I, Raymond A, Jackman SD, Pleasance S, Coope R, Taylor GA, Yuen MMS, Keeling CI, Brand D, Vandervalk BP, Kirk H, Pandoh P, Moore RA, Zhao Y, Mungall AJ, Jaquish B, Yanchuk A, Ritland C, Boyle B, Bousquet J, Ritland K, Mackay J, Bohlmann J, Jones SJ. 2013. Assembling the 20 Gb white spruce (*Picea glauca*) genome from whole-genome shotgun sequencing data. *Bioinformatics* 29:1492–1497. [https://doi.org/10.1093/bioinformatics/btt178](#).
- 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. [https://doi.org/10.1101/gr.214346.116](#).
- Mikheenko A, Pribelski A, Saveliev V, Antipov D, Gurevich A. 2018. Versatile genome assembly evaluation with QAST-LG. *Bioinformatics* 34:i142–i150. [https://doi.org/10.1093/bioinformatics/bty266](#).
- 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. [https://doi.org/10.1093/gbe/evv244](#).
- 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:e0163059. [https://doi.org/10.1371/journal.pone.0163059](#).
- 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. [https://doi.org/10.1186/s13742-015-0076-3](#).
- 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* 16:230. [https://doi.org/10.1186/s12859-015-0663-4](#).
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](#).
- 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. <https://doi.org/10.1371/journal.pone.0112963>.
15. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
16. 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:W6–W11. <https://doi.org/10.1093/nar/gkx391>.
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. <https://doi.org/10.1093/nar/gkt289>.