



Complete Chloroplast Genome Sequence of an Engelmann Spruce (*Picea engelmannii*, Genotype Se404-851) from Western Canada

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

^dForest Research Centre and Institute for Systems and Integrative Biology, Université Laval, Québec City, QC, Canada

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

ABSTRACT Engelmann spruce (*Picea engelmannii*) is a conifer found primarily on the west coast of North America. Here, we present the complete chloroplast genome sequence of *Picea engelmannii* genotype Se404-851. This chloroplast sequence will benefit future conifer genomic research and contribute resources to further species conservation efforts.

We sequenced, assembled, and annotated the complete chloroplast genome of Engelmann spruce (*Picea engelmannii*, genotype Se404-851). The Engelmann spruce dominates much of the large spruce forests of interior British Columbia, where it has been reported to hybridize with *Picea glauca* and *Picea sitchensis* (1), and its range extends southward to New Mexico. The tree has three different genomes, a nuclear genome, a mitochondrial genome, and a plastid genome (i.e., chloroplast). In general, chloroplast genomes are derived from the ancestral genomes of the microbial endosymbiont from which these organelles originated (2).

A tissue sample was collected from a 13-year-old Engelmann spruce grown at the Kalamalka Forestry Centre in British Columbia (50°14'38.4"N, 119°16'40.8"W; elevation, 450 m) and planted from a seed from Don Fernando Mountain, New Mexico (36°17'60"N, 105°24'0"W; elevation, 2,987 m). Genomic DNA was extracted from 60 g tissue by Bio S&T using an organelle exclusion method yielding 300 μg of high-quality purified nuclear DNA, as previously described (3). The sample was sequenced at Canada's Michael Smith Genome Sciences Centre.

To sequence the sample, a 900-bp whole-genome library was constructed following a previously described protocol (4, 5) with minor modifications. Briefly, 5 μg of genomic DNA was subjected to shearing by sonication (Covaris LE220) using a duty factor of 5 and peak incident power of 450 for 70 seconds. The sonicated DNA products were fractionated in a 6% PAGE gel to recover fragments greater than 700 bp for library preparation. These PCR-free libraries were sequenced with paired-end 150-base reads on an Illumina HiSeq X platform using V4 chemistry according to the manufacturer's recommendations. With this protocol, four libraries were generated, sequencing approximately 200 million reads from each of them.

To assemble the chloroplast genome, we subsampled the whole-genome shotgun sequencing reads of one lane of one library (i.e., 41,748,620 read pairs) to subsets of 0.75, 1.5, 3, 6, 12, 25, and 41 million read pairs and then assembled each subset with ABySS v2.1.1 (6) (*k*-mer size [*k*], 128; *k*-mer count [*kc*], 3). The ABySS assembly of the

Citation Lin D, Coombe L, Jackman SD, Gagalova KK, Warren RL, Hammond SA, McDonald H, Kirk H, Pandoh P, Zhao Y, Moore RA, Mungall AJ, Ritland C, Doerksen T, Jaquish B, Bousquet J, Jones SJM, Bohlmann J, Birol I. 2019. Complete chloroplast genome sequence of an Engelmann spruce (*Picea engelmannii*, genotype Se404-851) from western Canada. *Microbiol Resour Annot* 8:e00382-19. <https://doi.org/10.1128/MRA.00382-19>.

Editor Jason E. Stajich, University of California, Riverside

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 11 April 2019

Accepted 20 May 2019

Published 13 June 2019

sequence, and ran Sealer v2.1.1 (10), closing the “end” gap and removing overlapping sequences as previously described (11). Finally, the resulting assembly was polished using Pilon v1.22 (12) using the 3-million subset of read pairs aligned with the Burrows-Wheeler Aligner (BWA) v0.1.7 (13).

The complete *P. engelmannii* genotype Se404-851 chloroplast genome is 123,542 bp long with a 38.74% GC content. Using GeSeq v1.65 (14) and using other *Picea* chloroplast genomes as references (7, 11), we annotated 114 genes comprising 74 protein-coding genes, 36 tRNA-coding genes, and 4 rRNA-coding genes. We note that four genes (*rps12*, *petB*, *petD*, and *rpl16*) in this list were manually annotated. We used OrganellarGenomeDRAW v1.2 (15) to generate the map in Fig. 1.

The introduction of this new chloroplast genome will benefit conifer genomic research and inform future evolutionary studies.

Data availability. The complete chloroplast genome sequence of *Picea engelmannii* genotype Se404-851 is available under GenBank accession number [MK241981](https://www.ncbi.nlm.nih.gov/nuclink/MK241981), and the raw reads are available under SRA numbers [SRX5070635](https://www.ncbi.nlm.nih.gov/sra/SRX5070635) and [SRR8252852](https://www.ncbi.nlm.nih.gov/sra/SRR8252852). The annotations used as references were from *Picea abies* (GenBank accession number [NC_021456](https://www.ncbi.nlm.nih.gov/nuclink/NC_021456)), *Picea asperata* (GenBank accession number [NC_032367](https://www.ncbi.nlm.nih.gov/nuclink/NC_032367)), *Picea glauca* genotype PG29 (GenBank accession number [NC_028594](https://www.ncbi.nlm.nih.gov/nuclink/NC_028594)), *Picea morrissonicola* (GenBank accession number [NC_016069](https://www.ncbi.nlm.nih.gov/nuclink/NC_016069)), and *Picea sitchensis* (GenBank accession numbers [NC_011152](https://www.ncbi.nlm.nih.gov/nuclink/NC_011152) and [KU215903](https://www.ncbi.nlm.nih.gov/nuclink/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

- Sutton BCS, Flanagan DJ, Gawley JR, Newton CH, Lester DT, El-Kassaby YA. 1991. Inheritance of chloroplast and mitochondrial DNA in *Picea* and composition of hybrids from introgression zones. *Theor Appl Genet* 82:242–248. <https://doi.org/10.1007/BF00226220>.
- 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>.
- Biol 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>.
- Jones MR, Schrader KA, Shen Y, Pleasance E, Ch'ng 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. <https://doi.org/10.1093/annonc/mdw060>.
- 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:e20563. https://doi.org/10.1200/JCO.2017.35.15_suppl.e20563.
- Jackman SD, Vandervalk BP, Mohamadi H, Chu J, Yeo S, Hammond SA, Jahesh G, Khan H, Coombe L, Warren RL, Biol 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>.
- 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, Biol 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>.
- 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>.
- 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](https://doi.org/10.1016/S0022-2836(05)80360-2).
- Paulino D, Warren RL, Vandervalk BP, Raymond A, Jackman SD, Biol 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>.
- Coombe L, Warren RL, Jackman SD, Yang C, Vandervalk BP, Moore RA, Pleasance S, Coope RJ, Bohlmann J, Holt RA, Jones SJM, Biol 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>.
- 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>.
- 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>.
- 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>.
- 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>.