

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} 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* genotype

#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*, 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, part of ongoing research to understand their adaptation to environmental stress.

GENOME ANNOUNCEMENT

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: 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* (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). Genomic DNA was extracted from 60 gm tissue by BioS&T, using an organelle exclusion method yielding 300µg 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 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, 1µg of genomic DNA was sonicated (Covaris LE220) in 62.5µL to 400-bp, purifying 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 HiSeq2500 instrument in rapid mode. Using this protocol, four libraries were generated, sequencing approximately 400 million reads from each.

To assemble this genome, we generated various random subsamples of read pairs from one lane of one library (i.e. 42,881,319 read pairs) producing subsets with 21X, 43X, 88X, 172X, 345X, 711X, 1219X, and 5619X coverage of the chloroplast genome. Each subset was assembled with ABySS v2.1.0 (7) ($k=128$, $kc=3$). Due to the large number of chloroplasts per cell, the chloroplast genome will be sequenced at a very high coverage over a full lane of data. Therefore, by subsampling the full dataset, the coverages of the nuclear and 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 43X, 88X, and 172X subsets produced the best ABySS assemblies (N50 length = 3692, 1313, 949 bp respectively), as determined by a QUAST analysis (v5.0.0). For comparison, we used the white spruce admix (PG29) chloroplast genome (NCBI accession NC_028594.1; (8)), the most

64 closely related published chloroplast genome to the WS77111 genotype. The use of this admix as
65 a reference has been established previously (10), as it is a naturally occurring ingress of *Picea*
66 *glauca*, *Picea engelmannii*, and *Picea sitchensis* (5). We then performed additional ABySS
67 assemblies with varying *k* and *kc* parameters using these three subsets (*k*=96, 112, 128, 144, 160,
68 *kc*=3, 4). The assembly with the fewest aligning contigs (*n*=14) and fewest misassemblies (43X,
69 *k*=96, *kc*=3) was chosen for further scaffolding with the PG29 chloroplast genome, using LINKS
70 v1.8.5 (11), thereby joining the contigs into one piece. We then used Sealer v2.1.0 (12) to close
71 the scaffold gaps. To be consistent with previously published chloroplast genomes when
72 reporting gene annotations, we adjusted the start position of our assembly using BLAST v.2.7.1
73 (13), and polished the final assembly with Pilon v1.22 (14), using BWA v0.1.7 (15) for read
74 alignment.

75
76 The complete WS77111 chloroplast genome is 123,421-bp long, with 38.74% GC
77 content. Using GeSeq v1.65 (16) with several *Picea* chloroplast genomes as reference (8, 10), we
78 annotated 114 genes: 74 protein-coding, 36 tRNA-coding, and four rRNA-coding genes. Four
79 genes (*rps12*, *petB*, *petD*, *rpl16*, and *psbZ*) required manual annotation. The genome map in
80 Figure 1 was generated using OGDRAW v1.2 (17).

81
82 The assembly of this new chloroplast genome will enable further analysis of *Picea*
83 phylogeny and genetics.

84
85 **Accession number(s).** The complete chloroplast genome sequence of *Picea glauca*, genotype
86 WS77111 is available from Genbank under accession [MK174379](#), and the raw reads are in the
87 SRA under [SRX525336](#) and [SRR1259605](#). The annotations used as references were from *Picea*
88 *abies* ([NC_021456](#)), *Picea asperata* ([NC_032367](#)), *Picea glauca* genotype PG29 ([NC_028594](#)),
89 *Picea morrisonicola* ([NC_016069](#)), and *Picea sitchensis* ([NC_011152](#), [KU215903](#)).

90 91 **ACKNOWLEDGEMENTS**

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

95 96 97 **REFERENCES**

- 98 1. Li P, Beaulieu J, Bousquet J. 1997. Genetic structure and patterns of genetic variation
99 among populations in eastern white spruce (*Picea glauca*). Can J For Res 27:189-198.
- 100 2. Bouillé M, Bousquet J. 2005. Trans-species shared polymorphisms at orthologous
101 nuclear gene loci among distant species in the conifer *Picea* (*Pinaceae*): implications for the
102 long-term maintenance of genetic diversity in trees. Am J Bot 92:63-73

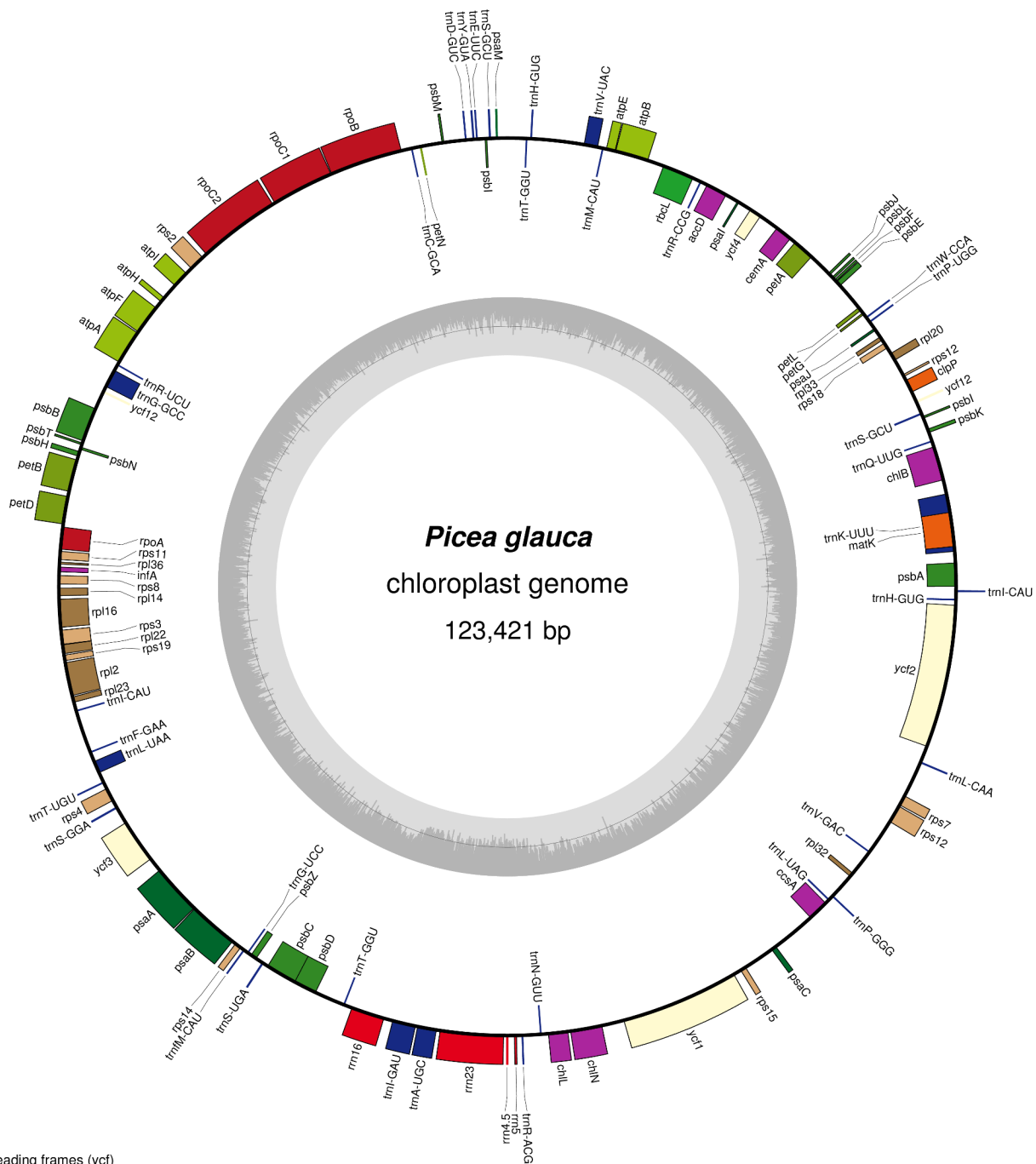
3. Kiss GK, Yanchuk AD. 1991. Preliminary evaluation of genetic variation of weevil resistance in interior spruce in British Columbia. *Can J For Res* 21:230–234.
4. Ku C, Nelson-Sathi S, Roettger M, Sousa FL, Lockhart PJ, Bryant D, Hazkani-Covo E, McInerney JO, Landan G, Martin WF. 2015. Endosymbiotic 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. 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.
7. 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.
8. 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.
9. Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. 2018. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* 34:i142–i150.
10. 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.
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. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760.

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.
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.

Figure 1. The complete chloroplast genome of *Picea glauca* genotype WS77111

The *Picea glauca* chloroplast genome was annotated using GeSeq v1.65 (16), and plotted using OGDRAW v1.2 (17). The inner grey circle illustrates the GC content of the genome.

Picea glauca
chloroplast genome
123,421 bp



- photosystem I
- photosystem II
- cytochrome b/f complex
- ATP synthase
- NADH dehydrogenase
- RubisCO large subunit
- RNA polymerase
- ribosomal proteins (SSU)
- ribosomal proteins (LSU)
- clpP, matK
- other genes
- hypothetical chloroplast reading frames (ycf)
- ORFs
- transfer RNAs
- ribosomal RNAs
- origin of replication
- polycistronic transcripts
- introns