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

Diana Lin, Lauren Coombe, Kristina Gagalova, René L. Warren, Stewart A. Hammond, Heather Kirk, Pawan Pandoh, Yongjun Zhao, Richard A. Moore, Andrew J. Mungall, Carol Ritland, Barry Jaquish, Jean Bousquet, Steven J.M. Jones, Joerg Bohlmann, Inanc Birol

March 20, 2019

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 *Picea glauca* (isolate WS77111), a dominant species in the Canadian boreal forest (1). Conifers such as *P. glauca* have demonstrated great endurance to external stressors, from extreme climates and natural disasters to infestations, tolerating ice ages (2), forest fires (3), and invasive species (4). With the current threat of climate change, their ability to adapt is crucial to the survival of the species. This work contributes to future analyses of adaptation and resistance, which can inform genomic selection in spruce breeding programs.

A *P. glauca* (isolate WS77111) needle tissue sample was collected in southern 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 (5-6). 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 (7) (k=128, kc=3). The 1.5M, 3M and 6M read subsets produced the best ABySS assemblies, as determined by comparing these assemblies to the white spruce admix (PG29) chloroplast genome (NCBI accession NC_028594.1) using QUAST v5.0.0 (8). We then performed additional ABySS assemblies with varying k and kc parameters of the software using these three subsets (k=96, 112, 128, 144, 160, kc=3, 4). The assembly with the fewest aligning contigs and fewest misassemblies (1.5M read pairs, k=96, kc=3) was chosen for further scaffolding. Scaffolding the assembly using LINKS v1.8.5 (9) and the PG29 chloroplast genome joined the contigs into one piece. We then used Sealer (10) to close the scaffold gaps. We modified the start position of our assembly to match the PG29 reference using BLAST v.2.7.1 (11), and polished the final assembly with Pilon v1.22 (12).

The complete WS77111 chloroplast genome is 123,421-bp long, with 38.74% GC content. Using GeSeq (13) with several other *Picea* chloroplast genomes as reference, 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 (14).

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*, isolate 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* isolate PG29 (NC_028594), *Picea morrisonicola* (NC_016069), and *Picea sitchensis* (NC_011152).

Figures and Data

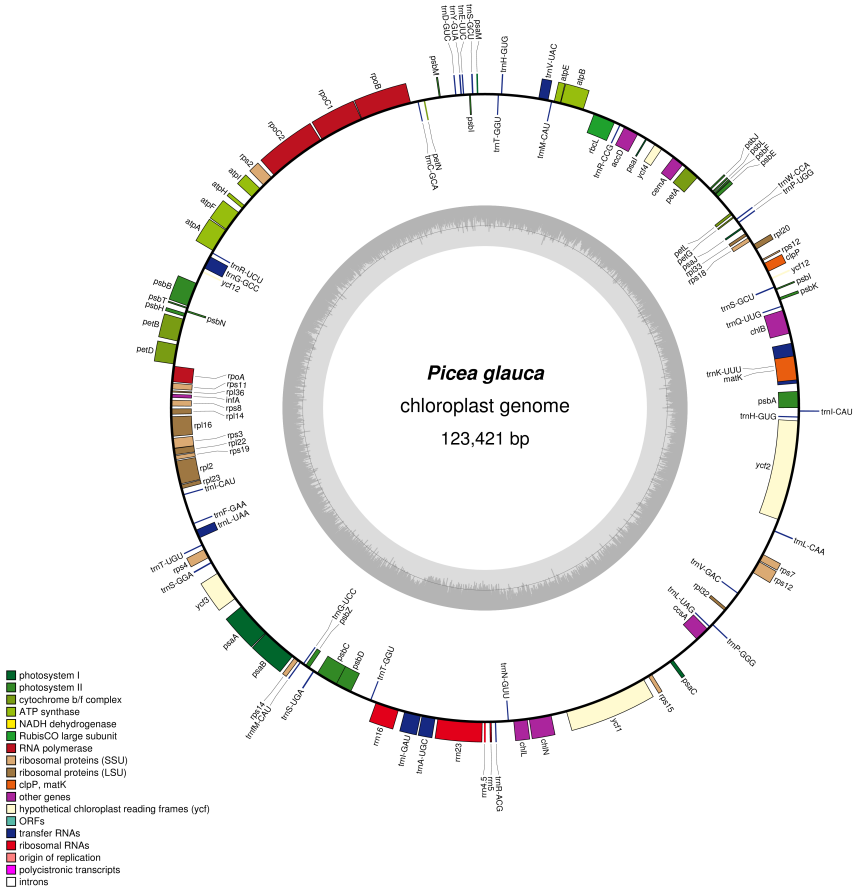

Figure 1: The complete chloroplast genome of *Picea glauca*, isolate WS77111. The *Picea glauca* chloroplast genome was annotated using GeSeq (13), and plotted using OGDRAW (14). The inner grey circle illustrates the GC content of the genome.

Acknowledgements

This work was supported by funds from Genome Canada and Genome BC [243FOR, 281ANV] as part of the SpruceUp (`www.spruce-up.ca`) project.

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. Anderson LL, Hu FS, Paige KN. 2010. Phylogeographic History of White Spruce During the Last Glacial Maximum: Uncovering Cryptic Refugia. J Hered 102:207–216.

3. Arbellay E, Stoffel M, Sutherland EK, Smith KT, Falk DA. 2014. Changes in tracheid and ray traits in fire scars of North American conifers and their ecophysiological implications. Ann Bot 114:223–232.

4. 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.

5. 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.

6. 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.

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. Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. 2018. Versatile genome assembly evaluation with QUAST-LG. Bioinformatics 34:i142–i150.

9. 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.

10. 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.

11. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403-410.

12. 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.

13. 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.

14. 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.