Complete chloroplast genome sequence of *Picea engelmannii*, isolate Se404-851 from western Canada

Diana Lin, Lauren Coombe, Kristina Gagalova, Rene 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

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*, isolate Se404-851. This chloroplast sequence will benefit future conifer genomic research and contribute resources to further species conservation efforts.

Genome Announcement

We sequenced, assembled and annotated the complete chloroplast genome of *Picea engelmannii* (isolate Se404-851). Engelmann spruce, the native host of the white pine weevil, has the highest susceptibility to insect infestation among the *Picea* genus (1). Research into the adaptation of *P. engelmannii* to insect resistance is key to help maintain the health of our forests.

A needle tissue sample was collected from a 13-year-old Engelmann spruce grown at the Kalamalka Forestry Centre in British Columbia ($36°17'60"$N, $105°24'0"$W; elevation: 298m), planted from a seed from Don Fernando Mountain, New Mexico ($50°14'38.4"$N $119°16'40.8"$W; elevation of 450m). The sample was sequenced at Canada's Michael Smith Genome Sciences Centre.

To sequence the sample, a 900bp whole genome library was constructed following the previously described protocol (2-3) with 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 the Illumina HiSeqX platform using V4 chemistry according to manufacturer recommendations.

The chloroplast reads were assembled by subsampling the reads to subsets of 0.75, 1.5, 3, 6, 12, 25, 46 million read pairs, then assembling each subset with ABySS v2.1.1 (4) (k=128, kc=3). The ABySS assembly of the 3M read pair subset resulted in a single 123,601-bp contig that aligned to the reference chloroplast sequence (*Picea glauca* admix genotype PG29, NCBI accession NC_028594), with zero misassemblies and no internal gaps, based on QUAST v5.0.0 (5) analysis.

For consistency with previously published conifer chloroplast genomes, we modified our assembly to match the strand and start position of the white spruce admix (PG29) chloroplast genome assembly using BLAST v2.7.1 (6). To ensure there were no missing sequences at the ends of our assembly we introduced a gap at the end, circularized the sequence and ran Sealer (7), closing the 'end' gap and removing overlapping sequences. Finally, the final assembly was polished using Pilon v1.22 (8).

The complete *P. engelmannii* Se404-851 chloroplast genome is 123,542-bp long, with 38.74% GC content. Using GeSeq (9) and other *Picea* chloroplast genomes as references (accessions below), we annotated 114 genes: 74 protein-coding, 36 tRNA-coding and 4 rRNA-coding. We note that 4 genes (*rps12*, *petB*, *petD*, and *rpl16*) were manually annotated. OGDRAW v1.2 (10) generated the map in Figure 1. The introduction of this new

chloroplast genome will benefit conifer genomic research and inform future evolutionary studies.

Accession number(s). The complete chloroplast genome sequence of *Picea engelmannii*, isolate Se404-851 is available under Genbank acession MK241981 and the raw reads in the SRA under SRX5070635. The annotation used as references were *Picea abies* (NC_021456), *Picea asperata* (NC_032367), *Picea glauca* isolate PG29 (NC_028594), *Picea morrisonicola* (NC_016069), and *Picea sitchensis* (NC_011152).
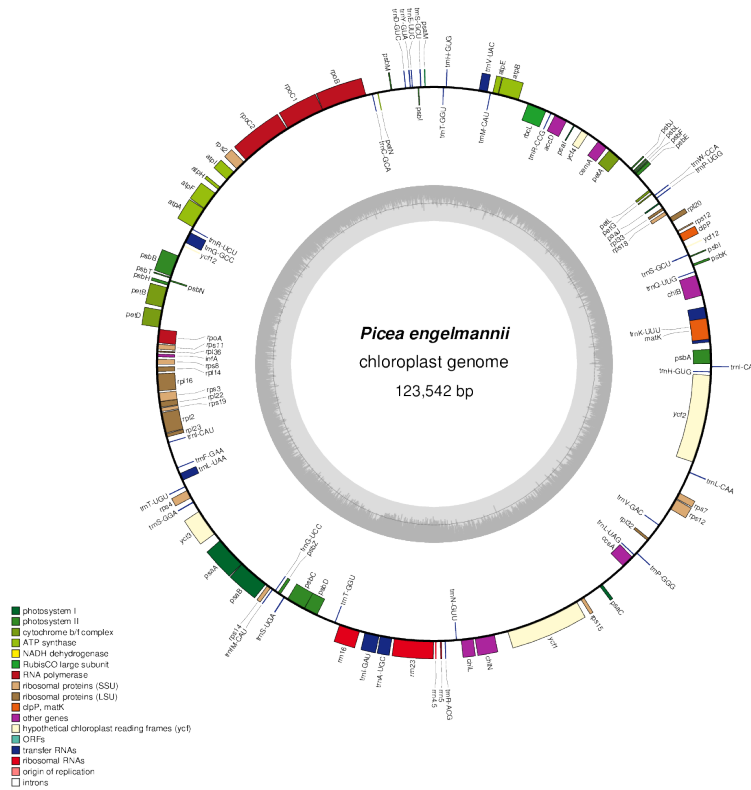
Figures and Data

Figure 1: The complete chloroplast genome of *Picea engelmannii*, isolate Se404-851. The *Picea engelmannii* chloroplast genome was annotated using GeSeq (9) and plotted using OGDRAW (10). The inner grey circle illustrates the GC content of the genome.

Acknowledgements

References

1. Vandersar, T.J.D. 1978. Resistance of western white pine to feeding and oviposition by *Pissodes strobi* peck in western Canada. J Chem Ecol 4:641.

2. 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.

3. 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.

4. 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.

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

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

7. 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.

8. 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.

9. 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.

10. 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.