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

TBD

February 1, 2019

Abstract

Picea engelmannii, is an Engelmann spruce tree grown in western Canada. Here we present the complete chloroplast genome sequence of the *Picea engelmannii*, isolate Se404-851. This sequence contributes data to the study of the evolutionary phylogeny of the *Picea* organisms, and consequently, facilitating the improvement of Canada's forestry industry.

Genome Announcement

As part of the SpruceUp (1) and SMarTForests (2) projects, we sequenced, assembled and annotated the chloroplast genome of *Picea engelmannii* isolate Se404-851. This work contributes to improving Canada's forestry industry through the the selection of spruce trees for breeding.

The needle tissue sample was collected from Kalamalka Forestry Centre in British Columbia (36°17'60"N, 105°24'0"W; elevation: 298 m). We then sequenced the sample at the British Columbia Genome Sciences Centre.

To sequence the sample, a modified version of TruSeq DNA PCR-Free kit (E6875-6877B-GSC, New England Biolabs) genome protocol was used to generate a 900-bp gap Illumina library on a Microlab NIMBUS liquid handling robot (Hamilton). Briefly, 5μg 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 concentrated with PCRClean DX magnetic beads (Aline Biosciences) and fractionated in 2 lanes of a 6% PAGE gel to recover fragments greater than 700-bp for library preparation. The isolated DNA fragments were end-repaired and bead-purified with a 1.8:1 ratio of beads, then A-tailed and ligated with full length indexed TruSeq adapters, and bead-purified. For quality check, an aliquot of the constructed library DNA was PCR amplified with Illumina universal primers to estimate the library gap size, using the Agilent 2100 Bioanalyzer HSDNA assay, while the library concentration was determined using KAPA qPCR Library Quantification kit (KK4824). The PCR-Free library was sequenced with paired-end 150 base reads on the Illumina HiSeqX platform using V4 chemistry according to manufacturer recommendations.

For assembly, we subsampled the reads into a few million read pairs. ABySS v2.1.1 (3) assembled each subset. Then, BWA v0.7.17 (4) filtered for contigs greater than 500-bp aligning with the reference chloroplast, *Picea glauca* isolate PG29. In the 3M assembly, a resultant contig (n=1; N_{50}=1743) had zero misassemblies and internal gaps, verified with QUAST v5.0.0 (5).

For consistency, we modified our assembly to match the reference using BLAST v2.7.1 (6). To ensure there were no missing sequences at the ends of our assembly we introduced a gap there, circularized the sequence and ran ABySS-sealer, closing the 'end' gap. Finally, Pilon v1.22 (7) polished the final assembly.

The Se404-851 chloroplast genome is 123,601-bp long, with 38.74% GC content. GeSeq (8) annotated 114 genes: 74 protein-coding, 36 tRNA-coding, 4 rRNA-coding genes, using other *Picea* chloroplast genomes as reference. *Rps12*, *petB*, *petD*, and *rpl16* were manu-

ally annotated. OGDRAW v1.2 (9) generated the genome map. Introduction of this new chloroplast genome, further analysis of phylogeny and evolution of these spruce trees can be conducted.

Accession number(s). The complete chloroplast genome sequence of *Picea engelmannii*, isolate Se404-851 can be found in Genbank under MK241981. The tissue samples used can be found under BioSample: SAMN10388286; BioProject: PRJNA504036. The annotation NCBI references are as follows: *Picea abies* (NC_021456), *Picea asperata* (NC_032367), *Picea glauca* isolate PG29 (NC_028594), *Picea morrisonicola* (NC_016069), and *Picea sitchensis* (NC_011152).

Acknowledgements

This work was conducted as a part of the SpruceUp and SMarTForests projects.

References

3. 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 research 27:768-777.

4. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics (Oxford, England) 25:1754-1760.

5. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics (Oxford, England) 29:1072-1075.

6. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. Journal of Molecular Biology 215:403-410.

7. 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.

8. 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 research 45:W6-W11.

9. 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 research 41:W575-W581.