

WHITE SPRUCE (*PICEA GLAUCA*) GENOME SEQUENCE ASSEMBLY

JOURNAL PUBLICATION and NCBI ACCESSION

The SMarTForests Project (<http://www.smartforests.ca/>) has now released a **third assembly of 20 Gb** based on the sequence data of the white spruce genome (*Picea glauca*) through a publicly accessible FTP site hosted at the British Columbia Cancer Agency, Canada's Michael Smith Genome Sciences Centre (ftp://ftp.bcgsc.ca/public/Picea_Glauca/). The first 20 Gb assembly can be found through NCBI ([BioProject PRJNA83435](http://www.ncbi.nlm.nih.gov/BioProject/PRJNA83435), Accession ALWZ0000000000) and the above FTP site. We are working with NCBI on the release of the updated assembly. The updated assembly reconstructs the genome in 4.2 million scaffolds, about 100 thousand of which capture half of the assembly, representing a scaffold N50 of over 34,000 bases. Subsequent analysis revealed that PG29 is not a pure white spruce (*Picea glauca*). The genome of PG29 has signatures of three different ancestors: *Picea glauca*, *P. sitchensis*, and *P. engelmannii*. Our analysis suggested the genome of PG29 to be 0.60 *P. glauca*, 0.31 *P. sitchensis*, and 0.09 *P. engelmannii*.

The SMarTForests project has also published in peer-reviewed form the details of the sequencing strategy, assembly details and an analysis of the first assembly; please see:

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, and SJM Jones (2013) **Assembling the 20 Gb white spruce (*Picea glauca*) genome from whole-genome shotgun sequencing data. *Bioinformatics*, published online May 22, 2013 as DOI:10.1093/bioinformatics/btt178**

Please cite the above mentioned paper and accession when using the published white spruce genome sequence data.

For inquiries about the data and for possible contributions to our ongoing analyses, please contact the Project Leaders: Dr. John MacKay (john.mackay@sbf.ulaval.ca) and Dr. Joerg Bohlmann (bohlmann@mssl.ubc.ca).

The current data release aims to support research and development in the scientific community according to the principles outlined in the Toronto workshop on pre-publication sharing of genomic data¹. We request anyone interested in using the *Picea glauca* genome sequence data for large-scale analyses to respect these principles.

The SMarTForests Project is funded by Genome Canada, Genome Québec, Genome British-Columbia, and Genome Alberta.

1. Toronto International Data Release Workshop Authors* (2009) Prepublication Data Sharing. *Nature*, 461: 168-170

*(author list: <http://www.nature.com/nature/journal/v461/n7261/extref/461168a-s1.pdf>).