

## Spruce Sequences and Assemblies Offer Hints of Genome Growth

May 23, 2013

By a [GenomeWeb staff reporter](#)

NEW YORK (GenomeWeb News) – The sequence and assembly of two spruce genomes, published in *Nature* and *Bioinformatics* yesterday, offer insight into gymnosperm biology and evolution. In particular, they offer clues as to how the expansive spruce genomes — coming in at about 20 gigabases in size — may have grown so large.

Researchers from Sweden and Canada characterized the Norway spruce and white spruce genomes — the first gymnosperms to be sequenced fully and the largest sequence assemblies produced thus far.

"The long reproductive cycles and large sizes of gymnosperms have made traditional, breeding-based analyses of these plants challenging," noted Ronald Sederoff from North Carolina State University in a [related News and Views article](#) in *Nature*.

"DNA-based technology that can bypass these limitations has been particularly useful in forest trees, enabling genomic mapping, gene sequencing, genomic selection, and genetic engineering.

The two genomes may also inform forestry and tree breeding efforts.

"These genome sequences allow us to develop innovative tools for tree breeding, addressing economically and ecologically important targets such as insect resistance, wood quality, growth rates, and adaptation to changing climate," the University of British Columbia's Joerg Bohlmann, a co-author on both studies, said in a statement.

The size of the spruce genomes also presented challenges to the researchers attempting to untangle their sequences.

The group led by Steven Jones at the Genome Sciences Centre of the British Columbia Cancer Agency, combined sequences generated by both the Illumina HiSeq 2000 and a modified MiSeq to piece together the white spruce genome, *Picea glauca*. Having both high-coverage and low-coverage data, the researchers noted, improved the assembly process and using a number of libraries and different fragment lengths would limit sampling biases and lead to an "even representation of the underlying genome."

Using the ABySS algorithm, they [assembled the white spruce genome](#) into 4.9 million scaffolds, as they reported online in *Bioinformatics*.

The Norway spruce genome, *Picea abies*, on the other hand, was assembled using a [combination of haploid and diploid whole-genome shotgun](#) and RNA sequencing data, as the other group led by Umeå University's Stefan Jansson reported in *Nature*.

The size of the Norway spruce genome, Jansson and his team said, appears to be due to its slow accumulation of long terminal repeat retrotransposon elements during the course of millions of years as well as of long intron stretches.

aA Type size: + -

Email

Printer-friendly version

RSS Feed

Jansson and his colleagues developed phylogenies for the reverse transcriptase genes of the Ty1/Copi and Ty3/Gypsy elements to track transposable elements through the history of vascular plants. From this, they found no recent — meaning within the last 5 million years — evidence of transposable element activity in *P. abies*. Further, *P. abies* and *P. glauca* share 63 insertions, and only five insertions appear to have arisen after their lineages split.

In addition, many of the transposable elements in the spruce genome appear to be removed from the genome by unequal recombination less frequently than they are in other plants. The researchers speculated that the mechanism that removes transposable elements in other organisms does not work as well in conifers.

"Taken together, these findings indicated that the extant set of transposable elements in *P. abies* accumulated slowly over tens or hundreds of millions years, mainly by the insertion of LTR-RT elements with limited transposable element removal," the investigators write.

Further, they discovered that the insertion of so many transposable elements has led to a high number of large introns and pseudogenes.

Each chromosome of the Norway spruce genome has grown to be about the same size, perhaps, they added, "limited by physical constraints on, for instance, chromosome replication."

"It is remarkable that the spruce is doing so well despite this unnecessary genetic load," said Pär Ingvarsson, a professor at the Umeå Plant Science Centre. "Of course, some of this DNA has a function, but it seems strange that it would be beneficial to have so very much."

Both spruce genome sequencing projects are associated with the SmartForest Project, which is developing marker systems for tree-breeding efforts.

"A genome-based marker system could serve to reduce the time of a spruce breeding cycle from currently 25 to as short as five years," added John MacKay, a professor at the Université Laval and a co-author of both studies.