



Conifer forests cover much of our planet's northern hemisphere. And yet we know rather little about their genetics — breeding is difficult owing to their long life cycles — or their genomes, which are some of the largest in the plant kingdom. Two reports now present the first drafts of gymnosperm genomes.

Although they are diploid, genome-wide analyses in conifers are a challenge: their genomes are very large and highly heterozygous, which makes genome assembly difficult. Nystedt *et al.* used a combination of a fosmid-pooling approach (which was developed to sequence the oyster genome), whole-genome shotgun sequencing and RNA sequencing (RNA-seq) to assemble a draft of the 20 Gb genome of the Norway spruce. In addition, Birol *et al.* have developed a new approach that combines recent improvements in sequence read length with a novel bioinformatics tool to assemble the similarly large genome of the white spruce.

To gain insights into plant genome evolution, Nystedt *et al.* also generated low-coverage draft genome assemblies of five other gymnosperms. Comparative genomics using these and other previously sequenced plant genomes revealed why the Norway spruce genome is so large. Although there was no evidence for a recent whole-genome duplication, the authors found a profusion of long terminal repeat transposable elements, which appear to have been accumulating over several tens of millions of years. This accumulation also probably accounts for the long introns and the large number of pseudogenes. The authors propose that mechanisms for transposable element removal have been less active in the Norway spruce than in other organisms. It

also appears that a class of small RNAs associated with methylation of repeats, which is known to limit transposition, is less abundant in the Norway spruce than in other genomes.

The genome also yielded some important insights into gymnosperm biology. Two important differences between angiosperms and gymnosperms — two major plant groups — are their contrasting reproductive development and the development of water-conducting xylem cells. It has previously been suggested, and is confirmed by this genome sequence, that gymnosperms lack *FLOWERING LOCUS T (FT)* genes, which encode key activators of flowering; instead, they contain genes that are predicted to suppress flowering. Furthermore, the Norway spruce contains only two members of the gene family that controls formation of water-conducting vessels (namely, *VASCULAR NAC DOMAIN*), whereas *Arabidopsis thaliana* has seven. The authors speculate that the expansion of this gene family might have been important for angiosperm evolution.

Now that the first gymnosperm genome has been cracked, other large genomes from this important group of plants will follow suit. The comparative analyses presented in this paper will improve understanding of plant biology and plant genome evolution. For conifers specifically, the genomes will be an invaluable resource for the breeders.

Magdalena Skipper, Senior Editor, Nature

ORIGINAL RESEARCH PAPERS Nystedt, B. *et al.* The Norway spruce genome sequence and conifer genome evolution. *Nature* 22 May 2013 (doi:10.1038/nature12211) | Birol, I. *et al.* Assembling the 20 Gb white spruce (*Picea glauca*) genome from whole-genome shotgun sequencing data. *Bioinformatics* 22 May 2013 (doi:10.1093/bioinformatics/btt178)