

Organellar Genomes of White Spruce (*Picea glauca*): Assembly and Annotation

Shaun D Jackman <sjackman@bcgsc.ca>
Genome Sciences Centre, British Columbia Cancer Agency
Vancouver, Canada

The genome sequences of the plastid and mitochondrion of white spruce (*Picea glauca*) were assembled from whole-genome shotgun sequencing data using ABySS. The sequencing data contained reads from both the nuclear and organellar genomes, and reads of the organellar genomes were abundant in the data as each cell harbors hundreds of mitochondria and plastids. Hence, assembly of the 123-kb plastid and 5.9-Mb mitochondrial genomes were accomplished by analyzing data sets primarily representing low coverage of the nuclear genome. The assembled organellar genomes were annotated for their coding genes, ribosomal RNA, and transfer RNA. Transcript abundances of the mitochondrial genes were quantified in three developmental tissues and five mature tissues using data from RNA-seq experiments. C-to-U RNA editing was observed in the majority of mitochondrial genes, and in four genes, editing events were noted to modify ACG codons to create cryptic AUG start codons. The informatics methodology presented in this study should prove useful to assemble organellar genomes of other plant species using whole-genome shotgun sequencing data.

Chloroplast genomes of gymnosperms, including conifers, are well studied, but little is known about the mitochondria of gymnosperms. In fact, only a single gymnosperm mitochondrion is found in NCBI GenBank. This nearest related mitochondrial sequence is of the Prince Sago palm (*Cycas taitungensis*) native to Taiwan, which diverged from the white spruce over a hundred million years ago. No conifer mitochondrion genomes are to be found in GenBank at all, until now.

Roughly one percent of the whole genome sequencing reads of white spruce are from its two organellar genomes: the chloroplast and mitochondrion. We assembled these reads using ABySS and found the mitochondrion genome to be nearly six megabases, which is unusually large for a mitochondrial genome. Although many genes typical of mitochondria were found in the genome, most open reading frames had no similarity to any known gene.

White spruce is an economically important species to the forestry industry of Canada. Insights into the conifer mitochondrial genome will provide relevant new information to reconstruct the evolution of this organelle genome relative to other plant lineages, and to identify which genes of a conifer are uniquely inherited through the mitochondria. As the mitochondrial genome is inherited maternally, and the plastid genome is inherited paternally, having a complete genome sequence for both organelles would enable classifying the maternal and paternal species of hybrid seed lots and determining the maternal and paternal lineage of saplings in breeding experiments.