

Complete mitochondrial genome of a gymnosperm, Sitka spruce (*Picea sitchensis*), indicates a complex physical structure

Shaun D. Jackman^{1,*}

Lauren Coombe^{1,*}

René L. Warren¹

Heather Kirk¹

Eva Trinh¹

Tina MacLeod¹

Stephen Pleasance¹

Pawan Pandoh¹

Yongjun Zhao¹

Robin J. Coope¹

Jean Bousquet²

Joerg Bohlmann³

Steven J. M. Jones¹

Inanc Birol¹

¹ Genome Sciences Centre, BC Cancer, Vancouver, BC V5Z 4S6, Canada

² Canada Research Chair in Forest Genomics, Institute for Systems and Integrative Biology, Université Laval, Quebec, Quebec G1V 0A6, Canada

³ University of British Columbia, Michael Smith Laboratories, Vancouver, BC V6T 1Z4, Canada

Author for Correspondence: Inanc Birol, Genome Sciences Centre, BC Cancer, Vancouver, BC V5Z 4S6, Canada. tel: 604-707-5900 x 675448 :: fax: 604-876-3561 *Authors contributed equally

Data deposition: The Oxford Nanopore MinION sequencing data is available at the SRA under accessions SRX5081713, and SRR8264752 through SRR8264769. The Illumina HiSeq sequencing data has SRA accession SRR5028199. The Sitka spruce (*Picea sitchensis*) genotype Q903 complete mitochondrial genome is available at NCBI GenBank under accessions MK697696 through MK697708.

Abstract

Plant mitochondrial genomes vary widely in size. Although many plant mitochondrial genomes have been sequenced and assembled, the vast majority are of angiosperms, and few are of gymnosperms. Most plant mitochondrial genomes are smaller than a megabase, with a few notable exceptions. We have sequenced and assembled the complete 5.5 Mbp mitochondrial genome of Sitka spruce (*Picea sitchensis*), to date, one of the largest mitochondrial genomes of a gymnosperm. We sequenced the whole genome using Oxford Nanopore MinION, and then identified contigs of mitochondrial origin assembled from these long reads based on sequence homology to the white spruce mitochondrial genome. The assembly graph shows a multipartite genome structure, composed of one smaller 168 kbp circular segment of DNA, and a larger 5.4 Mbp single component with a branching structure. The assembly graph gives insight into a putative complex physical genome structure, and its branching points may represent active sites of recombination.

Key words: gymnosperms, Sitka spruce, organelle, genome assembly, sequencing, ABySS.

Introduction

Plant mitochondrial genomes are amazingly diverse and complex (Mower et al. 2012). Land plant mitochondrial genomes range in size from 66 kilobases (kbp) for the parasitic angiosperm *Viscum scurruloideum* (Skippington et al. 2015) to more than 11 megabases (Mbp) in the case of the flowering plant *Silene conica* (Sloan et al. 2012). Although their genome structure is often portrayed as a circle, the true physical structure of their genome

appears to be a variety of circles, linear molecules, and complex branching structures (Backert et al. 1997; Backert & Börner 2000). While many species have a single master circle representation of their mitochondrial genome, others are composed of more than a hundred circular chromosomes (Sloan et al. 2012). The precise mechanism of how plant mitochondria replicate and maintain their DNA is not yet fully understood (Cupp & Nielsen 2014). It is hypothesized that recombination-dependent replication plays a role, giving a functional role to the repeat sequences often observed in mitochondria (Gualberto et al. 2014). This model does not fully explain how genomic copy number is regulated and maintained (Oldenburg & Bendich 2015), particularly in multipartite genomes (Vlcek et al. 2010). Although angiosperm mitochondrial genomes are well studied with numerous complete genomes available, gymnosperm mitochondrial genomes were scarcer until recently: one from each of the cycads (Chaw et al. 2008), ginkgos, gnetales (Guo et al. 2016), and conifers (Jackman et al. 2015). This year saw a number of mitochondrial genomes published in short succession (Sullivan et al. 2020; Kan et al. 2020; Guo et al. 2020). Whereas other gymnosperm mitochondrial genomes are smaller than a megabase, conifer mitochondrial genomes can exceed five megabases (Jackman et al. 2015), larger than many bacteria. Compared to other plants, conifers also tend to have very large nuclear genomes (De La Torre et al. 2014), particularly the spruce species (Birol et al. 2013; Nystedt et al. 2013; Warren et al. 2015). As plant mitochondrial genomes typically have fewer than 100 genes, what role this expanse of DNA serves, if any, remains mysterious.

Assembling plant mitochondrial genomes is difficult due to the presence of large (up to 30 kbp) perfect repeats, which may be involved in active recombination, and hypothesized recombination-dependent replication (Gualberto et al. 2014). A hybrid

assembly of both long reads, which are able to span most repeats, and accurate short sequencing reads, which correct indel errors, is well suited to tackle these challenging genome features. Hybrid assembly of long and short reads has been applied to assemble the plastid genome of *Eucalyptus pauciflora* (Wang et al. 2018), as well as plant mitochondrial genomes (Kover et al. 2018; Kozik et al. 2019).

Annotating plant mitochondrial genomes is also challenging, due to numerous features of plant mitochondria that are not typical of most organisms. For one, RNA editing of C-to-U is pervasive, and this process creates AUG start codons by editing ACG to AUG (Hiesel et al. 1989) or by editing GCG to GUG, an alternative start codon used by some plant mitochondrial genes (Sakamoto et al. 1997). RNA editing can also create stop codons in a similar fashion. Further complicating annotation using available bioinformatics pipelines, the typical GU-AG splice site expected by most splice-aware alignments tools is instead GNGCG-AY (Y denotes C or T) for group II introns (Lambowitz & Zimmerly 2010). Also, trans-spliced genes are common in mitochondrial genomes (Kamikawa et al. 2016; Guo et al. 2020), and no purpose-built software tool exists for identifying and annotating trans-spliced genes. To add further difficulty, trans-spliced exons may be as small as 22 bp, as is *nad5* exon 3 of gymnosperms (Guo et al. 2016) and other vascular plant mitochondria (Knoop et al. 1991). For these reasons, annotating a plant mitochondrial genome remains a laborious and manual task.

In this study, we report on the sequencing, assembly and annotation of the mitochondrial genome of Sitka spruce (*Picea sitchensis*, Pinaceae), a widely distributed conifer in the coastal regions of the Pacific Northwest of North America. We show that this

mitochondrial genome is one of the largest among plants and exhibits a multipartite genome structure.

Materials and Methods

Genome Sequencing and Assembly

Genomic DNA was extracted from young Sitka spruce (*Picea sitchensis* [Bong.] Carrière, genotype Q903) needles, as described in Coombe et al. (2016). We constructed 18 Oxford Nanopore 1D sequencing libraries, 16 by ligation of 1 to 7 µg of lightly needle-sheared genomic DNA and 2 by rapid transposition of 0.6 µg of unsheared genomic DNA, and sequenced these on 18 MinION R9.4 flow cells. This whole genome sequencing produced 98 Gbp in 9.6 million reads (SRA accession SRX5081713), yielding 5-fold depth of coverage of the roughly 20 Gbp nuclear genome, and 26-fold depth of coverage of the mitochondrial genome. We first obtained a rough but computationally efficient assembly using Miniasm (Li 2016), after trimming adapter sequences with Porechop (Wick et al. 2017a) and polished the resulting assembly with Racon (Vaser et al. 2017). We selected contigs with homology to the white spruce (*Picea glauca*, interior white spruce genotype PG29) mitochondrial genome (Jackman et al. 2015) using Bandage (Wick et al. 2015), retaining contigs with at least one ≥ 5 kbp alignment to the white spruce mitochondrion by BLASTN (Altschul et al. 1990). We then proceeded to align back the nanopore reads to our draft Sitka spruce assembly with minimap2 (Li 2016), segregate aligned reads and assemble them *de novo* first with Unicycler (Wick et al. 2017b), and then with Flye (Kolmogorov et al. 2018), as described (Supplementary methods). The final genome sequence was obtained

by polishing with Pilon (Walker et al. 2014) using Illumina HiSeq sequencing reads of the same DNA extraction (Supplementary Figure S1).

Annotation

We annotated coding genes and non-coding rRNA and tRNA genes using automated methods where possible, and performed manual inspection to refine these automated annotations. We used Prokka (Seemann 2014), which uses Prodigal (Hyatt et al. 2010) to identify single-exon coding genes and open reading frames (ORFs). We used MAKER (Holt & Yandell 2011), which uses BLASTP and Exonerate (Slater & Birney 2005) to identify cis-spliced coding genes. Prokka used *ab initio* predictions from Prodigal. MAKER used evidence alignments only. Genes in GenBank from *Viridiplantae* mitochondria were used as evidence. The annotations of Prokka and Maker were combined, along with the manually annotated *cis*-spliced and trans-spliced genes. No species-specific TE libraries were used.

We used tRNAscan-SE (Lowe & Eddy 1997) and Aragorn (Laslett 2004) to identify tRNA. We used RNAmmer (Lagesen et al. 2007) and Barrnap (Seemann 2014) to identify rRNA. We used RNAweasel (Lang et al. 2007) and Infernal (Nawrocki et al. 2009) to identify group II introns. RFAM motif Domain-V (RM00007) represents domain V of group II introns, and RFAM family Intron_gpII (RF00029) represents both domains V and VI (Kalvari et al. 2017).

Results and Discussion

Complete Genome Assembly

The complete mitochondrial genome of Sitka spruce is 5.52 Mbp assembled in 13 segments, and has a GC content of 44.7%. The assembly statistics are summarized in Table 1. The genome assembly is composed of two components: a 168 kbp circular segment (labeled 10), and a larger 5.36 Mbp component composed of 12 segments, as visualized in Figure 1 using Bandage (Wick et al. 2015). The eleven largest segments, ranging in size from 84 kbp to 1.65 Mbp, have similar depth of coverage, assumed to represent single-copy genomic segments. The two smallest segments (27 kbp and 24 kbp labeled 12 and 13 in Figure 1, respectively, representing less than 1% of the mitochondrial genome) exhibit an estimated copy number of two based on their depth of sequencing coverage. No sequence variation is evident in these two repeats. An absence of variation in the repeat implies that they may be involved in active recombination (Maréchal & Brisson 2010). Though 10% of reads are larger than 24 kbp, no reads fully span these repeats.

Table 1. Summary of assembly statistics.

Assembly statistics	
Number of contigs	13
Assembly N50	547 kbp
Undetermined bases	None
Largest contig	1.65 Mbp
Total reconstruction	5.53 Mbp
GC content	44.7%

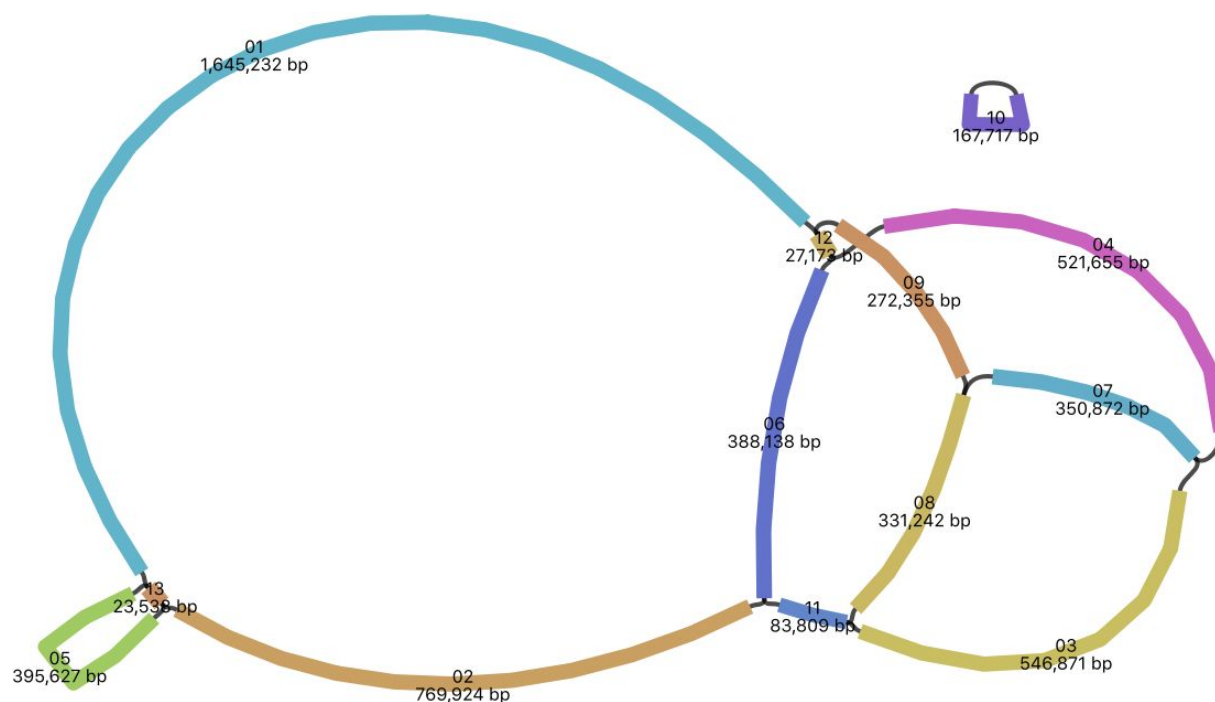


Fig. 1. The assembly graph of the mitochondrial genome of Sitka spruce. Each colored segment is labeled with its size and named 01 through 13 by rank of size. Only segment 12 and 13 representations are inferred as repeats. All segment adjacencies are supported by the long reads, indicating a complex branching genomic structure.

The complete mitochondrial genome assembly of Sitka spruce assembled from Oxford Nanopore sequencing is composed of 13 contigs larger than 20 kbp with an N50 length of 547 kbp. The use of long reads was critical in achieving this contiguity and completeness. In contrast, the draft mitochondrial genome assembly of white spruce was assembled from paired-end and mate-pair Illumina sequencing and is composed of 117 contigs larger than 2 kbp arranged in 36 scaffolds with a contig and scaffold N50 length of 102 and 369 kbp, respectively (Jackman et al. 2015). The fragmented state of the white spruce mitochondrial assembly provided little information as to the structure of the mitochondrial genome, while the Sitka spruce assembly graph (Figure 1) suggests a multipartite genome structure.

The complete genome is composed of 1.7% (93 kbp) of genes with known function, 28.0% (1,545 kbp) of 6,806 ORFs (each of at least 90 bp), 3.7% (205 kbp) of repeats, and 66.6% unclassified sequences. Of the ORFs, 1,039 are at least 300 bp (100 amino acids) in size and compose 7.2% (400 kbp) of the genome. Aligning the ORFs with BLASTP, 63 ORFs (17 ORFs of at least 300 bp) have a significant ($E < 0.001$) hit to the nr database. Plastid-derived sequences compose 0.25% (14 kbp) of the genome spread across 24 segments.

The nuclear repeats LTR/Gypsy compose 51% of the repeat sequence, and LTR/Copia compose 7%. The genome also has non-transposable element repeats; simple repeat sequences compose 34%, low complexity sequences compose 3%, and 5% are other repeat sequences. The 36-bp Bpu repeat sequence is present in roughly 500 copies in *Cycas taitungensis* and roughly 100 copies in *Ginkgo biloba* (Guo et al. 2016). We find only a single full-length copy with four mismatches in Sitka spruce, similar to *Welwitschia mirabilis*.

Genes

The mitochondrial genome of Sitka spruce has 41 distinct protein coding genes with known function, 3 distinct rRNA genes (Supplementary Table 1), and 27 distinct tRNA genes representing 18 distinct anticodons (Supplementary Table 2). The relative order and orientation of these genes are shown in Figure 2. The 41 known protein coding genes found in the gymnosperm mitochondria *Cycas taitungensis* (Chaw et al. 2008) and *Ginkgo biloba* (Guo et al. 2016) are also found in Sitka spruce. The 29 introns, 16 cis-spliced and 13 trans-spliced, are found in 10 protein coding genes, two pseudogenes, and one plastid-derived tRNA (Supplementary Table 3).

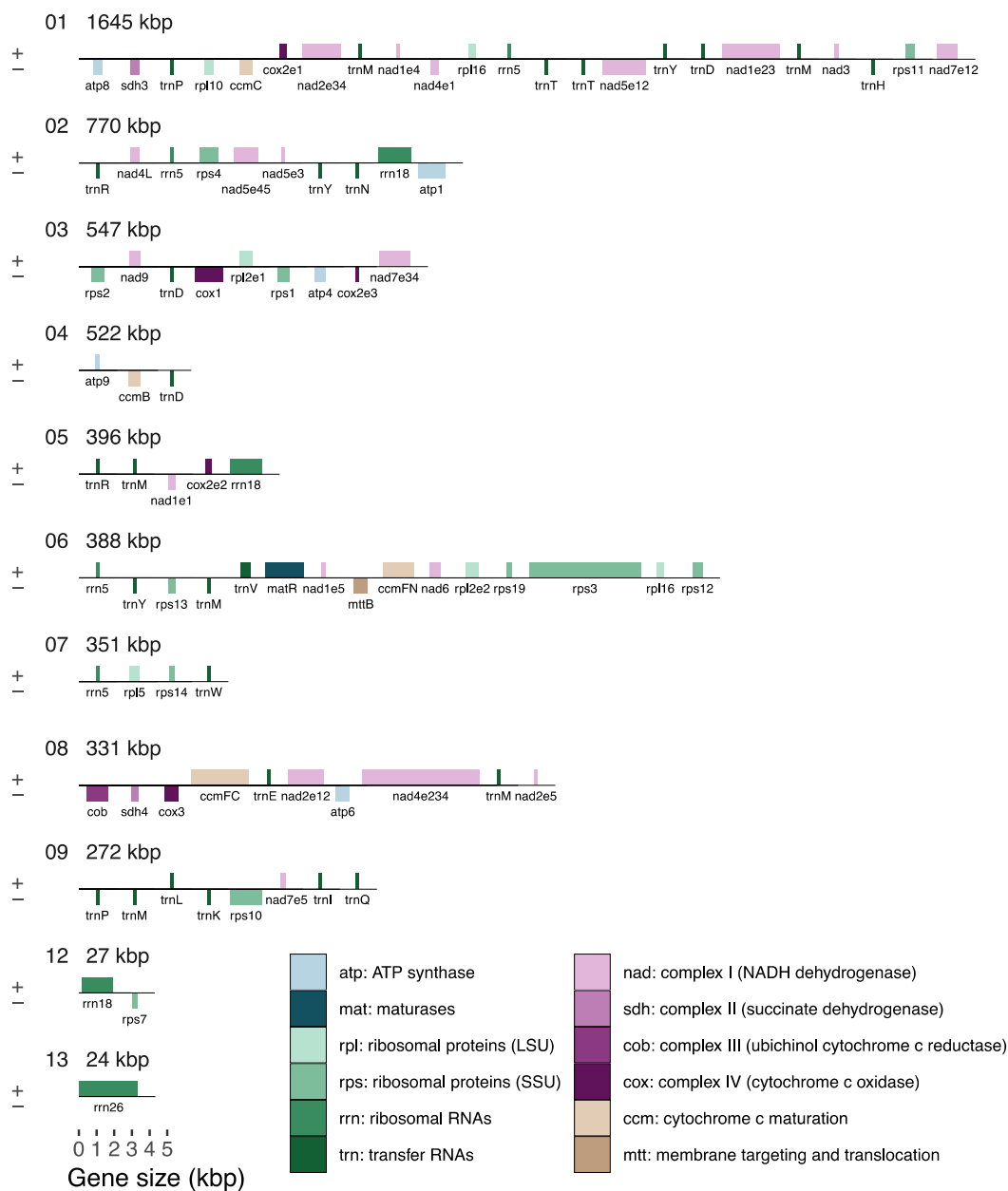


Fig. 2. The order, orientation, and size of the genes within the Sitka spruce mitochondrial genome. Each box is proportional to the size of the gene including introns, except that genes smaller than 200 nucleotides are shown as 200 nucleotides. Intergenic regions are not to scale. Contigs 10 and 11 did not contain any annotated genes and are not shown.

Four ORFs (937 [Q903MT_gene558], 507 [Q903MT_gene3172], 371 [Q903MT_gene4186], and 221 [Q903MT_gene4187] amino acids) contain an organellar DNA polymerase type B (DNA_pol_B_2) Pfam family (El-Gebali et al. 2018). The largest one also contains a segment homologous to a putative cell-wall binding lipoprotein (YkyA) Pfam domain. These ORFs have homology to DNA polymerase genes found in the mitochondria of *Picea glauca* and the angiosperms *Cocos nucifera*, *Daucus carota*, *Helianthus annuus*, and *Silene vulgaris*. Two ORFs (781 [Q903MT_gene4185] and 560 [Q903MT_gene6929] amino acids) contain an RNA polymerase (RNA_pol) domain. These ORFs have homology to DNA-dependent RNA polymerase genes found in the mitochondria of *Picea glauca* and the angiosperms *Beta vulgaris*, *Cocos nucifera*, *Daucus carota*, and *Phoenix dactylifera*. The two largest genes of the Sitka spruce mitochondrial genome are these putative DNA and RNA polymerase genes.

The *matR* gene [Q903MT_gene5018] and three additional ORFs (476 [Q903MT_gene679], 197 [Q903MT_gene3613], and 163 [Q903MT_gene5889] amino acids) contain a reverse transcriptase, or RNA-dependent DNA polymerase, (RT_like) NCBI conserved protein domain with similarity to the group II intron reverse transcriptase/maturase (group_II_RT_mat) NCBI conserved protein domain (Marchler-Bauer et al. 2016). We hypothesize that these ORFs may be additional maturases involved in splicing (Matsuura 2001). These ORFs have homology to mitochondrial genes of *Picea glauca* and the angiosperm *Utricularia reniformis*.

Conclusion

The 5.5 Mbp mitochondrial genome of Sitka spruce is among the largest ones in plants, and is the largest complete mitochondrial genome reported to date for a gymnosperm. It follows the trend seen for spruce and conifer nuclear genomes, which are also among the largest in plants (De La Torre et al. 2014). The physical structure of the Sitka spruce mitochondrial genome is not the typical circularly-mapping single chromosome, but multipartite. The larger component of the assembly graph exhibits a rosette-like structure, mirroring the rosette-like structures observed in electron micrographs of mitochondrial DNA (Backert & Börner 2000). The intricate structure and the discovery of large repeat elements suggest the presence of active sites for hypothesized recombination-dependent replication of the mitochondrial genome (Gualberto et al. 2014; Sullivan et al. 2020). Considering heteroplasmy resulting from naturally hybridizing species of spruce and paternal leakage of mitochondria, intermolecular recombination would result in interspecific hybridization of the mitochondrial genome, which has been reported to occur in natural spruce populations (Jaramillo-Correa & Bousquet 2005). The complete mitochondrial genome of Sitka spruce, a gymnosperm, should prove invaluable to future investigations into the genome structure and mechanism of replication of conifer mitochondrial genomes.

Funding

This work was supported by Genome Canada; Genome Quebec; Genome British Columbia; and Genome Alberta [grant number 243FOR] for the Spruce-Up project (<http://www.spruce-up.ca>).

Competing Interests

The authors declare no competing interests.

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology*. 215:403–410.
- Backert S, Börner T. 2000. Phage T4-like intermediates of DNA replication and recombination in the mitochondria of the higher plant *Chenopodium album* (L.). *Current Genetics*. 37:304–314.
- Backert S, Lynn Nielsen B, Börner T. 1997. The mystery of the rings: structure and replication of mitochondrial genomes from higher plants. *Trends in Plant Science*. 2:477–483.
- Birol I et al. 2013. Assembling the 20 Gb white spruce (*Picea glauca*) genome from whole-genome shotgun sequencing data. *Bioinformatics*. 29:1492–1497.

- Chaw S-M et al. 2008. The Mitochondrial Genome of the Gymnosperm *Cycas taitungensis* Contains a Novel Family of Short Interspersed Elements, Bpu Sequences, and Abundant RNA Editing Sites. *Molecular Biology and Evolution*. 25:603–615.
- Coombe L et al. 2016. Assembly of the Complete Sitka Spruce Chloroplast Genome Using 10X Genomics' GemCode Sequencing Data Budak, H, editor. *PLOS ONE*. 11:e0163059.
- Crooks GE. 2004. WebLogo: A Sequence Logo Generator. *Genome Research*. 14:1188–1190.
- Cupp JD, Nielsen BL. 2014. Minireview: DNA replication in plant mitochondria. *Mitochondrion*. 19:231–237.
- De La Torre AR et al. 2014. Insights into Conifer Giga-Genomes. *PLANT PHYSIOLOGY*. 166:1724–1732.
- El-Gebali S et al. 2018. The Pfam protein families database in 2019. *Nucleic Acids Research*. 47:D427–D432.
- Goruyunov DV et al. 2018. Complete mitochondrial genome sequence of the 'copper moss' *Mielichhoferia elongata* reveals independent nad7 gene functionality loss. *PeerJ*. 6:e4350.
- Gualberto JM et al. 2014. The plant mitochondrial genome: Dynamics and maintenance. *Biochimie*. 100:107–120.
- Guo W et al. 2016. Ginkgo and Welwitschia Mitogenomes Reveal Extreme Contrasts in Gymnosperm Mitochondrial Evolution. *Molecular Biology and Evolution*. 33:1448–1460.

Guo W et al. 2020. Extensive shifts from cis to trans splicing of gymnosperm mitochondrial introns. *Molecular Biology and Evolution*. msaa029, <https://doi.org/10.1093/molbev/msaa029>

Hiesel R, Wissinger B, Schuster W, Brennicke A. 1989. RNA editing in plant mitochondria. *Science*. 246:1632–1634.

Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics*. 12.

Hyatt D et al. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*. 11.

Jackman SD et al. 2015. Organellar Genomes of White Spruce (*Picea glauca*): Assembly and Annotation. *Genome Biology and Evolution*. 8:29–41.

Jaramillo-Correa JP, Bousquet J. 2005. Mitochondrial Genome Recombination in the Zone of Contact Between Two Hybridizing Conifers. *Genetics*. 171:1951–1962.

Kalvari I et al. 2017. Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Research*. 46:D335–D342.

Kamikawa R, Shiratori T, Ishida K-I, Miyashita H, Roger AJ. 2016. Group II Intron-Mediated Trans-Splicing in the Gene-Rich Mitochondrial Genome of an Enigmatic Eukaryote, *Diphylleia rotans*. *Genome Biology and Evolution*. 8:458–466.

Kan SL et al. 2020. The complete mitochondrial genome of *Taxus cuspidata* (Taxaceae): eight protein-coding genes have transferred to the nuclear genome. BMC Evol Biol. 20(1):10.

Knoop V, Schuster W, Wissinger B, Brennicke A. 1991. Trans splicing integrates an exon of 22 nucleotides into the nad5 mRNA in higher plant mitochondria. The EMBO Journal. 10:3483–3493.

Kolmogorov M, Yuan J, Lin Y, Pevzner P. 2019. Assembly of Long Error-Prone Reads Using Repeat Graphs. Nat Biotechnol. 37:540-546.

Kovar L et al. 2018. PacBio-Based Mitochondrial Genome Assembly of *Leucaena trichandra* (Leguminosae) and an Intrageneric Assessment of Mitochondrial RNA Editing. Genome Biol Evol. 10(9):2501-2517.

Kozik A et al. 2019. The alternative reality of plant mitochondrial DNA: One ring does not rule them all. PLoS Genet. 15(8):e1008373.

Lagesen K et al. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Research. 35:3100–3108.

Lambowitz AM, Zimmerly S. 2010. Group II Introns: Mobile Ribozymes that Invade DNA. Cold Spring Harbor Perspectives in Biology. 3:a003616–a003616.

Lang BF, Laforest M-J, Burger G. 2007. Mitochondrial introns: a critical view. Trends in Genetics. 23:119–125.

- Laslett D. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Research*. 32:11–16.
- Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 34:3094–3100.
- Li H. 2016. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics*. 32:2103–2110.
- Lowe TM, Eddy SR. 1997. tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. *Nucleic Acids Research*. 25:0955–964.
- Marchler-Bauer A et al. 2016. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Research*. 45:D200–D203.
- Maréchal A, Brisson N. 2010. Recombination and the maintenance of plant organelle genome stability. *New Phytologist*. 186:299–317.
- Matsuura M. 2001. Mechanism of maturase-promoted group II intron splicing. *The EMBO Journal*. 20:7259–7270.
- Mower JP, Sloan DB, Alverson AJ. 2012. Plant Mitochondrial Genome Diversity: The Genomics Revolution. *Plant Genome Diversity Volume 1*. 123–144.
- Nawrocki EP, Kolbe DL, Eddy SR. 2009. Infernal 1.0: inference of RNA alignments. *Bioinformatics*. 25:1335–1337.
- Nystedt B et al. 2013. The Norway spruce genome sequence and conifer genome evolution. *Nature*. 497:579–584.

- Oldenburg DJ, Bendich AJ. 2015. DNA maintenance in plastids and mitochondria of plants. *Frontiers in Plant Science*. 6.
- Sakamoto W, Tan S-H, Murata M, Motoyoshi F. 1997. An Unusual Mitochondrial atp9-rpl16 Cotranscript Found in the Maternal Distorted Leaf Mutant of *Arabidopsis thaliana*: Implication of GUG as an Initiation Codon in Plant Mitochondria. *Plant and Cell Physiology*. 38:975–979.
- Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 30:2068–2069.
- Skippington E1, Barkman TJ2, Rice DW1, Palmer JD. 2015. Miniaturized mitogenome of the parasitic plant *Viscum scurruloideum* is extremely divergent and dynamic and has lost all nad genes. *Proc Natl Acad Sci U S A*. 112(27):E3515-24.
- Slater G, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*. 6:31.
- Sloan DB et al. 2012. Rapid Evolution of Enormous, Multichromosomal Genomes in Flowering Plant Mitochondria with Exceptionally High Mutation Rates Gray, MW, editor. *PLoS Biology*. 10:e1001241.
- Sullivan AR et al. 2020. The Mitogenome of Norway Spruce and a Reappraisal of Mitochondrial Recombination in Plants. *Genome Biol Evol*. 12(1):3586-3598.
- Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Research*. 27:737–746.

Vlcek C, Marande W, Teijeiro S, Lukeš J, Burger G. 2010. Systematically fragmented genes in a multipartite mitochondrial genome. *Nucleic Acids Research*. 39:979–988.

Walker BJ et al. 2014. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement Wang, J, editor. *PLoS ONE*. 9:e112963.

Wang W et al. 2018. Assembly of chloroplast genomes with long- and short-read data: a comparison of approaches using *Eucalyptus pauciflora* as a test case. *BMC Genomics*. 19.

Warren RL et al. 2015. Improved white spruce (*Picea glauca*) genome assemblies and annotation of large gene families of conifer terpenoid and phenolic defense metabolism. *The Plant Journal*. 83:189–212.

Wick RR, Judd LM, Gorrie CL, Holt KE. 2017a. Completing bacterial genome assemblies with multiplex MinION sequencing. *Microbial Genomics*. 3.

Wick RR, Judd LM, Gorrie CL, Holt KE. 2017b. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads Phillippy, AM, editor. *PLOS Computational Biology*. 13:e1005595.

Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics*. 31:3350–3352.