

Tight Association of Genome Rearrangements With Gene Expression in Conifer Plastomes

Chung-Shien Wu

Academia Sinica <https://orcid.org/0000-0003-0000-6253>

Edi Sudianto

Academia Sinica

Shu-Miaw Chaw (✉ smchaw@sinica.edu.tw)

Academia Sinica <https://orcid.org/0000-0003-2499-7071>

Research article

Keywords: Conifer, plastid transcriptome, plastomic rearrangement, strand-specific RNAseq, RNA-editing

Posted Date: November 10th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-101511/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published on January 8th, 2021. See the published version at <https://doi.org/10.1186/s12870-020-02809-2>.

Abstract

Background: Our understanding of plastid transcriptomes is limited to a few model plants whose plastid genomes (plastomes) have a highly conserved gene order. Consequently, little is known about how gene expression changes in response to genomic rearrangements in plastids. This is particularly important in the highly rearranged conifer plastomes.

Results: We sequenced and reported the plastomes and plastid transcriptomes of six conifer species, representing all six extant families. Strand-specific RNAseq data shows a nearly full transcription of both plastomic strands and detects C-to-U RNA-editing sites at both sense and antisense transcripts. We demonstrate that the expression of plastid coding genes is strongly functionally dependent among conifer species. However, the strength of this association declines as the number of plastomic rearrangements increases. This finding indicates that plastomic rearrangement influences gene expression.

Conclusions: Our data provide the first line of evidence that plastomic rearrangements not only complicate the plastomic architecture but also drive the dynamics of plastid transcriptomes in conifers.

Background

Conifers are a group of cone-bearing seed plants. They comprise ca. 630 species in two clades, Pinaceae (conifers I clade) and cupressophytes (conifers II clade, consisting of five families). Conifers dominate temperate forests, especially in the Northern hemisphere, and significantly contribute to photosynthesis and biomass production. They provide shelters for wildlife and important resources for humans, such as solid wood fuel, valuable timber, edible seeds, and essential oils [1].

Plastid gene transcription is a complex process, involving both prokaryotic- and eukaryotic-type systems [2]. Most plastid genes are presumably transcribed as polycistronic mRNAs which then undergo various post-transcriptional modifications [3]. These processes generate tremendously elaborate transcriptomes with an unprecedented diversity of non-coding RNAs [4], multiple loci for transcriptional initiation and termination [5, 6], a full or nearly full transcription of the genome [7, 8], and varying frequencies of RNA-editing sites [9].

Plastid genomes (plastomes) of land plants are highly conserved in their gene content and order. Functionally related genes are commonly found in clusters and are likely co-transcribed as operons [10]. These operons may be conserved due to selective constraints rather than slow rates of neutral chromosomal rearrangements [11]. However, mounting evidence indicates that many taxa, including conifers (the largest group of gymnosperms), have highly rearranged plastomes [12–14]. Some of these rearrangements resulted in disruption of canonical operons and creation of novel co-transcriptional units. An example is the disruption of *rps2* operons in *Sciadopitys* and *Callitris* [15, 16]. We have long been puzzled by these findings because it is then unclear whether plastomic rearrangements affect plastid gene transcription. If they do, what are the underlying mechanisms and consequences of such changes?

In this study, we sequenced both plastomic DNA and RNA from one representative genus in each of the six extant conifer families. Strand-specific RNA libraries have the advantage of allowing for the discrimination of sense and antisense transcripts [17]. We took advantage of this to (1) investigate the full transcription capability of both plastomic strands, (2) estimate the relative number of plastid coding and antisense transcripts, and (3) identify plastid C-to-U RNA-editing sites separately at sense and antisense transcripts in conifers. We also compared plastid gene expression levels among conifers and demonstrated a strong association between gene expression and plastomic rearrangements. We discuss possible mechanisms underlying this association.

Results

Both plastomic strands are fully transcribed in conifers

The six newly assembled plastomes are illustrated as linear molecules to facilitate pairwise comparisons (Fig. 1a). A plastomic inversion was detected in the sampled *K. davidiana* individual when it was compared to the conspecific reference (NC_011930; Fig. S1a). This polymorphic inversion is flanked by Pinaceae Type I repeats (Wu et al. 2011[18]), which are capable of triggering homologous recombination to generate predominant and substoichiometric plastomic isomers in *K. davidiana* (Fig. S1b).

RNAseq coverage across the six sampled conifer plastomes is represented as histograms in Fig. 1. RNAseq coverage of rRNAs is low, indicating effective depletion of rRNA transcripts prior to sequencing. We also found that over 94.2% of the plastome sequences were covered by RNAs generated from a specific single DNA strand. The coverage ratio increased to over 99.9% after RNAseq reads from both strands were combined (Fig. S2). Overall, our data reveal almost full transcription of both plastomic strands, indicating that intergenic, intronic, and antisense transcripts are ubiquitous in conifer plastids.

Our data also show that CDS sense transcripts are generally more abundant than their antisense counterparts, although there are several exceptions (Fig. S3). For example, *psbN*, a photosynthetic system II gene, is located on the strand opposite to the *psbB* operon, a well-known polycistronic transcription unit that comprises four genes: *psbB*, *psbH*, *petB*, and *petD* [19]. Therefore, the transcripts antisense to *psbN* are likely overrepresented due to the strongly expressed *psbB* operon. This finding suggests that the positions of plastid genes strongly affect antisense RNA expression.

Influence of plastomic rearrangements on CDS expression

Among the six conifer plastomes, 31 syntenic blocks were identified to estimate plastomic rearrangements (Fig. S4a). Pairwise comparisons reveal 2 – 14 inversions among the sampled conifers (Fig. S4b). To normalize expression levels, RNAseq reads mapped to CDSs were collected and combined to calculate transcripts per million (TPM). Figure 2 compares the TPM scores between orthologous plastid CDSs that retain equivalent functions across conifer species. We found that (1) *psbA* and *rbcl* are the two most highly expressed genes in the presence of light and (2) TPM scores of these orthologous genes are significantly correlated (Pearson's correlation coefficients = 0.733 to 0.914, all $P < 0.001$),

suggesting that their expression levels are strongly functionally dependent. However, these correlation coefficients are inversely associated with the number of plastomic rearrangements (PR) between species (Pearson's $\rho = -0.626$, $P = 0.013$). Taken together, our results demonstrate that plastomic rearrangements reduce the strength of functionally-dependent association of plastid gene expression. In other words, these rearrangements influence gene expression in conifers.

Plastid RNA editing occurs in both sense and antisense transcripts

We detected 78 C-to-U RNA-editing sites in *K. davidiana* plastids, 42 in *A. dammara*, 23 in *N. nagi*, 35 in *S. verticillata*, 32 in *Ce. wilsoniana*, and 21 in *Cu. konishii* (Fig. 3a; Table S1). Notably, the majority (76.2–96.9%) of these edited sites cause non-silent editing, introducing non-synonymous changes in amino acid sequences. In contrast, silent-editing sites at synonymous codon positions occur in only 0–14.3% of the sites. In addition, editing efficiency at silent-editing sites is nearly always less than 50%, with two exceptions: *ndhE* of *A. dammara* and *psbA* of *S. verticillata* (Fig. 3b; Table S1). We also discovered one to three editing sites in antisense transcripts of CDSs from each conifer species. These sites are partially edited, with efficiency less than 50% (Fig. 3b).

We further investigated the intersection among these edited sites based on their alignments. In Fig. 3c, edited sites are designated as “shared” when they appear at the same alignment position in two or more species. Those found only in a single species are designated as “specific” sites. Most silent and antisense edited sites are species-specific. Only one site—located in the *rps8* transcript—is shared by all conifers, suggesting that it originated in the common ancestor of all conifers more than 300 million years ago [20]. In addition, *K. davidiana* plastids contain more species-specific RNA-editing sites than any other species we examined, with the proportion of “specific” sites exceeding other conifers by more than twofold (Fig. S5). This finding implies that Pinaceae has evolved a distinctive set of plastid RNA-editing sites after diverging from cupressophytes.

Discussion

We used strand-specific RNAseq data to explore plastid transcriptomic profiles across all six conifer families. Our data indicate that conifer plastomes transcribe nearly full sequences of both DNA strands, reinforcing the viewpoint that full transcription of plastomic sequences is the norm rather than an exception among seed plants [8]. In addition, we identified a number of C-to-U edited sites in sense and antisense transcripts. These results suggest that strand-specific RNAseq improves the detection of RNA-editing sites by not only removing antisense contamination during mapping but also allowing for the exploration of editing events in antisense transcripts. Notably, all antisense sites are edited inefficiently, implying that they are likely accidental or tissue-specific [21–23].

We also discovered numerous RNAseq reads mapping onto introns, intergenic spacers (IGSs), and the regions antisense to CDSs. Plastid non-coding RNAs were proposed to regulate gene expression [24]. In some model plants, plastid CDS and IGS transcripts have similar expression levels [7]. We did not

compare transcript abundance between CDSs and IGSs because the latter's transcriptional orientation was uncertain, making it difficult to identify the corresponding RNAseq reads in a strand-specific manner. Nonetheless, we did observe numerous transcripts antisense to CDSs. In plastids, antisense transcripts were hypothesized to bind to the 3' end of mRNAs and stabilize them [25]. This stabilization mechanism is likely active for all CDS transcripts since their antisense counterparts are prevalent in conifer plastids.

It has long been known that transcription termination of most plastid genes is inefficient as it results in abundant and diverse read-through transcripts that must be post-transcriptionally processed [26]. In a recent study [27], the mechanism of read-through transcription, which affects the transcription of downstream genes, resulted in extreme accumulation of *accD* transcripts when transcription termination of the upstream gene, *rbcL*, was inactivated. Here, we propose that read-through transcription also helps interpret our finding that plastomic rearrangements influence gene expression. Relocating a gene involves reconfiguring its neighboring loci and thus altering the read-through transcription effect from the upstream gene. This ultimately changes the number of transcripts of the relocated gene and its downstream neighbors. However, without environmental stress treatments, it is difficult to link altered gene expression from plastomic rearrangements with a biological adaptation. Fortunately, inter- and intra-specific plastomic inversions have been documented in several conifer lineages [this study; 18, 28 – 31], providing ideal material to study the association between plastomic rearrangements and biological adaptation in the future.

Conclusion

It has long been known that plastomic rearrangements occur frequently in conifers. However, gene expression dynamics in the relocated plastid gene (after rearrangements) and its downstream neighbors has not been investigated. In this pivotal study, we show that in conifers (1) both plastomic strands are fully transcribed, (2) increased plastomic arrangements reduce the strength of functionally-dependent association of plastid gene expression, (3) RNA editing occurs in both sense and antisense transcripts, and (4) the Pinaceae have evolved a distinctive set of plastid RNA-editing sites after diverging from cupressophytes. The tight association of plastomic rearrangements with gene expression leads us to propose that read-through transcription is likely the key to make this association. Additional studies and molecular biology validation are needed to better understand the biological adaptation of plastomic rearrangements in conifers.

Methods

Plant materials, DNA and RNA extraction and sequencing

The six representative conifer species (i.e., *Keteleeria davidiana* for Pinaceae, *Agathis dammara* for Araucariaceae, *Nageia nagi* for Podocarpaceae, *Sciadopitys verticillata* for Sciadopityaceae, *Cephalotaxus wilsoniana* for Taxaceae, and *Cunninghamia konishii* for Cupressaceae) were collected and identified by Dr. Chung-Shien Wu (Biodiversity Research Center, Academia Sinica). Permission was

not necessary for collecting these plants. The voucher specimens were deposited at the Herbarium, Biodiversity Research Center, Academia Sinica, Taipei (HAST; Table S2).

For DNA and RNA extraction, approximately 30 cm of fresh young shoots were collected from the six conifer species. To reduce potential variability due to different growth conditions, shoots were grown hydroponically in a growth chamber (GC-550R, Yihder Company, New Taipei City) at 25 °C with a light intensity of $100 \mu\text{mol m}^{-2} \text{s}^{-1}$. After 24 hours, fresh leaves on the shoots were harvested for DNA and RNA extraction using the methods described in [32] and [33], respectively. The extracted DNA was sequenced at Genomics BioSci & Tech (New Taipei City, Taiwan) on an Illumina HiSeq 4000 system. We also performed strand-specific RNAseq using the same system after DNase I (Invitrogen) treatment, rRNA depletion (Illumina Ribo-Zero rRNA Removal kits, Plant Leaf version), and library construction with dUTP and random hexamers. Table S2 details the information on sampling locality, voucher numbers, GenBank accessions, and DNAseq and RNAseq read counts used in this study.

Plastome assembly and RNA mapping analysis

Plastome assembly was initially conducted using SPAdes 3.13 [34] with the option of “careful” and a range of k-mer sizes (21, 33, 55, 77, and 99). Plastomic contigs were identified using NCBI-blast 2.2.18 [35] against in-house databases. Gaps between contigs were closed using GapCloser 1.12 [36]. This yielded complete plastomes for all sampled conifers, except *K. davidiana* because of its long Pinaceae Type I repeats [18]. We subsequently designed specific primers (Fig. S1) to amplify the corresponding regions and perform genome finishing in the latter species.

For each conifer species, 20 million paired-end RNAseq reads were randomly extracted and mapped to the corresponding plastome using TopHat 2.1.1 [37] with the parameters: library-type = fr-firststrand, read-mismatches = 15, read-gap-length = 0, and read-edit-dist = 15. Samtools 1.9 [38] was used to sort, filter, and combine the mapped reads. The resulting BAM files were imported into Geneious 11.1.5 (<https://www.geneious.com>) to calculate read counts and conduct downstream analyses. The RNAseq coverage, which refers to mapped read counts per base, was calculated using 100-bp non-overlapping sliding windows across plastomes, followed by transformation with the formula: $\text{Log}_{10}(\text{coverage} + 1) / \text{Log}_{10}(\text{maximum coverage} + 1)$.

Identification and visualization of RNA-editing sites

To identify C-to-U RNA-editing sites, the “Find Variations” option implemented in Geneious 11.1.5 was employed with the threshold: minimum coverage = 50, minimum variant frequency = 0.1, and maximum variant *P*-value = 10^{-6} . Editing efficiencies were estimated by calculating the ratio of edited to unedited bases in mapped reads. Intersection among edited sites from the six conifers were evaluated using UpSetR [39]. Plastome maps, transcriptional profiles, and RNA-editing sites were visualized using Circos 0.67 [40].

Abbreviations

CDS

Coding sequence; IGS: Intergenic spacer; TPM: Transcripts per million; PR: Plastomic rearrangements; IR: Inverted repeat; Rho (ρ): Pearson correlation coefficient

Declarations

Acknowledgments

We thank *Taipei Botanical Garden and Taipei Floriculture Experiment Center* for providing the plant materials.

Authors' contributions

CSW and SMC conceived and designed the study. CSW performed the experiments and analyzed the data. CSW, ES, SMC wrote the manuscript. All authors read and approved the manuscript.

Funding

This work was supported by research grants from Ministry of Science and Technology Taiwan (MOST 106-2311-B-001-005) and from Biodiversity Research Center of Academia Sinica to S.-M.C. The funding bodies were not involved in the design of the

study, collection, analysis, and interpretation of data, and in writing the manuscript.

Availability of data and materials

All DNA data and their annotations were deposited into DDBJ DataBank under accession numbers LC571739–LC571883.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Author details

¹Biodiversity Research Center, Academia Sinica, Taipei 11529, Taiwan.

References

1. Mofikoya OO, Mäkinen M, Jänis J. Chemical fingerprinting of conifer needle essential oils and solvent extracts by ultrahigh-resolution fourier transform ion cyclotron resonance mass spectrometry. *ACS Omega*. 2020; 5:10543–10552 <http://doi.org/10.1021/acsomega.0c00901>.
2. Pfannschmidt T, Blanvillain R, Merendino L, Courtois F, Chevalier F, Liebers M, et al. Plastid RNA polymerases: orchestration of enzymes with different evolutionary origins controls chloroplast biogenesis during the plant life cycle. *J Exp Bot*. 2015; 66: 6957–6973 <http://doi.org/10.1093/jxb/erv415>.
3. del Campo EM. Post-transcriptional control of chloroplast gene expression. *Gene Regul Syst Bio*. 2009; 3:31–47 <http://doi.org/10.1104/pp.125.1.142>.
4. Hotto AM, Schmitz RJ, Fei Z, Ecker JR, Stern DB. 2011. Unexpected diversity of chloroplast noncoding RNAs as revealed by deep sequencing of the *Arabidopsis* G3. 2011; 1:559–570 <http://doi.org/10.1534/g3.111.000752>
5. Zhelyazkova P, Sharma CM, Förstner KU, Liere K, Vogel J, Börner T. The primary transcriptome of barley chloroplasts: numerous noncoding RNAs and the dominating role of the plastid-encoded RNA polymerase. *Plant Cell*. 2012; 24:123–136 <http://doi.org/10.1105/tpc.111.089441>.
6. Castandet B, Germain A, Hotto AM, Stern DB. Systematic sequencing of chloroplast transcript termini from *Arabidopsis thaliana* reveals >200 transcription initiation sites and the extensive imprints of RNA-binding proteins and secondary structures. *Nucleic Acids Res*. 2019; 47:11889–11905 <http://doi.org/10.1093/nar/gkz1059>.
7. Shi C, Wang S, Xia EH, Jiang JJ, Zeng FC, Gao LZ. Full transcription of the chloroplast genome in photosynthetic eukaryotes. *Sci Rep*. 2016; 6: 30135 <http://doi.org/10.1038/srep30135>.
8. Sanitá Lima M, Smith DR. Pervasive transcription of mitochondrial, plastid, and nucleomorph genomes across diverse plastid-bearing species. *Genome Biol Evol*. 2017; 9:2650–2657 <http://doi.org/10.1093/gbe/evx207>.
9. Ishibashi K, Small I, Shikanai T. Evolutionary model of plastidial RNA editing in angiosperms presumed from genome-wide analysis of *Amborella Trichopoda*. *Plant Cell Physiol*. 2019; 60:2141–2151 <http://doi.org/10.1093/pcp/pcz111>.
10. Raubeson L, Jansen R. Chloroplast genomes of plants. In: Henry R, editor. *Plant diversity and evolution: genotypic and phenotypic variation in higher plants*. Cambridge, MA: CABI Publishing. 2005. p. 45–68 <http://doi.org/10.1079/9780851999043.0045>.
11. Cui L, Leebens-Mack J, Wang LS, Tang J, Rymarquis L, Stern DB, et al. Adaptive evolution of chloroplast genome structure inferred using a parametric bootstrap approach. *BMC Evol Biol*. 2006; 6:13 <https://doi.org/10.1186/1471-2148-6-13>.
12. Wu CS, Chaw SM. Highly rearranged and size-variable chloroplast genomes in conifers II clade (cupressophytes): evolution towards shorter intergenic spacers. *Plant Biotechnol J*. 2014; 12:344–353 <http://doi.org/10.1111/pbi.12141>.
13. Chaw SM, Wu CS, Sudianto E. Evolution of gymnosperm plastid genomes. In Chaw SM, Jansen RK, editors. *Advances in botanical research*. Cambridge, MA: Academic Press. 2018. p. 195–222

<http://doi.org/10.1016/bs.abr.2017.11.018>.

14. Sudianto E, Wu CS, Chaw SM. The origin and evolution of plastid genome downsizing in southern hemispheric cypresses (Cupressaceae). *Front. Plant Sci.* 2020; 11:901 <http://doi.org/10.3389/fpls.2020.00901>.
15. Hsu CY, Wu CS, Chaw SM. Birth of four chimeric plastid gene clusters in Japanese umbrella pine. *Genome Biol Evol.* 2016; 8:1776–1784 <http://doi.org/10.1093/gbe/evw109>.
16. Wu CS, Chaw SM. Large-scale comparative analysis reveals the mechanisms driving plastomic compaction, reduction, and inversions in conifers II (Cupressophytes). *Genome Biol Evol.* 2016; 8:3740–3750 <http://doi.org/10.1093/gbe/evw278>.
17. Mills JD, Kawahara Y, Janitz M. Strand-specific RNA-Seq provides greater resolution of transcriptome profiling. *Curr Genomics.* 2013;14:173–181 <http://doi.org/10.2174/1389202911314030003>.
18. Wu CS, Lin CP, Hsu CY, Wang RJ, Chaw SM. Comparative chloroplast genomes of Pinaceae: insights into the mechanism of diversified genomic organizations. *Genome Biol Evol.* 2011; 3:309–319 <http://doi.org/10.1093/gbe/evr026>.
19. Westhoff P, Herrmann RG. Complex RNA maturation in chloroplasts. The *psbB* operon from spinach. *Eur J Biochem.* 1988; 171:551–564 <http://doi.org/10.1111/j.1432-1033.1988.tb13824.x>.
20. Leslie AB, Beaulieu J, Holman G, Campbell CS, Mei W, Raubeson LR, et al. An overview of extant conifer evolution from the perspective of the fossil record. *Am J Bot.* 2018; 105:1531–1544 <http://doi.org/10.1002/ajb2.1143>.
21. Sloan DB, MacQueen AH, Alverson AJ, Palmer JD, Taylor DR. Extensive loss of RNA editing sites in rapidly evolving *Silene* mitochondrial genomes: selection vs. retroprocessing as the driving force. *Genetics.* 2010; 185:1369–1380 <http://doi.org/10.1534/genetics.110.118000>.
22. Bentolila S, Oh J, Hanson MR, Bukowski R. 2013. Comprehensive high-resolution analysis of the role of an Arabidopsis gene family in RNA editing. *PLoS Genet.* 2013; 9: e1003584 <https://doi.org/10.1371/journal.pgen.1003584>.
23. Chen TC, Liu YC, Wang X, Wu CH, Huang CH, Chang CC. Bot Stud. Whole plastid transcriptomes reveal abundant RNA editing sites and differential editing status in *Phalaenopsis aphroditeformosana*. *Bot Stud.* 2017; 58:38 <http://doi.org/10.1186/s40529-017-0193-7>.
24. Hotto AM, Germain A, Stern DB. Plastid non-coding RNAs: emerging candidates for gene regulation. *Trends Plant Sci.* 2012; 17:737–744 <http://doi.org/10.1016/j.tplants.2012.08.002>.
25. Manavski N, Schmid LM, Meurer J. RNA-stabilization factors in chloroplasts of vascular plants. *Essays Biochem.* 2018; 62:51–64 <http://doi.org/10.1042/EBC20170061>.
26. Stern DB, Goldschmidt-Clermont M, Hanson MR. Chloroplast RNA metabolism. *Annu Rev Plant Biol.* 2010; 61:125–155 <http://doi.org/10.1146/annurev-arplant-042809-112242>.
27. Ji D, Manavski N, Meurer J, Zhang L, Chi W. 2019. Regulated chloroplast transcription termination. *Biochim Biophys Acta Bioenerg* 1860:69–77 <http://doi.org/10.1016/j.bbabi.2018.11.011>.

28. Guo W, Grewe F, Cobo-Clark A, Fan W, Duan Z, Adams RP, et al. 2014. Predominant and substoichiometric isomers of the plastid genome coexist within *Juniperus* plants and have shifted multiple times during cupressophyte evolution. *Genome Biol Evol.* 2014; 6:580–590 <http://doi.org/10.1093/gbe/evu046>.
29. Sullivan AR, Schiffthaler B, Thompson SL, Street NR, Wang XR. Interspecific plastome recombination reflects ancient reticulate evolution in *Picea* (Pinaceae). *Mol Biol Evol.* 2017; 34:1689–1701 <http://doi.org/10.1093/molbev/msx111>.
30. Qu XJ, Wu CS, Chaw SM, Yi TS. Insights into the existence of isomeric plastomes in Cupressoideae (Cupressaceae). *Genome Biol Evol.* 2017; 9:1110–1119 <http://doi.org/10.1093/gbe/evx071>.
31. Fu CN, Wu CS, Ye LJ, Mo ZQ, Liu J, Chang YW, et al. Prevalence of isomeric plastomes and effectiveness of plastome super-barcodes in yews (*Taxus*) worldwide. *Sci Rep.* 2019; 9:2773 <http://doi.org/10.1038/s41598-019-39161-x>.
32. Stewart CN Jr, Via LE. A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. *Biotechniques.* 1993; 14:748–750 [http://doi.org/10.1016/0167-7012\(94\)00045-9](http://doi.org/10.1016/0167-7012(94)00045-9).
33. Kolosova N, Miller B, Ralph S, Ellis BE, Douglas C, Ritland K, et al. Isolation of high-quality RNA from gymnosperm and angiosperm trees. *Biotechniques.* 2004; 36:821–824 <http://doi.org/10.2144/04365ST06>.
34. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 2012; 19:455–477 <https://doi.org/10.1089/cmb.2012.0021>.
35. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. 2009. BLAST+: architecture and applications. *BMC Bioinformatics.* 2009; 10:421 <https://doi.org/10.1186/1471-2105-10-421>.
36. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience.* 2012; 1:18 <http://doi.org/10.1186/2047-217X-1-18>.
37. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 2013; 14:R36 <http://doi.org/10.1186/gb-2013-14-4-r36>.
38. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics.* 2009; 25:2078–2079 <http://doi.org/10.1093/bioinformatics/btp352>.
39. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics.* 2017; 33:2938–2940 <http://doi.org/10.1093/bioinformatics/btx364>.
40. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. *Genome Res.* 2009; 19:1639–1645 <http://doi.org/10.1101/gr.092759.109>.

Supplementary Information

Additional file 1 Table S1. Plastid C-to-U editing sites detected in this study.

Additional file 2 Table S2. Conifer species, voucher numbers, DNaseq, and RNAseq data used in this study.

Additional file 3 Figure S1. Plastomic isomers in *K. davidiana*. (a) Comparison of plastomes shows an intraspecific inversion flanked by the Pinaceae type I inverted repeat (IR). Primers used to detect specific isomers are indicated. (b) Semi-quantitative PCR demonstrates the coexistence of two isomers containing distinctive copy numbers.

Additional file 4 Figure S2. Percentages of plastomic sequences covered by stranded RNAseq reads.

Additional file 5 Figure S3. RNAseq coverage of CDS transcripts and their antisense counterparts in the conifer plastids. Coverage scores were transformed using $\text{Log}_{10}(1 + \text{coverage})$. Dashed lines denote diagonal lines. CDSs are indicated if the coverage scores of their transcripts are smaller than those of their antisense counterparts.

Additional file 6 Figure S4. Extensive plastomic rearrangements in conifers. (a) Thirty-one syntenic regions (color boxes) identified in the six sampled conifer plastomes. (b) Number of plastomic inversions between species.

Additional file 7 Figure S5. Proportion of specific and shared RNA-editing sites in the six representative conifer plastids.

Figures

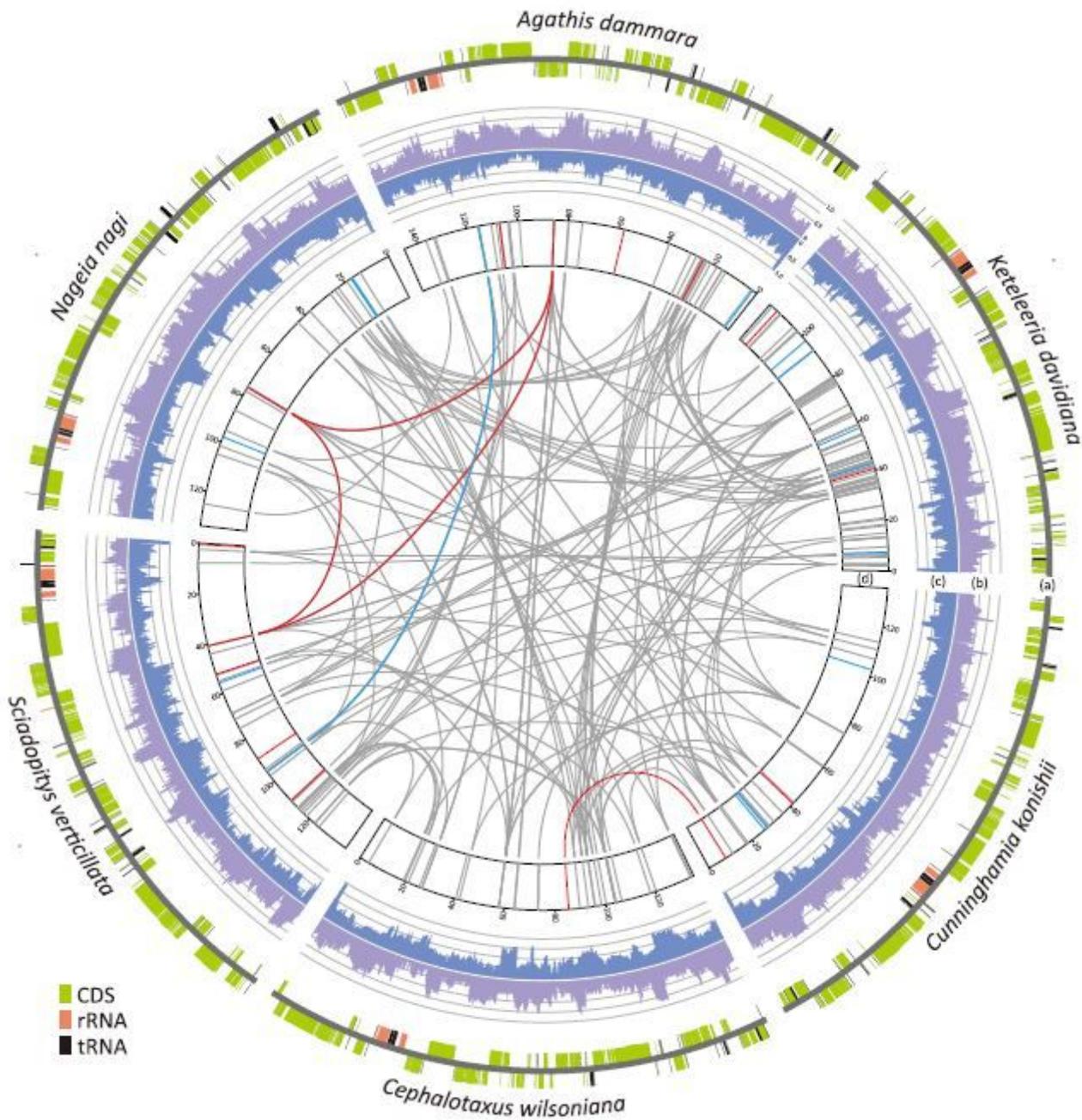


Figure 1

Plastid transcriptomic profiles of the six sampled conifer species. (a) Plastomic maps with genes in outer and inner strands transcribed clockwise and counterclockwise, respectively. Transcriptomic profiles where outer (b) and inner (c) histograms represent RNAseq coverage (read counts per base) after transformation by the formula: $\text{Log}_{10}(\text{coverage} + 1) / \text{Log}_{10}(\text{maximum coverage} + 1)$. (d) Distribution of RNA-editing sites where red, blue, and grey lines denote anti-sense, silent, and non-silent editing, respectively. Shared edited sites are linked by lines.

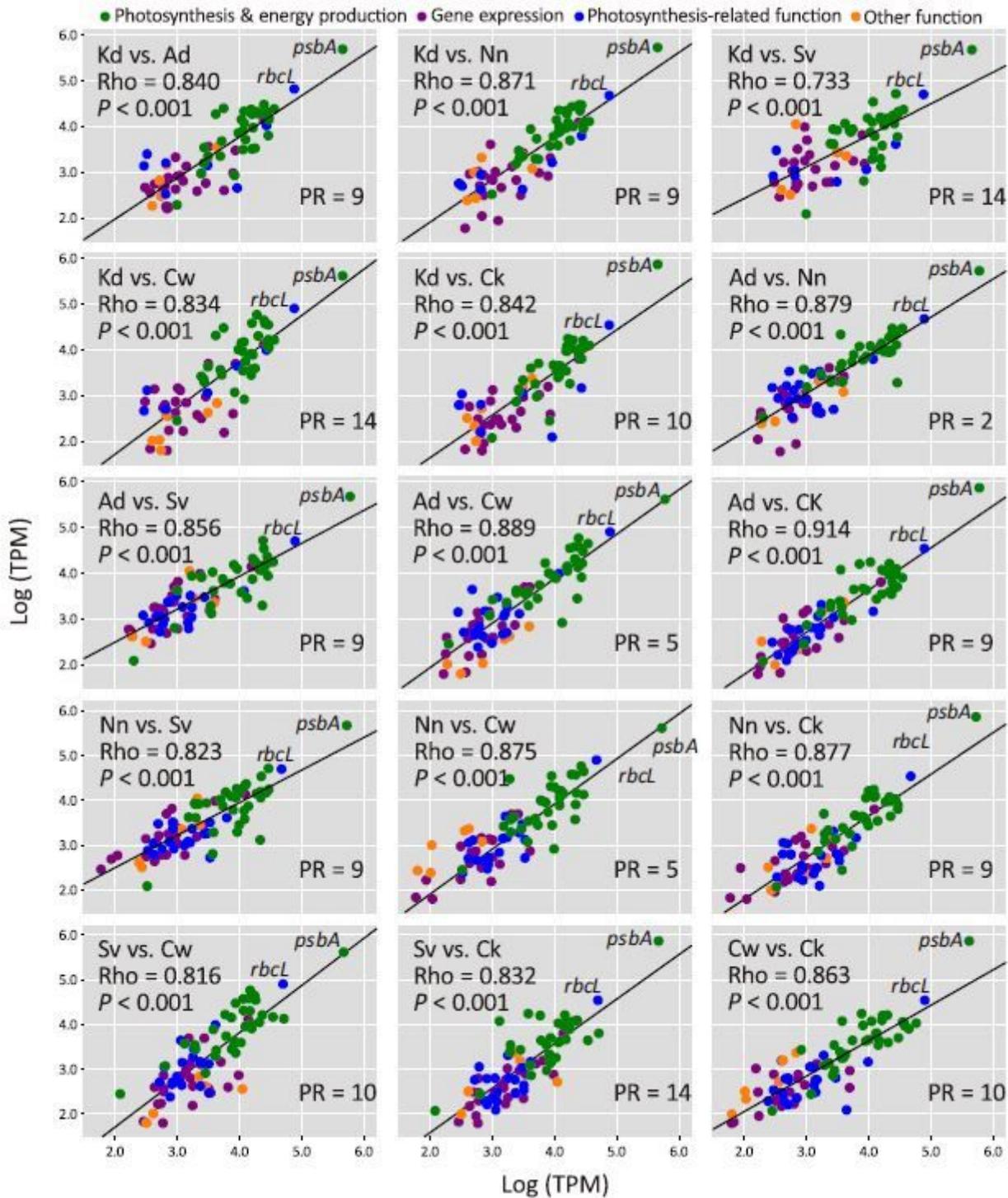


Figure 2

Strong correlations among plastid CDS expression levels in the sampled conifer species. Kd, *K. davidiana*; Ad, *A. dammara*; Nn, *N. nagi*; Sv, *S. verticillata*; Cw, *Ce. wilsoniana*; Ck, *Cu. konishii*; Rho (ρ), Pearson correlation coefficient; PR, numbers of plastomic rearrangements between species.

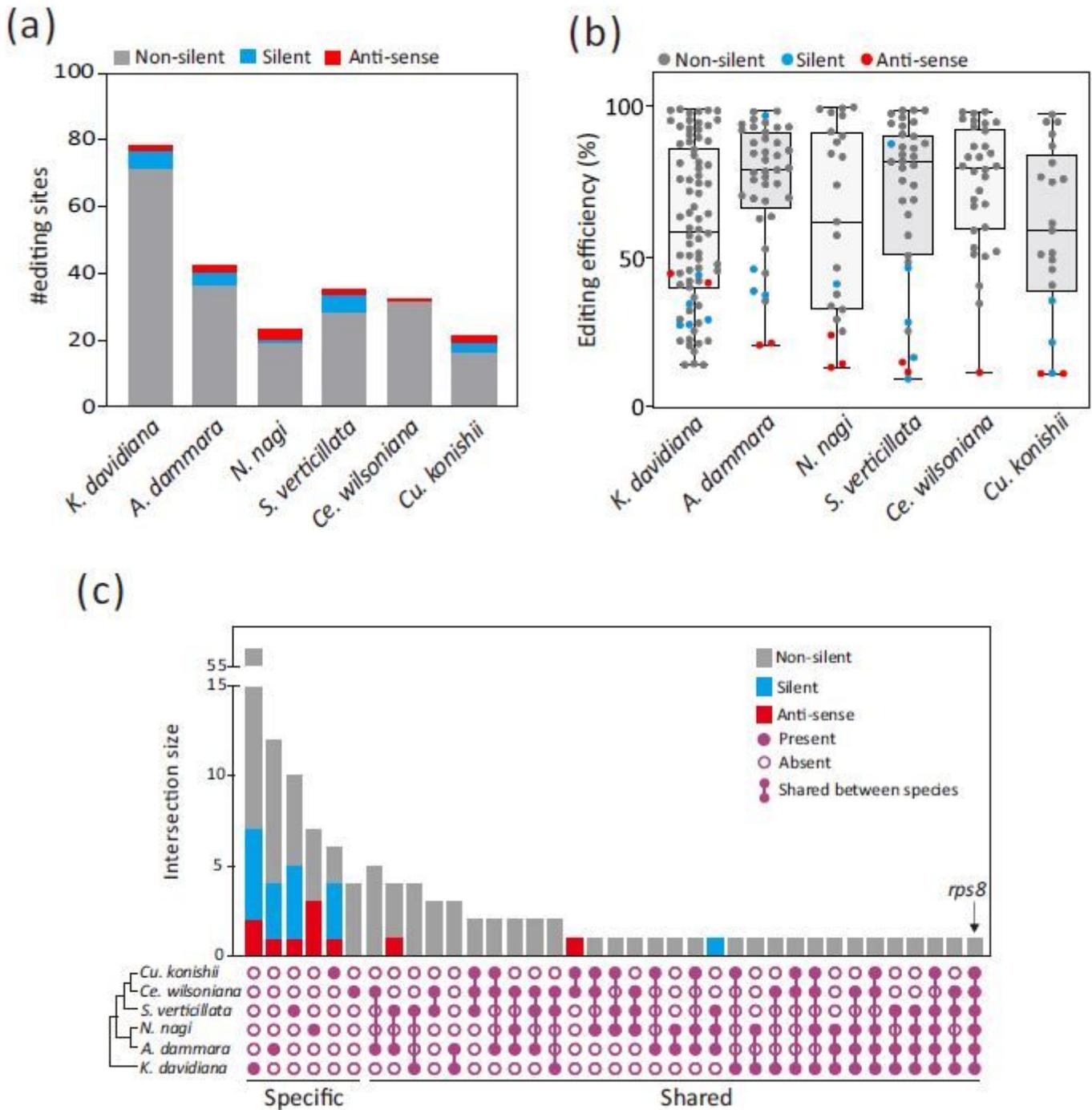


Figure 3

Plastid RNA-editing sites detected in the six sampled conifers. (a) RNA-edited sites, and (b) their editing efficiencies. (c) UpSet plot illustrating the intersections of these edited sites.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementarymaterials1009.pdf](#)