bioRxiv preprint doi: https://doi.org/10.1101/097519; this version posted December 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Interspecific plastome recombination reveals ancient reticulate evolution in *Picea* (Pinaceae)

Alexis R. Sullivan[1,2], Bastian Schiffthaler[3], Stacey Lee Thompson[1,4], Nathaniel R. Street[3], Xiao-Ru Wang[1]

1 – Umeå Plant Science Center, Department of Ecology and Environmental Science, Umeå University, SE-901 87 Umeå, Sweden

2 – Bergvik Skog AB, SE-791 71 Falun, Sweden

3 – Umeå Plant Science Center, Department of Plant Physiology, Umeå University, SE-901 87 Umeå, Sweden

4 – Department of Biology, Dalhousie University, Halifax, NS B3H 4R2, Canada

Corresponding author: Alexis R. Sullivan alexis.sullivan@umu.se

bioRxiv preprint doi: https://doi.org/10.1101/097519; this version posted December 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Plastid sequences are a cornerstone in plant systematic studies but key aspects of their evolution, such as uniparental inheritance and absent recombination, are often treated as axioms. While exceptions to these assumptions can profoundly influence evolutionary inference, detecting them can require extensive sampling, abundant sequence data, and detailed testing. Using advancements in high-throughput sequencing, we analyzed the whole plastomes of 65 accessions of *Picea*, a genus of ~35 coniferous forest trees, to test for deviations from canonical plastome evolution. Using complementary hypothesis and data-driven tests, we found evidence for plastomes generated by interspecific hybridization and recombination in the clade comprising Norway spruce (*P. abies*) and ten other species. Support for interspecific recombination remained after controlling for sequence saturation, positive selection, and potential alignment artifacts. These results reconcile conflicting plastid-based phylogenies reported in previous studies and strengthen the mounting evidence of reticulate evolution in *Picea.* Given the relatively high frequency of biparental plastid inheritance and hybridization in plants, we suggest interspecific plastome recombination may be more widespread than currently appreciated and could underlie reported cases of discordant plastid phylogenies.

Key words: hybridization, chloroplast genomes, phylogenetic incongruence, recombination, reticulate evolution, plastid

bioRxiv preprint doi: https://doi.org/10.1101/097519; this version posted December 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

Chloroplasts and other plastids retain a compact haploid genome derived from their cyanobacterium ancestor. While nuclear genomes vary in size and chromosome number by orders of magnitude, a typical land plant plastome contains around 115 unique genes distributed over 120-160 kb (Bock 2007). Genes are typically highly conserved among species and encode the two photosystems, the large subunit of the carbon fixing enzyme RuBisCO, and ~30 housekeeping and other proteins. Unambiguous orthology, coupled with a smaller effective population size relative to the nuclear genome, have made the plastome a workhorse of historical inference. Underpinning many studies of plant evolution are key assumptions of unbroken uniparental inheritance, strong functional constraint, presence of a single haplotype within an individual, and an absence of recombination (Wolfe & Randle 2004).

Whole plastome sequencing is rapidly becoming routine due to its utility in plant systematics. As a result, recent comparative plastome analyses within genera have found pronounced changes in plastome inheritance, architecture, and divergence across relatively short timescales. For example, *Silene* plastomes have undergone extensive structural changes (Sloan et al. 2012), which contrasts with the conservation observed across most angiosperms. In *Citrus*, biparental plastid inheritance and heteroplasmy is surprisingly frequent and associated with interspecific hybridization (Carbonell-Caballero et al. 2015). While such studies have documented deviations from canonical plastome evolution that may have important implications for plant evolutionary research, deep taxon sampling within genera is uncommon, especially in gymnosperms and non-crop species (but see Parks et al. 2009; Whittall et al. 2010; Ruhsam et al. 2015).

Gymnosperms, and the Pinaceae in particular, differ markedly in plastome organization and inheritance from most land plants. Most angiosperm plastomes contain two large inverted repeats (IRs) and two single copy regions, but the IRs are highly reduced or lost entirely in conifers (Strauss et al. 1988; Wu et al. 2011). Rather than the typical quadripartite structure shared by most angiosperms, Pinaceae plastomes contain three single copy regions (F1, F2, and large fragments), which are intervened by large repeats (Wu et al. 2011, see Fig. 1). Both dispersed repeats and IR loss are associated with genome instability in land plants (Strauss et al. 1988), likely due to increased recombination (Day & Madesis 2007). In the spruce genus *Picea* (Pinaceae), large re-arrangements involving the repeat regions have been recently documented among several species (Wu et al. 2011; Nystedt et al. 2013), which may indicate a recombination-prone plastome. Conifers also differ from most angiosperms in their predominately paternal plastid inheritance (e.g. Neale & Sederoff 1989), although the significance of this for plastome evolution is not yet clear (Crosby & Smith 2012). These novelties relative to angiosperms make *Picea* a useful system for studying plastome evolution.

Picea comprises about 35 commonly recognized species widely distributed across Northern Hemisphere (Farjon 1991; Eckenwalder 2009). Fundamental aspects of *Picea* evolution remain unclear due to their broad morphological similarity, large nuclear genomes (Birol et al. 2013; Nystedt

bioRxiv preprint doi: https://doi.org/10.1101/097519; this version posted December 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

et al. 2013), rapid and recent radiation (Lockwood et al. 2013; Ran et al. 2015), and potentially high levels of introgression (e.g. Jaramillo-Correa et al. 2005; Du et al. 2011; Hamilton & Aitken 2013). Recent plastid-based phylogenies have yielded conflicting topologies depending on the locus analyzed (cf., Bouillé et al. 2011; Ran et al. 2015), a result Bouillé et al. (2011) attributed to interspecific recombination. In support of this hypothesis, *Picea* are widely interfertile and hybridization is linked to rare biparental plastid inheritance in Pinaceae and some angiosperms (Sutton et al. 1991; Szmidt et al. 1987; Hansen et al. 2007; Ellis et al. 2008). Hybrids could inherit two plastome copies with divergent ancestry, which may recombine and be transmitted to future offspring. However, robust evidence for chimeric plastomes in natural populations has been lacking (but see Marshall et al. 2001; Huang et al. 2001; D'Alelio & Ruggiero 2015), in part due to poor taxon sampling and the lack of complete plastome sequences.

In this study, we analyzed 65 whole plastomes from all ~35 extant species of *Picea*. We describe a simple and efficient sequencing strategy, characterize genome-wide sequence divergence, and reconstruct the evolutionary history of the *Picea* plastome. Focusing on the role dispersed repeats play in homologous recombination (Day & Madesis 2007), we tested the hypothesis that phylogenetic incongruence, if present, should occur among the three plastome fragments in *Picea* that are separated by large repeats (see Fig. 1). Combining this hypothesis-based phylogenetic analysis with purely data-driven tests, we identified at least one interspecific plastome recombination event, which resulted in a chimeric genome reflecting two divergent lineages. Over one-third of the genus inherited this chimeric plastome, which reconciles previous discordant plastid phylogenies (i.e. Bouillé et al. 2011; Ran et al. 2015). Given the prevalence of hybridization among plants, our results illustrate the importance of identifying non-bifurcating evolution, even in predominantly uniparentally inherited, non-recombining genomes.

Results

Plastome assembly and divergence

We used whole-genome shotgun sequencing to generate 61 draft plastomes, which represent all 35 commonly-recognized *Picea* species. Total genomic DNA was sequenced on three different Illumina platforms at differing multiplexing levels and read lengths (See Materials and Methods), but all assemblies had high coverage and large scaffolds (Table 1). *De novo* plastomes averaged 121,885 bp (± 2,861) or about 98% of reported reference-quality assemblies (Nystedt et al. 2013; Jackman et al. 2015; Wu et al. 2011; Yang et al. 2016). Assemblies generally comprised five large non-overlapping scaffolds, which correspond to the intervening regions between the two highly-reduced inverted repeats (IRa, IRb) and the three large Pinaceae-specific dispersed repeats (Fig. 1). Plastomes assembled from two paired-end libraries were more contiguous despite lower sequencing depth and shorter reads (Table 1), possibly because of improved resolution of repetitive sequences (Prjibelski et al. 2014) or differences in Illumina chemistry. Assembly statistics, detailed locality information, and

bioRxiv preprint doi: https://doi.org/10.1101/097519; this version posted December 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Genbank and Short Read Archive (SRA) accession numbers for each individual are reported in Supplementary Table S1.

Subsequent analyses included four additional published *Picea* plastomes (Nystedt et al. 2013; Jackman et al. 2015; Wu et al. 2011; Yang et al. 2016), yielding a total of 65 assembled genomes. Spruce plastomes were highly similar, averaging 97.9% sequence identity in pair-wise comparisons. Gene content was conserved and comprised 74 protein-coding genes, 36 tRNA genes, and four rRNAs. No evidence of rearrangements within scaffolds was found, but we were unable to evaluate structural changes spanning multiple scaffolds, such as the inversion of the F2 fragment between *P. morrisonicola* and *P. abies* (Nystedt et al. 2013). Indels were found in only three genes: *acc*D, *chl*N, and *ycf*1.

Nucleotide diversity (π) across the plastome was relatively low (0.0030; Table 2). Sliding-scale estimates of π in non-overlapping 350 bp windows revealed three 'hot-spots' of diversity occurring near the boundaries of the structural fragments (Fig. 1). Five genes (*psb*H, *psb*T, *rpl*23, *ycf*1, *ycf*2) were more polymorphic than intergenic spacers on average. Notably, *ycf*1 and *ycf*2 together contain 57% of all coding sequence polymorphisms (Supplementary Table S2). Average diversity at protein-coding genes and four-fold (4D) synonymous sites were similar (Table 2) but substantially lower than in intergenic spacers. Given the essential role the plastome plays in photosynthesis and energy metabolic pathways, the observed constraint acting on 4D sites is likely not limited to spruces.

Mean synonymous and non-synonymous divergence was generally low (dN = 0.017 ± 0.024, dS = 0.040 ± 0.031) but varied markedly among genes (Fig. 2). Mean substitution rates were significantly elevated rates in the *ycf* genes (GLM, $t = 2.828$, $p < 0.001$), a pattern driven entirely by *ycf*1 and *ycf*2, and in the photosystem II complex *psb* (GLM, $t = 2.103$, $p < 0.05$; Fig. 2). Six genes comprised solely non-synonymous substitutions while eight contained only synonymous substitutions (Fig. 2; Supplementary Table S2).

The variation in substitution rates among genes prompted us to test for selection acting on them. Codon-based models of nucleotide substitution strongly supported ($p \leq 0.001$) positive evolution ($\omega > 1$) in 12 out of 70 genes (Table 3). Five additional genes were significant at $0.001 < p \leq 0.05$ but we did not consider them further (Supplementary Table S2). Most genes had one or two positively selected codons, but the massive *ycf*1 gene contained 73 sites (3.5% of the predicted amino acid) potentially under selection. However, many of these sites were located within repetitive sequences and poor alignment may have generated false positives. After excluding these regions, positive selection was still supported at 21 codons (presented in Table 3).

Phylogenetic discordance in the Picea *plastome*

Phylogenetic analysis of the 65 whole plastomes using maximum likelihood resolved five major clades: i) the North American *P. engelmannii*, *P. glauca*, and *P. mexicana* (hereinafter *P.*(*glauca*)), ii) *P. abies* and predominately northeast Asian taxa (*P.*(*abies*)), iii) *P. mariana* and four geographically-

bioRxiv preprint doi: https://doi.org/10.1101/097519; this version posted December 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

disparate species (*P.(mariana*)), iv) taxa from the Qinghai-Tibetan Plateau and surrounding regions (*P.*(*likiangensis*)), and v) a diverse clade comprising species spanning the entire range of genus (*P.*(*chihuahuana*); Fig. 3). Relationships within most clades were strongly supported with the notable exception of the *P.*(*abies*), in which several nodes had less than 50% support. The uncertain topology is likely due to a rapid and recent radiation, likely within the last two million years (Lockwood et al. 2013, Ran et al. 2015). In contrast, the backbone topology was poorly resolved, with 52% of trees supporting a *P.*(*glauca*)-*P.*(*abies*) sister relationship and 39% supporting a most recent common ancestor of *P.*(*glauca*), *P.*(*abies*), and *P.*(*mariana*).

We analyzed bipartition frequencies in bootstrap replicate trees to assess for the presence of alternate topologies at these two weakly-supported backbone nodes. Out of 1,000 replicates, *P.*(*glauca*) occurred in a bipartition with *P. sitchensis* and *P. breweriana* at even greater frequency than the relationship included on the consensus tree in Fig. 3 (46 vs. 39% of bootstrap replicates). Both *P.*(*mariana*) and *P.*(*glauca*) were grouped with *P.*(*abies*) to the exclusion of the other spruces with almost equal frequency (52% vs. 48%). This result is consistent with prior plastid-based studies, which discovered conflicting sister clades for *P.*(*abies*) depending on the locus analyzed (cf. Bouillé et al. 2011; Ran et al. 2015). These results suggest the presence of highly-supported yet conflicting topologies in the whole-plastome alignment.

Phylogenetic discord caused by homoplasy or recombination could explain the near-equal occurrence of conflicting bipartitions in the whole-plastome dataset and the disagreement among earlier studies. To assess incongruence across the plastome, we estimated phylogenies from five different subsets: 1) protein-coding exons; 2) concatenated intergenic spacers; and 3-5) subsets corresponding to the three major structural fragments (F1, F2, and large fragments) of Pinaceae plastomes, which may serve as substrates for homologous recombination (Wu et al. 2011; see Fig. 1). We reason these repeats could also facilitate recombination between divergent plastome copies if biparental plastid inheritance and fusion occurred. If the entire *Picea* plastome shares a similar phylogenetic history, then we expect no significant discordance among data subsets. Discordance among structural fragments could indicate recombination whereas discordance between genic and intergenic regions would be consistent with convergence or saturation.

We implemented randomization tests of Robinson-Foulds (RF) tree distances, quartet distances (QD), and the parsimony-based incongruence length difference (ILD) metric to test the significance of phylogenetic incongruence between *a priori* data subsets. Incongruence between intergenic spacers and protein-coding exons was only supported by the ILD test (Table 4), which potentially has elevated Type I and II error rates under a variety of circumstances (Planet 2006). The lack of strongly-supported discordance between intergenic and coding sites reflects the absence of pervasive differences in phylogenetic signal, even if some sites are genuinely discordant. In contrast, all three tests strongly supported incongruence between the large fragment versus the F1 and F2 fragments (Table 4, Fig. 4). While the placement of *P.*(*mariana*) is uncertain in the large fragment

bioRxiv preprint doi: https://doi.org/10.1101/097519; this version posted December 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

topology, it is excluded from the bipartition comprising *P.*(*glauca*) and *P.*(*abies*) in 99% of bootstrap replicates. In the F1 and F2 phylogenies, *P.*(*mariana*) is resolved as the sister clade of *P.*(*abies*) with high support (Fig. 4). Together, these results suggest spatial variation in the evolutionary history of sites among the *Picea* plastome structural fragments.

Differences in gene content, nucleotide diversity, and substitution rates across the fragments could potentially explain the variation in topologies across fragments. Therefore, we tested if phylogenetic discordance between fragments could be explained by differences in sequence evolution by further comparing only 1) intergenic sites, 3) protein-coding exons, and 3) four-fold synonymous sites (4D) on each fragment. Because we found no evidence for discordance between the F1 and F2 fragments, we concatenated them for comparisons against the large fragment. Each of these comparisons found strong evidence for phylogenetic discord between the large and F1 and F2 fragments, with the exception of the ILD test on 4D sites (Table 4). As there are only ~100 parsimony informative 4D sites across the entire plastome, the ILD test may have insufficient power to detect discordance (Planet 2006). Each subset of sites (e.g., large fragment 4D sites) yielded a similar topology as the entire fragment (Fig. 4), albeit with poorer resolution (Supplementary Fig. S1). Therefore, phylogenetic discordance encompasses intergenic, synonymous, and non-synonymous sites and topological differences among the three fragments cannot be adequately explained by saturation or selection.

Recombination tests

Phylogenetic discordance among fragments is consistent with our *a priori* hypothesis of interspecific recombination in the plastome. To further support this inference, we employed RDP4 (Martin et al. 2015) to implement unguided tests for recombination across the whole plastome. This two-step approach helps safeguard against potentially spurious results from exploratory tests yet also allows the data to refute our hypothesis that recombination should involve the repeats subtending the three large structural fragments.

Seven tests implemented in RDP4 identified two putative recombination events, both with breakpoints estimated near fragment junctions (Fig. 5, Table 5). All tests supported a recombinant *P.*(*abies*) plastome originating from a *P.*(*glauca*) and a *P.*(*mariana*) parent (Table 5). While the majority of the plastome suggested a close relationship of *P.*(*abies*) to *P.*(*glauca*), a smaller region corresponding to the F2 fragment was highly similar to *P. jezoensis*, a *P.*(*mariana*) species distributed from central Japan to far northeastern Siberia. This close relationship between *P.*(*abies*) and *P. jezoensis* is also evident in our *a priori* analysis, which supports a shared ancestor of the F2 fragment between *P.*(*abies*) and *P. jezoensis* (Fig. 4). While maximum likelihood recombination breakpoints were placed near the F2 fragment junctions, the very low divergence of *Picea* plastomes likely hindered precise estimation and broad confidence intervals also encompassed most of the F1 fragment (Fig. 5). Although the phylogenetic discordance and recombination tests differ somewhat in the extent

bioRxiv preprint doi: https://doi.org/10.1101/097519; this version posted December 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

of the exchange, both support a recombinant *P.*(*abies*) plastome with at least the F2 fragment inherited from a *P.*(*mariana*)-like parent.

A second putative recombination event spanning the F1 and F2 fragments involved *P.*(*likiangensis*), a clade of Qinghai-Tibetan Plateau taxa, and an unknown or extinct species similar to *P. pungens,* a *P.*(*mariana*) species (Table 5, Fig. 5). However, recombinant plastomes in this lineage were not strongly supported by the fragment consensus phylogenies (Fig. 4). The large fragment topology supported a sister relationship between *P.*(*likiangensis*) and *P.*(*mariana*) with 68% bootstrap support (Supplementary Fig. S2), whereas the F1 and F2 fragments robustly supported a sister relationship between *P.*(*chihuahuana*) and *P.*(*likiangensis*) (Fig. 4). As the RF, QD, and ILD randomization tests are global measures of discordance, this more equivocal recombination event likely still contributed to the highly significant results in Table 4.

Discussion

Uniparental inheritance, strong functional constraints, and absent recombination are key but often untested assumptions in plant evolutionary studies (Wolfe & Randle 2004). Using whole genome assemblies and comprehensive taxon sampling, we found that *Picea* plastomes are highly conserved in sequence and structure but also shaped by interspecific recombination and punctuated by putative positive selection. Our results add to the growing body of evidence that exceptions to typical plastid evolution are not rare on evolutionary timescales and can generate misleading phylogenetic signals.

Sequence divergence

Plastome substitution rates are typically low in land plants (Wolfe et al. 1987). Consequently, interspecific distances in the *Picea* plastome are of the same magnitude as intraspecific variation in Pinaceae nuclear genes (e.g. Mosca et al. 2012). Genome-wide substitution rates are not commonly reported in genus-level whole plastome studies (e.g. Parks et al. 2009; Ruhsam et al. 2015; Carbonell-Caballero et al. 2015), however, substitution rates in *Picea* are similar to those reported from more limited sets of loci in *Pinus* (Willyard et al. 2007; Wang & Wang 2014). Both genera have ancient origins in the Cretaceous (Wang et al. 2000; Klymiuk & Stockey 2012), but the diversification of extant pine and spruce lineages occurred relatively recently (Willyard et al. 2007; Lockwood et al. 2013; Ran et al. 2015; Wang & Wang 2014), perhaps facilitated by a cooling climate after the Miocene optimum. Although we did not estimate absolute substitution rates explicitly, the similarity of the *Pinus* and *Picea* relative substitution rates and divergence times together suggests a similar per-generation mutation rate in these two Pinaceae genera (Willyard et al. 2007).

Positive evolution in Picea *plastid genes*?

Positive selection on plastid genes has been widely inferred in angiosperms (e.g. Carbonell-Caballero et al. 2015; Hu et al. 2015) and in some cases potentially adaptive changes have been demonstrated at the protein or phenotype level (Mengistu et al. 2000). Gymnosperm plastomes are less studied. An

bioRxiv preprint doi: https://doi.org/10.1101/097519; this version posted December 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

analysis of *Pinus* plastomes found more limited evidence for positive selection (6/71 genes; Parks et al. 2009), in contrast to the 20% of *Picea* genes apparently affected. Only two genes, *psb*H and *ycf*1, were inferred to have codons with elevated ω in both *Pinus* and *Picea*. To our knowledge, *psb*H has not been widely implicated in positive evolution in angiosperms, and high ω at this gene may represent a unique attribute of Pinaceae. In contrast, *ycf*1 is exceptionally divergent across land plants (Dong et al. 2015) and shows elevated ω in some angiosperm lineages (e.g. Hu et al. 2015).

Consistent with an adaptive explanation, positively selected sites in *Picea* were disproportionately inferred in genes involved in photosynthesis (Table 3). However, non-synonymous mutations may also become fixed through non-adaptive processes. Small effective population sizes, rare recombination, and strong mutational bias are characteristic of plastomes and can facilitate the fixation of slightly deleterious alleles (Kimura 1983; Felsenstein 1974; Moran 1996). On the other hand, plastids have highly efficient DNA repair systems, which could be expected to purge deleterious mutations (Day & Madesis 2007). Site-directed mutagenesis and in vitro characterization of mutant proteins (e.g. Lan et al. 2013) could help clarify the role of adaptive and non-adaptive processes in shaping plastid substitution rates.

Reticulate evolution and chimeric plastomes

Recombination between divergent haplotypes is often invoked to explain phylogenetic discordance among plastid loci (Wolfe & Randle 2004; Erixon & Oxelman 2008; Bouillé et al. 2011). Although recombination has been demonstrated in somatic hybrids (e.g. Medgyesy et al. 1985), evidence from natural populations is rare and based on limited data (Marshall et al. 2001; Huang et al. 2001; D'Alelio & Ruggiero 2015). Using hypothesis and data-driven analyses, we found evidence for at least one interspecific recombination event in *Picea* plastomes. Although a robust, coalescent species tree is not yet available, but topologies inferred for the large plastome fragment closely match those from mitochondrial loci (Ran et al. 2015) and nuclear loci (Zou et al. 2016; J-Q Liu, unpublished data), which suggests *P.*(*glauca*) and *P.*(*abies*) are related by vertical descent. In contrast, the F2 and potentially also the F1 fragment of the *P.*(*abies*) plastome was inherited laterally from a member of *P.(mariana)*, most likely *P. jezoensis* or an extinct relative. Formation of chimeric plastomes via interspecific recombination would require at a minimum: 1) hybridization, 2) temporary breakdown of uniparental inheritance, and 3) plastid fusion and intermolecular recombination.

Like many plant taxa, extant *Picea* are widely interfertile and many members of *P.(abies*) successfully hybridize with *P.*(*mariana*) in controlled crosses (Wright 1955). In addition, many parapatric spruces hybridize readily (e.g. Jaramillo-Correa et al. 2005; Du et al. 2011), including distantly-related species (Hamilton & Aitken 2013). Comparison of phylogenies inferred from mitochondrial and plastid loci, which are differentially inherited in gymnosperms, also support widespread introgression throughout the genus (Bouillé et al. 2011; Ran et al. 2015). In particular, *P. jezoensis* has a *P.*(*mariana*)-type plastome but more closely resembles *P.*(*abies*) at mitochondrial and

bioRxiv preprint doi: https://doi.org/10.1101/097519; this version posted December 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

some nuclear loci (Ran et al. 2015; Lockwood et al. 2013). Biogeographic reconstructions suggest the most recent common ancestor of *P.*(*abies*) emerged in eastern Asia in the Pleistocene, after the dispersal of the *P. jezoensis* lineage across the Bering Land Bridge from a North American ancestor (Ran et al. 2015). Fossil and genetic evidence suggest the continuous presence of *P. jezoensis* and its extinct relative *P. hondoensis* in northeastern Russia through the quaternary glaciations (Aizawa et al. 2007). Introgression between a *P. jezoensis*-like species and the most recent common ancestor of *P.*(*abies*) may have occurred here, prior to the clade's diversification and colonization of Eurasia.

Hybrids can potentially inherit plastids from two different species. Mechanisms maintaining uniparental inheritance are varied (Bock 2007), but hybridization or interpopulation crosses have been linked to a breakdown of these processes in *Picea* (Sutton et al. 1991), *Larix* (Szmidt et al. 1987), and some angiosperms (e.g. Hansen et al. 2007; Ellis et al. 2008). In *Passiflora* and *Helianthus,* rates of plastid leakage vary by family or species but biparental inheritance occurs in ~4-2% of offspring overall, respectively. As biparental inheritance is not uncommon on an evolutionary timescale, fusion of the two parental plastids is likely the limiting step preventing more widespread interspecific recombination (Day & Madesis 2007). To our knowledge, plastid fusion has not been observed directly in natural populations but has been demonstrated experimentally in somatic hybrids (e.g. Medgyesy et al. 1985). Once fused, recombination should occur readily between highly similar genomes because it is the dominant form of plastid DNA repair and may also play a role in plastome replication (Day & Madesis 2007). In our analysis, both hypothesis and data-driven tests supported recombination near dispersed repeats intervening the structural fragments, which are likely substrates for homologous recombination (Day & Madesis 2007). Sorting of plastomes due to the strong genetic bottleneck brought on at every cell-division could result in a mature tree possessing only copies of the recombinant plastome genome. Spread of the recombinant plastome through the ancestor of *P.*(*abies*) may have proceeded through entirely neutral processes, such as allele surfing during range expansion (Excoffier & Ray 2008).

Implications for low-level taxonomic studies

Plastomes effectively provide a single gene estimate of the species phylogeny. Incomplete lineage sorting can cause the plastome to markedly vary from the true species tree (Doyle 1992), and introgression can cause further discordance between plastome and species history. In the absence of a robust species tree estimate, we caution against interpreting plastome phylogenies as an adequate proxy for species history. Nevertheless, resolving relationships among closely related species using whole plastomes is a common objective in plant systematics. In this respect, whole plastomes yielded high resolution within most clades and shallow relationships were largely unaffected by interspecific recombination. Multiple accessions of the same nominal species generally formed monophyletic groups within the *P.*(*glauca*) and *P.*(*mariana*) clades, but plastomes were shared among species within *P.*(*abies*), *P.*(*likiangensis*) and *P.*(*chihuahuana*). Shared haplotypes may arise from incomplete lineage

bioRxiv preprint doi: https://doi.org/10.1101/097519; this version posted December 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

sorting, hybridization, cryptic and poorly-delimited species, or misidentified material. Our data provide a framework for future population-level plastome sequencing, which will help elucidate the contribution of each factor to plastome sharing in *Picea*.

Conclusions

We used whole plastome sequences and comprehensive taxon sampling to characterize sequence divergence, positive selection, and evolutionary history in *Picea.* Most significantly, we found chimeric plastomes formed by ancient introgression and recombination between divergent clades. Chimeric plastomes explain the conflicting results of earlier plastid-based studies in *Picea*: phylogenies inferred from different loci alternately reflect vertical descent and reticulate evolution. As whole plastome sequences become increasingly available, we suggest future studies utilize structural features, such as large repeats, to develop phylogenetic hypotheses to test for recombination. Given that hybridization occurs in ~15% of analyzed plant genera (Ellstrand et al. 1996; Whitney et al. 2010) and the potential for leaky plastid inheritance in up to 17% (Corriveau & Coleman 1988; Zhang et al. 2003), interspecific plastome recombination could be a relatively widespread phenomenon. Together, our results add to an emerging portrait of the plastome as a highly dynamic genome, even on short evolutionary time scales, which is likely to have a pronounced effect on the interpretation of plant evolutionary studies.

Materials and Methods

Plant material

A total of 61 representatives from the 35 commonly recognized *Picea* species (Farjon 1991; Eckenwalder 2009) were sourced from natural populations and cultivated collections. Multiple accessions were sequenced for broadly distributed species or those with spatially-disjunct or fragmented ranges. For cultivated trees, we preferred individuals obtained as seed or seedlings from natural populations and selected wild accessions outside of known hybrid zones whenever possible (see Supplementary Table S1 for provenances and voucher numbers). Additional complete chloroplast genomes for *P. abies* (Nystedt et al. 2013), *P. glauca* (Jackman et al. 2015), *P. morrisonicola* (Wu et al. 2011), and *P. jezoensis* (Yang et al. 2016) were downloaded from Genbank (Supplementary Table S1).

Genome sequencing and assembly

Total genomic DNA was isolated from leaf material using the Macherey-Nagel NucleoSpin® Plant II kit according to the manufacturer instructions (Düren, Germany). Paired-end libraries with ~350 bp inserts were constructed for 77 accessions for this and another study and sequenced on Illumina HiSeq 2000 and 2500 lanes using 2×101 bp and 2×126 bp chemistry, respectively, at the National Genomics Infrastructure hosted at Science for Life Laboratory (SciLifeLab, Stockholm, Sweden). Paired-end libraries with ~350 bp inserts for 17 accessions for this and another study were multiplexed on a single Illumina HiSeq 4000 lane at the University of Texas at Austin (Austin, Texas, USA) using 2×150 bp

bioRxiv preprint doi: https://doi.org/10.1101/097519; this version posted December 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

chemistry. Finally, an entire lane of 2×101 bp reads from the Illumina MiSeq platform was available for one accession. We developed a *de novo* assembly strategy by first comparing assemblies produced by SPAdes v. 3.7.0 (Bankevich et al. 2012; Prjibelski et al. 2014) using either whole-genome reads or reads pre-filtered for similarity to the plastome. For read filtering, we used local alignment in bowtie2 v. 2.2.7 (Langmead & Salzberg 2012) using the *P. abies* plastome as a reference and tested both the 'very fast' and 'very sensitive' presets. The most contiguous assemblies with the highest depth of coverage were obtained by using pre-filtered reads in and results were nearly identical for both local alignment presets (results not shown). Therefore, we used 'very fast' local alignment followed by assembly using the 'careful' parameter and default error correction and kmer selection. Sequencing methods, assembly metrics, and Genbank accession numbers for each accession reported in Supplementary Table S1.

Annotation and alignment

We tested for structural rearrangements in *de novo* scaffolds of each accession versus the *P. abies* reference genome using nucmer, which were visualized as dot plots with mummerplot as contained within the software package MUMmer v. 3.0 (Kurtz et al. 2004). The scaffolds of each accession were aligned against the *P. abies* reference and pseudomolecules were generated in Geneious v. 9.0.4 for further analysis. Multiple sequence alignment of pseudomolecules was conducted in MAFFT v. 7.245 (Katoh & Standley 2013) using the fast Fourier transformation approximation option, a partition size of 1,000 and three iterative refinements. Sequences were annotated according to the *P. abies* reference genome using the BLAST-like 'transfer annotations' tool in Geneious using an identity cut-off of 60%. All annotations were manually curated to ensure accurate identification of genes, pseudogenes, and reading frames. Plastome maps were created with CIRCOS (Krzywinski et al. 2009).

Sequence polymorphism and divergence

Nucleotide diversity (π) was estimated using the R package PopGenome v. 2.1.6 (Pfeifer et al. 2014) by genomic context and in 350 bp sliding windows. Standard errors were estimated using 500 bootstrap replicates. Average synonymous (*d*S) and non-synonymous (*d*N) substitution rates were estimated per-gene using the maximum likelihood method (Goldman & Yang 1994) implemented in the codeml program in PAML v. 4.9 (Yang 2007). We inferred natural selection at the codon-level from the ratio of *d*N to *d*S (ω) using nested random site models (M1a & M2a; Wong et al. 2004; Yang et al. 2005) as implemented in codeml. The maximum-likelihood tree inferred from concatenated protein-coding exons was used as the input tree with relationships between the five major clades and any branches with < 70% bootstrap support constrained to polytomies. Both copies of *ycf*12 and *psb*I were excluded because they are located within large repeats (white arrows in Fig. 1) and were not uniformly recovered across all accessions.

bioRxiv preprint doi: https://doi.org/10.1101/097519; this version posted December 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Phylogenetic analyses

Sequence alignments were prepared for phylogenetic analysis by stripping sites with more than 20% missing data and removing tandem repeats using Phobos v. 3.3.12 (<http://www.ruhr-uni-bochum.de/spezzoo/cm/cm_phobos.htm>). We used PartitionFinder v. 1.1.1 to identify best-fit partitioning schemes of nucleotide evolution (Lanfear et al. 2012, 2014). We defined an initial model giving each protein-coding gene and codon position a partition, and tRNAs, rRNAs, introns were each assigned a single partition, respectively. We constrained the model of nucleotide evolution to GTR+Γ and used the linked branch lengths model. Partition searches were performed using the 'rcluster' clustering algorithm with the default percentage. Optimal partitioning schemes identified using the whole genome data were subsequently applied to each of the five data subsets: 1) protein-coding exons, 2) concatenated intergenic spaces, and 3-5) the three structural fragments (F1, F2, and large) of the *Picea* plastome. Phylogenies were inferred from the partitioned datasets in RAxML v. 8.2.4 (Stamatakis 2014) using rapid bootstrapping under the autoMRE convergence criterion and were summarized on an extended majority rule consensus tree (Aberer et al. 2010). Trees are rooted by *P. breweriana* for display, which has been inferred to be the sister lineage to all other extant spruces in previous studies (Bouillé et al. 2011; Lockwood et al. 2013; Ran et al. 2015).

We tested for phylogenetic discordance using character and tree-based methods. The character-based ILD test was performed using PAUP* v. 4.0a149 (Swofford 2002) using 100 replicates with a heuristic starting seed, random-stepwise additions, and nearest-neighbor interchange branch swapping with the search limited to 100,000 re-arrangements for computational feasibility. We also implemented permutation tests based on tree QD (Bryant et al. 2000) and RF (Robinson & Foulds 1981) distances to test if phylogenies estimated from different *a priori* subsets are more different than expected by chance. To estimate tree distances under the null hypothesis, we calculated QD and RF values between 10,000 trees estimated from randomly sampled plastome sites. Sets of random sites were selected to be comparable in size to the subsets of interest. Then, we calculated RF and QD distances between 10,000 randomly selected pairs of trees sampled from the bootstrap replicates of the *a priori* subsets. We implemented the permutation test using the R package 'perm' (Fay & Shaw 2010) using 2,000 replications and estimated *p*-values using the default algorithm. QDs were calculated using the program rqtdist v. 1.0 (Sand et al. 2013) and RF distances in the R packages 'phangorn' (Schliep 2011) and 'ape' (Paradis et al. 2004).

Recombination tests

Unguided tests for recombination were conducted in the RDP4 package (Martin et al. 2015) using the RDP (Martin & Rybicki 2000), GENECOV, BootScan (Salminen et al. 1995), MaxChi (Maynard Smith 1992), Chimaera (Posada & Crandall 2001), SiScan (Padidam et al. 1999), and 3seq (Boni et al. 2007) methods. After identifying a recombination event, RDP4 implements a hidden Markov model to estimate breakpoint positions. Then, RDP4 identifies the probable recombinant sequence using the

bioRxiv preprint doi: https://doi.org/10.1101/097519; this version posted December 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

PHYLPRO (Weiller 1998), VISRD (Lemey et al. 2009), and EEEP (Beiko & Hamilton 2006) methods. Bonferrroni corrections were applied to set the family-wise error rate to 0.05. We tested the whole-plastome alignment for recombination and an alignment stripped of gaps and poorly aligned regions but the results did not differ. In addition, altering window sizes and indel processing options did not alter the results so default values were used for the final runs. Szmidt et al. 1987

Acknowledgements

We are grateful to Daniel Luscombe (Bedgebury National Pinetum), Peter Brownless (Royal Botanic Garden Edinburgh) Charles Keith (Charles Keith Arboretum), Erik Dahl Kjær and Albin Lobo (University of Copenhagen, Denmark), Piotr Banaszczak (Rogów Arboretum), Johnny Schimmel (Arboretum Norr), Dan Crowley (Westonbirt, The National Arboretum), László Csiba and Rhinaixa Duque-Thüs (Royal Botanic Gardens, Kew), Jaakko Saarinen (Mustila Arboretum), and Jianquan Liu (Lanzhou University and Sichuan University) for contributing samples to this study. We acknowledge support from Science for Life Laboratory and the National Genomics Infrastructure, Sweden for providing assistance in massive parallel sequencing**.** Computations were performed on resources provided by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under Project SNIC 2015-6-67. Douglas Scofield at UPPMAX provided bioinformatics support. This study was supported by the Trees and Crops for the Future (TC4F) program.

bioRxiv preprint doi: https://doi.org/10.1101/097519; this version posted December 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

List of figures

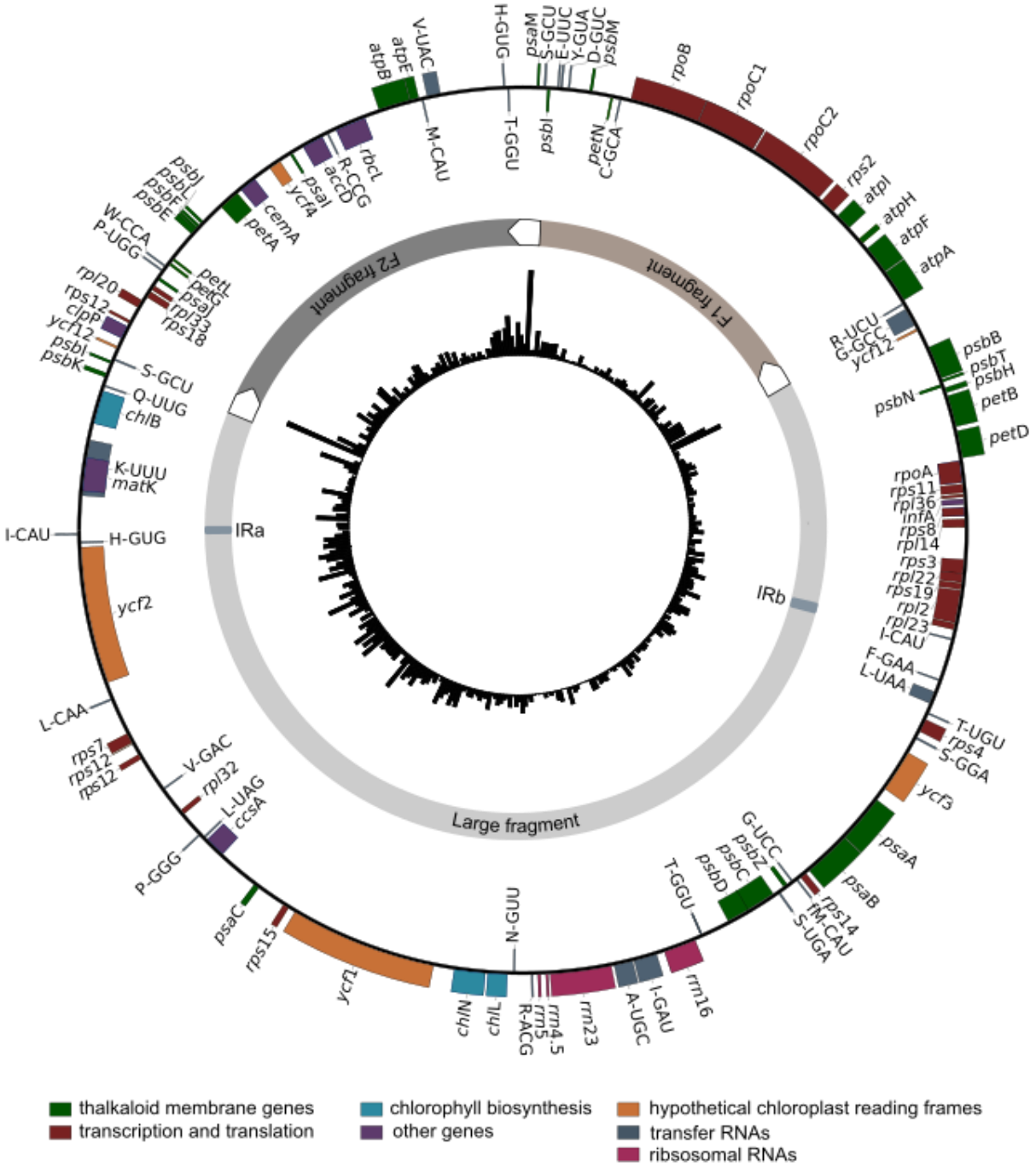


Figure 1. Consensus map of the *Picea* plastome. The outermost ring displays genes colored by functional group: genes on the inside and outside of the circle are transcribed in the clockwise and counterclockwise directions, respectively. The second circle displays the boundaries of the three major fragments (F1, F2 and large) of the plastome, which are intervened by large repeats as indicated by white arrows (not to scale). The innermost circle depicts average pairwise nucleotide diversity (π) estimated in 350 bp moving windows.

bioRxiv preprint doi: https://doi.org/10.1101/097519; this version posted December 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

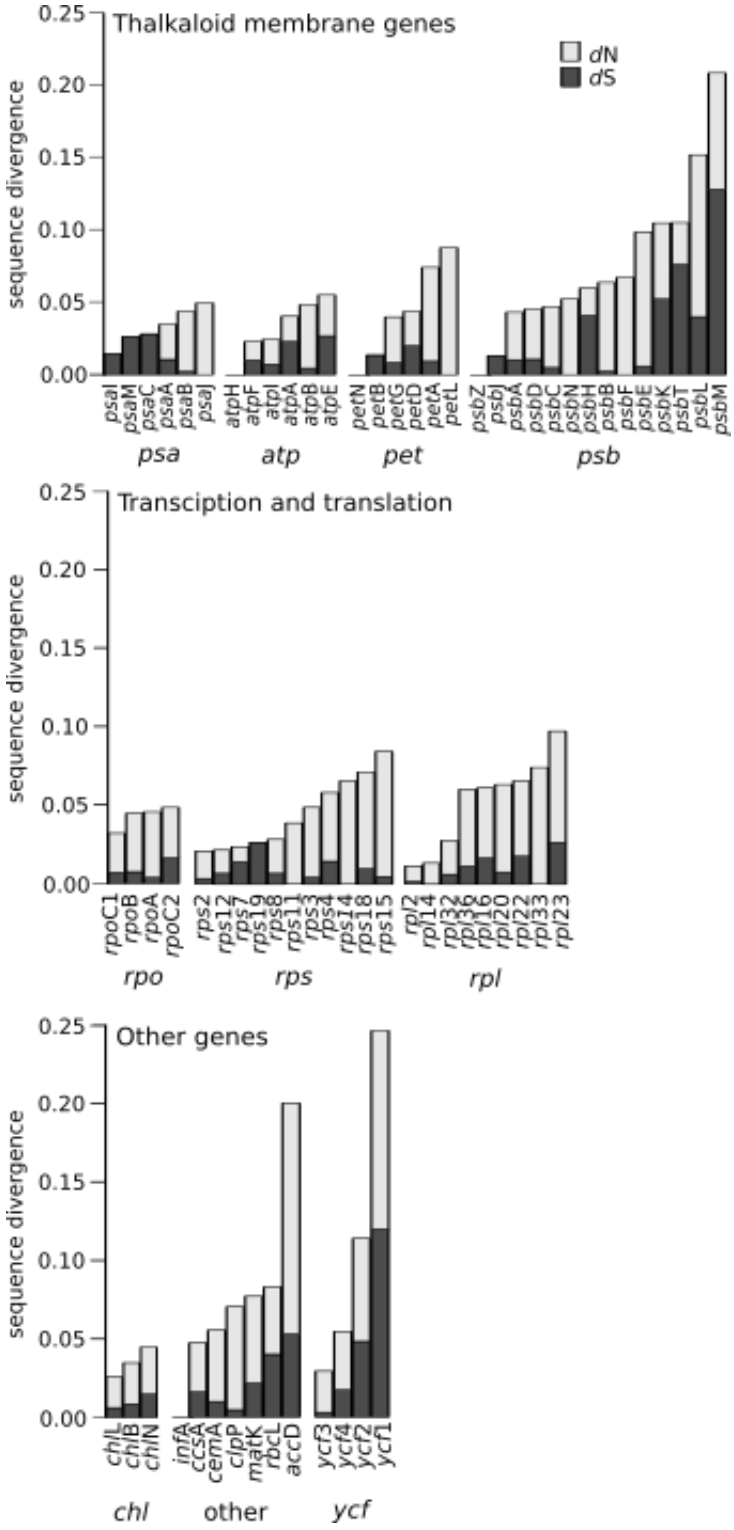
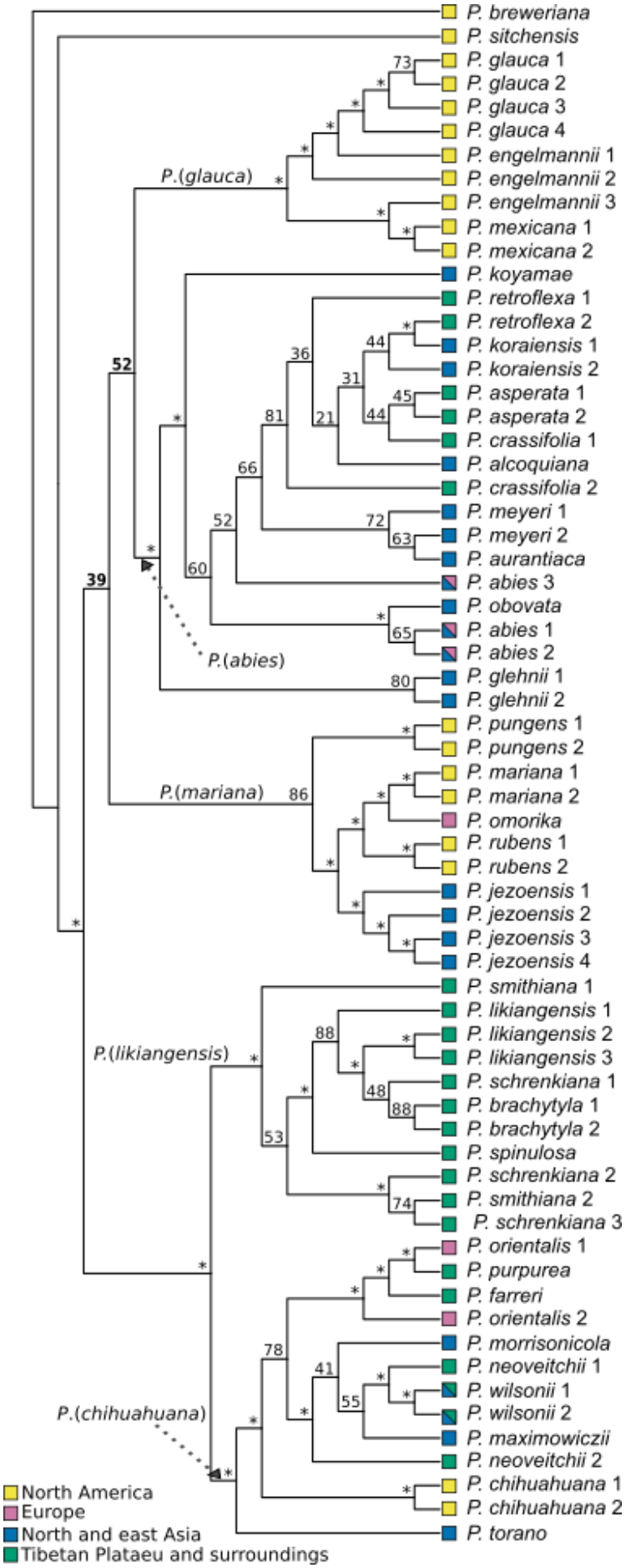


Figure 2. Synonymous (*d*S) and non-synonymous (*d*N) substitution rates for *Picea* plastome genes grouped by complex and function. Rates were using the maximum likelihood method implemented in PAML (Yang 2007). Substitution rates were significantly ($p < 0.05$) elevated in the photosystem II and hypothetical open reading frame genes. Per-gene estimates are reported in Supplementary Table S2.

bioRxiv preprint doi: https://doi.org/10.1101/097519; this version posted December 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/097519; this version posted December 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 3. Maximum likelihood analysis of the whole plastome alignment of 65 *Picea* accessions provides strong support for many shallow nodes but the relationships between *P.*(*glauca*), *P.*(*abies*), and *P.*(*mariana*) are unresolved. The majority rules extended consensus tree is presented and node labels denote the proportion of bootstrap replicate trees supporting the bipartition. Nodes with ≥ 90% support are indicated by an asterisk. Colored squares denote the geographic distribution of the species. Numbers after species names correspond to the accession information in Supplementary Table S1.

bioRxiv preprint doi: https://doi.org/10.1101/097519; this version posted December 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

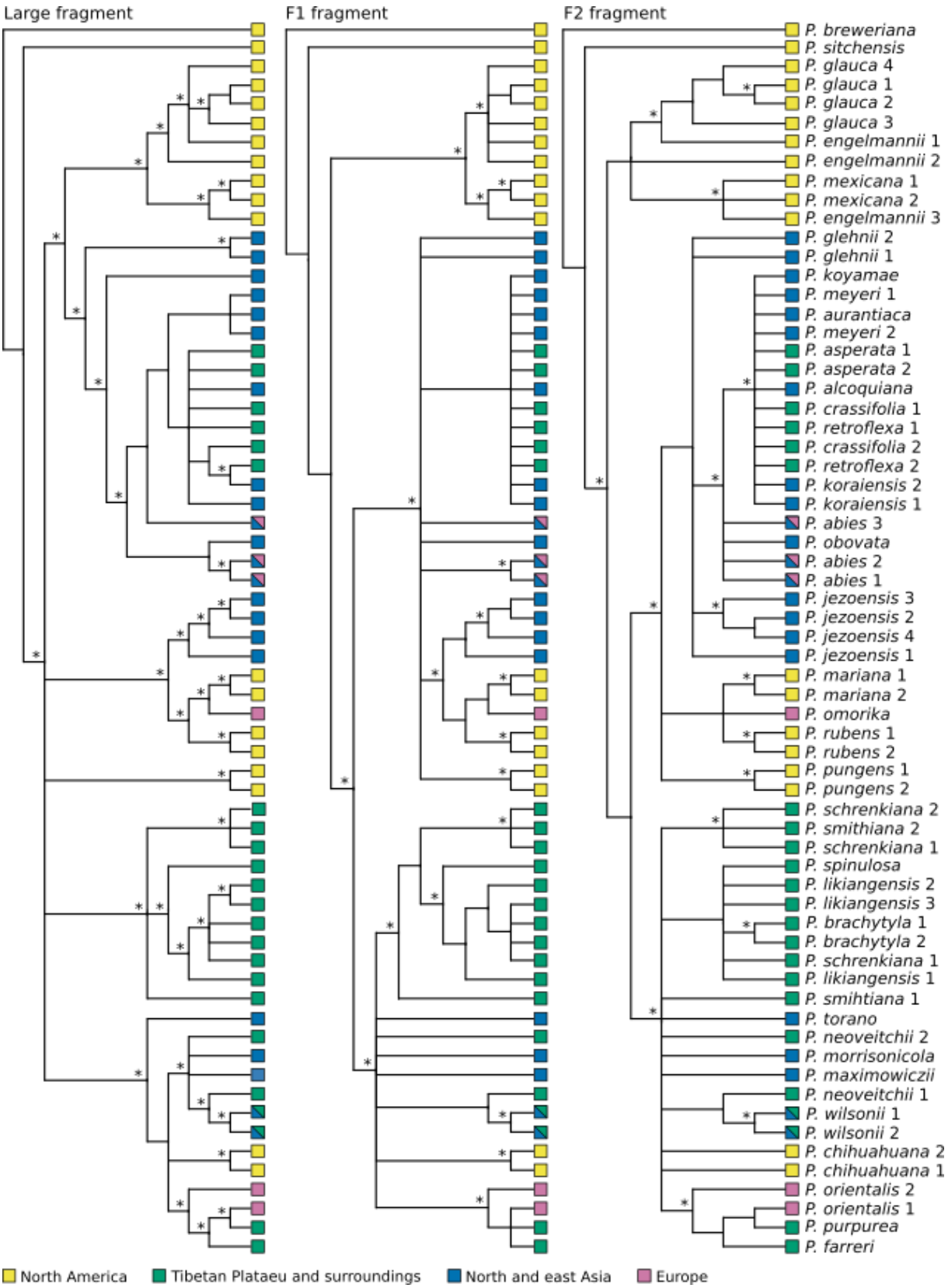


Figure 4. Phylogenies inferred from the structural subunits of the *Picea* plastome are strongly discordant according to topology and character based incongruence tests. While the large fragment supports a sister relationship between *P.*(*glauca*) and *P.*(*abies*), the F2 and F1 fragments suggest a closer relationship between *P.*(*abies*) and *P.*(*mariana*). Majority rules extended consensus trees are

bioRxiv preprint doi: https://doi.org/10.1101/097519; this version posted December 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

presented. Nodes with ≥ 90% support are indicated by an asterisk and those with < 70% support are collapsed. Colored squares denote the geographic distribution of the species. Numbers after species names correspond to the accession information in Supplementary Table S1 and uncollapsed trees are presented in Supplementary Fig. S2. \

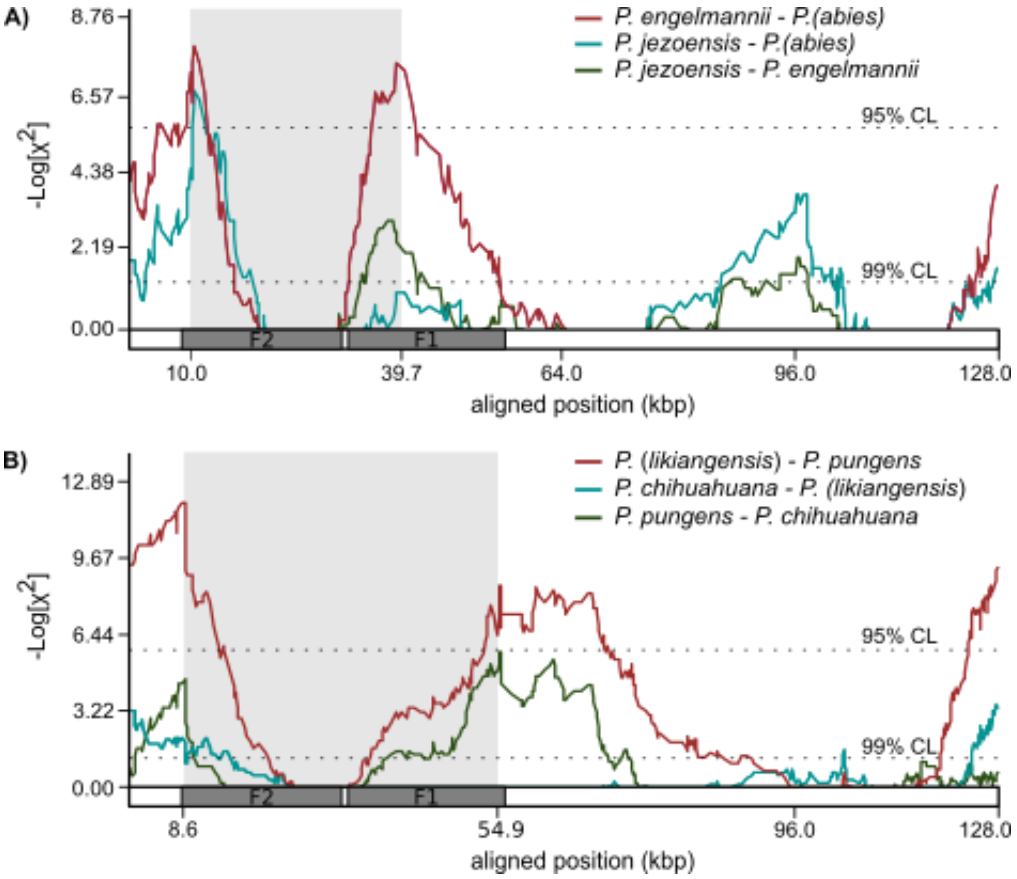


Figure 5. Unguided tests for recombination in the *Picea* whole plastome alignment, as illustrated by the MaxChi method (Maynard Smith 1992). A) All seven recombination tests supported a recombinant plastome shared by the entire *P.*(*abies*) clade originating from interspecific recombination with a *P. jezoensis*-like species. On the y-axis, lines represent the 2x2 contingency χ^2 value of the number of polymorphic sites on either side of a given window. Breakpoints are inferred to occur where this value maximized and the recombinant region is shaded in grey. A schematic representing the structural components of the plastome is superimposed on the x-axis. B) Five tests supported a recombinant *P.*(*chihuahuana*) plastome and estimated breakpoints occurred near the F2 and F1 fragment junctions.

bioRxiv preprint doi: https://doi.org/10.1101/097519; this version posted December 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Tables

Table 1. Sequencing and assembly metrics for 61 *Picea* draft plastomes. Standard errors are reported in parentheses.

Sequencing system	**Number of Scaffolds**	**Scaffold N50 (bp)**	**Maximum scaffold size (bp)**	**Mean read depth**
Illumina HiSeq 2000 & *2500* 2x100 bp, 2x125 bp	5.77 (1.76)	56,475 (2,528)	66,112 (13,562)	46.83 (23.32)
Illumina HiSeq 4000 2x150 bp	6.75 (1.35)	22,939 (1,162)	55,847 (6,949)	413.51 (239.18)
Illumina MiSeq 2x100 bp	7	23,241	55,799	47.17

Table 2. Alignment length, number of polymorphic sites, and nucleotide diversity (π) by genomic context estimated from 65 *Picea* plastomes. Standard errors are reported in parentheses.

Context	**Alignment length (bp)**	**Bi-allelic sites**	**π**
Protein-coding exons	61,256	430	0.002 (1.6×10^{-4})
Four-fold synonymous sites	7,720	180	0.0025 (2.9×10^{-4})
tRNAs	2,874	12	0.0015 (1.6×10^{-4})
rRNAs	4,522	13	0.0002 (7.6×10^{-5})
Introns	9,929	197	0.0023 (3.2×10^{-4})
Intergenic spacers	49,373	1,681	0.0048 (1.1×10^{-4})
Whole plastome	127,954	2,333	0.0030 (6.6×10^{-5})

Table 3. Positive selection inferred from the M1a and M2a random site models implemented in PAML. Positively selected sites have ω > 1 with posterior probability ≥ 0.95 according to the Bayes empirical Bayes method. Amino acids variants observed at each putative selected sites are listed in parenthesis. Site number refers to the position in the *Picea* alignment for that protein.

Complex	**Gene**	**-ln(M1a)**	**-ln(M2a)**	**Sites**
ATP synthase	*atpA*	-2210.70	-2182.54	185(P,L), 232(P,S), 253(K,N), 299(S,L), 314(V,L), 427(A,T)
	atpE	-617.32	-578.11	119(P,L)
Photosystem I	*psaC*	-352.21	-342.13	35(K,R)
Photosystem II	*psbA*	-1517.02	-1481.65	155(T,S,A)
	psbD	-1538.38	-1525.49	156(P,S)
	psbH	-346.92	-337.02	16(S,A), 21(R,T)
	psbT	-198.05	-168.65	30(P,L,Q,S)
Large ribosomal subunit	*rpl22*	-683.02	-664.74	7(N,Y,K)
RNA polymerase	*rpoC2*	-5556.69	-5540.32	326(N,K), 870(R,I)
Small ribosomal subunit	*rps19*	-399.85	-388.39	20(K,Q)
Hypothetical open reading frame	*ycf1*	-13094.20	-12574.89	68(A,T), 176(V,A), 502(S,Y), 612(L,I), 619(H,Y), 739(R,K), 959(F,L), 1054(A,S), 1083(L,F), 1139(N,K), 1149(S,A), 1223(F,L), F1306(F,L), 1378(L,F), 1383(S,L), 1507(S,L), 1633(L,F), 1708(D,E), 1709(K,Q), 1790(V,T), 1912(R,E)
	ycf4	-815.00	-799.87	118(L,P)

bioRxiv preprint doi: https://doi.org/10.1101/097519; this version posted December 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 4. Significance of phylogenetic incongruence between plastome subsets according to the incongruence length difference (ILD), quartet distance (QD), and Robinson-Foulds (RF) metric permutation tests. F1, F2 and large fragments refer to the three major structural regions of the *Picea* plastome (see Fig. 1). The F1 and F2 fragments are concatenated in site-specific comparisons.

Comparison	ILD	QD	RF
PCE[1] *vs* intergenic	0.05	0.40	0.46
F1 *vs* Large fragment	0.01	< 0.05	< 0.01
F2 *vs* Large fragment	0.01	< 0.001	< 0.01
F2 *vs* F1 fragment	0.20	0.17	0.23
Large vs F1+F2 intergenic	0.01	< 0.001	< 0.001
Large *vs* F1+F2 PCE	0.01	< 0.001	< 0.001
Large *vs* F1+F2 4D[2]	0.86	< 0.001	< 0.001

1 – Protein-coding exons

2 – Fourfold synonymous sites

Table 5. Bonferonni corrected p-values for the two recombination events detected by unguided tests as implemented in RDP4.

Recombination Test	*P.*(*abies*)	*P.*(*likiangensis*)
RDP	6.71×10^{-09}	2.36×10^{-5}
GENECOV	1.89×10^{-11}	n.s.
BootScan	6.71×10^{-5}	n.s.
MaxChi	1.81×10^{-7}	6.11×10^{-6}
Chimaera	1.83×10^{-4}	6.87×10^{-8}
SiScan	9.46×10^{-11}	1.55×10^{-3}
3Seq	4.67×10^{-4}	2.57×10^{-9}

n.s. – not significant

bioRxiv preprint doi: https://doi.org/10.1101/097519; this version posted December 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary material

Supplementary Table S1. Locality and accession information for all 65 included *Picea* plastomes. Assembly statistics are reported for the 61 *de novo* assemblies.

Supplementary Table S2. Per gene estimates of nucleotide diversity (π), dN, dS, and codon-based maximum likelihood tests for positive selection.

Supplementary Figure S1. Extended majority rules consensus trees estimated from fourfold synonymous sites, intergenic spaces, and protein coding exons from each fragment (F1, F2, and large), respectively.

Supplementary Figure S2. Extended majority rules consensus trees estimated from the entire sequence of the large, F1, and F2, fragments, respectively. Nodes with 100% bootstrap support are denoted with an asterisk.

References

Aberer AJ, Pattengale ND, Stamatakis A. 2010. Parallelized phylogenetic post-analysis on multi-core architectures. J Comput Sci. 1:107–114.

Aizawa M et al. 2007. Phylogeography of a northeast Asian spruce, *Picea jezoensis*, inferred from genetic variation observed in organelle DNA markers. Mol Ecol. 16:3393–3405.

Bankevich A et al. 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 19:455–477.

Beiko RG, Hamilton N. 2006. Phylogenetic identification of lateral genetic transfer events. BMC Evol Biol. 6:15–17.

Birol I et al. 2013. Assembling the 20 Gb white spruce (*Picea glauca*) genome from whole-genome shotgun sequencing data. Bioinformatics. 29:1492–1497.

Bock R. 2007. Structure, function, and inheritance of plastid genomes. In: *Cell and Molecular Biology of Plastids*. Bock, R, editor. Springer Berlin Heidelberg: Berlin, Heidelberg pp. 29–63.

Boni MF, Posada D, Feldman MW. 2007. An exact nonparametric method for inferring mosaic structure in sequence triplets. Genetics. 176:1035–1047.

Bouillé M, Senneville S, Bousquet J. 2011. Discordant mtDNA and cpDNA phylogenies indicate geographic speciation and reticulation as driving factors for the diversification of the genus *Picea*. Tree Gene Genomes. 7:469–484.

Bryant D, Tsang J, Jearney PE, Li M. 2000. Computing the quartet distance between evolutionary

bioRxiv preprint doi: https://doi.org/10.1101/097519; this version posted December 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

trees. In: Proceedings of the Eleventh Annual ACM-SIAM Symposium on Discrete Algorithms. ACM Press: New York City pp. 285–286.

Carbonell-Caballero J et al. 2015. A phylogenetic analysis of 34 chloroplast genomes elucidates the relationships between wild and domestic species within the genus *Citrus*. Mol Biol Evol. 32:2015–35.

Corriveau, JL, Coleman AW. 1988. Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperms. Am J Bot. 75:1443–1458.
Crosby K, Smith DR. 2012. Does the mode of plastid inheritance influence plastid genome architecture? PLoS One. 7:1–8.

D'Alelio D, Ruggiero MV. 2015. Interspecific plastidial recombination in the diatom genus *Pseudo-nitzschia*. J Phycol. 51:1024–1028.

Day A, Madesis P. 2007. DNA replication, recombination, and repair in plastids. In: *Cell and Molecular Biology of Plastids*. Bock, R, editor. Springer Berlin Heidelberg: Berlin, Heidelberg pp. 65–119.

Dong W et al. 2015. *ycf*1, the most promising plastid DNA barcode of land plants. Sci Rep. 5:8348.

Doyle JJ. 1992. Gene trees and species trees: molecular systematics as one-character taxonomy. Syst Bot. 17:144–163.

Du FK et al. 2011. Direction and extent of organelle DNA introgression between two spruce species in the Qinghai-Tibetan Plateau. New Phytol. 192:1024–1033.

Eckenwalder JE. 2009. *Conifers of the world: the complete reference*. Timber Press, Inc.: Portland, OR.

Ellis JR, Bentley KE, McCauley DE. 2008. Detection of rare paternal chloroplast inheritance in controlled crosses of the endangered sunflower *Helianthus verticillatus*. Heredity. 100:574–580.

Ellstrand NC, Whitkus R, Rieseberg LH. 1996. Distribution of spontaneous plant hybrids. Proc Natl Acad Sci U. S. A. 93:5090–5093.

Erixon P, Oxelman B. 2008. Reticulate or tree-like chloroplast DNA evolution in *Sileneae* (Caryophyllaceae)? Mol Phylogenet Evol. 48:313–325.

Excoffier L, Ray N. 2008. Surfing during population expansions promotes genetic revolutions and structuration. Trends Ecol Evol. 23:347–351.

Farjon A. 1991. *Pinaceae: Drawings and Descriptions of the Genera :* Abies, Cedrus, Pseudolarix, Keteleeria, Nothotsuga, Tsuga, Cathaya, Pseudotsuga, Larix *and* Picea. Koeltz Scientific Books:

bioRxiv preprint doi: https://doi.org/10.1101/097519; this version posted December 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Königstein.

Fay MP, Shaw P a. 2010. Exact and asymptotic weighted logrank tests for interval censored data: the interval R package. J Stat Softw. 36:1–34.

Felsenstein J. 1974. The evolutionary advantage of recombination. Genetics. 78:737–765.

Goldman N, Yang Z. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol. 11:725–736.

Hamilton JA, Aitken SN. 2013. Genetic and morphological structure of a spruce hybrid (*Picea sitchensis* x *P. glauca*) zone along a climatic gradient. Am J Bot. 100:1651–1662.

Hansen AK, Escobar LK, Gilbert LE, Jansen RK. 2007. Paternal, maternal, and biparental inheritance of the chloroplast genome in *Passiflora* (Passifloraceae): Implications for phylogenetic studies. Am J Bot. 94:42–46.

Hu S et al. 2015. Plastome organization and evolution of chloroplast genes in *Cardamine* species adapted to contrasting habitats. BMC Genomics. 16:306.

Huang S, Chiang YC, Schaal BA, Chou CH, Chiang TY. 2001. Organelle DNA phylogeography of *Cycas taitungensis*, a relict species in Taiwan. Mol Ecol. 10:2669–2681.

Jackman SD et al. 2015. Organellar genomes of white spruce (*Picea glauca*): assembly and annotation. Genome Biol Evol. 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.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol. 30:772–780.

Kimura M. 1983. *The Neutral Theory of Molecular Evolution:* Cambridge University Press: Cambridge.

Klymiuk AA, Stockey RA. 2012. A lower cretaceous (Valanginian) seed cone provides the earliest fossil record for *Picea* (Pinaceae). Am J Bot. 99:1069–1082.

Krzywinski MI et al. 2009. Circos: An information aesthetic for comparative genomics. Genome Res. 19:1639–1645.

Kurtz S et al. 2004. Versatile and open software for comparing large genomes. Genome Biol. 5:R12.

Lan T, Wang XR, Zeng QY. 2013. Structural and functional evolution of positively selected sites in

bioRxiv preprint doi: https://doi.org/10.1101/097519; this version posted December 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

pine glutathione S-transferase enzyme family. J Biol Chem. 288:24441–24451.

Lanfear R, Calcott B, Ho SYW, Guindon S. 2012. PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol Biol Evol. 29:1695–1701.

Lanfear R, Calcott B, Kainer D, Mayer C, Stamatakis A. 2014. Selecting optimal partitioning schemes for phylogenomic datasets. BMC Evol. Biol. 14:82.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods. 9:357–359.

Lemey P, Lott M, Martin DP, Moulton V. 2009. Identifying recombinants in human and primate immunodeficiency virus sequence alignments using quartet scanning. BMC Bioinformatics. 10:1–18.

Lockwood JD et al. 2013. A new phylogeny for the genus *Picea* from plastid, mitochondrial, and nuclear sequences. Mol Phylogenet Evol. 69:717–727.

Marshall HD, Newton C, Ritland K. 2001. Sequence-repeat polymorphisms exhibit the signature of recombination in lodgepole pine chloroplast DNA. Mol Biol Evol. 18:2136–2138.

Martin D, Rybicki E. 2000. RDP: detection of recombination amongst aligned sequences. Bioinformatics. 16:562–563.

Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. 2015. RDP4: Detection and analysis of recombination patterns in virus genomes. Virus Evol. 1:1–5.

Maynard Smith J. 1992. Analyzing the mosaic structure of genes. J Mol Evol. 34:126–129.

Medgyesy P, Fejes E, Maliga P. 1985. Interspecific chloroplast recombination in a *Nicotiana* somatic hybrid. Proc Natl Acad Sci U. S. A. 82:6960–6964.

Mengistu LW, Mueller-Warrant GW, Liston A, Barker RE. 2000. psbA Mutation (valine 219 to isoleucine) in *Poa annua* resistant to metribuzin and diuron. Pest Manag Sci. 56:209–217.
Moran NA. 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proc Natl Acad Sci U. S. A. 93:2873–2878.

Mosca E et al. 2012. Contrasting patterns of nucleotide diversity for four conifers of Alpine European forests. Evol Appl. 5:762–775.

Neale DB, Sederoff RR. 1989. Paternal inheritance of chloroplast DNA and maternal inheritance of mitochondrial DNA in loblolly pine. Theor Appl Genet. 77:212–216.

Nystedt B et al. 2013. The Norway spruce genome sequence and conifer genome evolution. Nature.

bioRxiv preprint doi: https://doi.org/10.1101/097519; this version posted December 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

497:579–84.

Padidam M, Sawyer S, Fauquet CM. 1999. Possible emergence of new geminiviruses by frequent recombination. Virology. 265:218–225.

Paradis E, Claude J, Strimmer K. 2004. APE: Analyses of phylogenetics and evolution in R language. Bioinformatics. 20:289–290.

Parks M, Cronn R, Liston A. 2009. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. BMC Biol. 7:84.

Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ. 2014. PopGenome: An efficient swiss army knife for population genomic analyses in R. Mol Biol Evol. 31:1929–1936.

Planet PJ. 2006. Tree disagreement: Measuring and testing incongruence in phylogenies. J Biomed Inform. 39:86–102.

Posada D, Crandall KA. 2001. Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. Proc Natl Acad Sci U. S. A. 98:13757–13762.

Prjibelski AD et al. 2014. ExSPAnder: A universal repeat resolver for DNA fragment assembly. Bioinformatics. 30:293–301.

Ran JH, Shen TT, Liu WJ, Wang PP, Wang XQ. 2015. Mitochondrial introgression and complex biogeographic history of the genus *Picea.* Mol Phylogenet Evol. 93:63–76.

Robinson DF, Foulds LR. 1981. Comparison of phylogenetic trees. Math Biosci. 53:131–147.

Ruhsam M et al. 2015. Does complete plastid genome sequencing improve species discrimination and phylogenetic resolution in *Araucaria*? Mol Ecol.Resour. 15:1067–1078.

Salminen MO, Carr JK, Burke DS, McCutchan FE. 1995. Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. AIDS Res Hum Retroviruses. 11:1423–1425.

Sand A et al. 2013. Algorithms for computing the triplet and quartet distances for binary and general trees. Biology (Basel). 2:1189–209.

Schliep KP. 2011. phangorn: Phylogenetic analysis in R. Bioinformatics. 27:592–593.

Sloan DB, Alverson AJ, Wu M, Palmer JD, Taylor DR. 2012. Recent acceleration of plastid sequence and structural evolution coincides with extreme mitochondrial divergence in the angiosperm genus *Silene*. Genome Biol Evol. 4:294–306.

bioRxiv preprint doi: https://doi.org/10.1101/097519; this version posted December 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30: 1312-1313.

Strauss SH, Palmer JD, Howe GT, Doerksen a H. 1988. Chloroplast genomes of two conifers lack a large inverted repeat and are extensively rearranged. Proc Natl Acad Sci U. S. A. 85:3898–3902.

Sutton BCS, Flanagan D., El-Kassaby YA. 1991. A simple and rapid method for estimating representation of species in spruce seedlots using chloroplast DNA restriction fragment length polymorphism. Silvae Genet. 40:119–123.

Swofford DL. 2002. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Assoc. Sunderland, Massachusetts.

Szmidt AE, Aldén T, Hällgren JE. 1987. Paternal inheritance of chloroplast DNA in *Larix*. Plant Mol Biol. 9:59–64.

Wang B, Wang X-R. 2014. Mitochondrial DNA capture and divergence in *Pinus* provide new insights into the evolution of the genus. Mol Phylogenet Evol. 80:20–30.

Wang X-Q, Tank DC, Sang T. 2000. Phylogeny and divergence times in Pinaceae: evidence from three genomes. Mol Biol Evol. 17:773–781.

Weiller GF. 1998. Phylogenetic profiles: a graphical method for detecting genetic recombinations in homologous sequences. Mol Biol Evol. 15:326–335.

Whitney KD, Ahern JR, Campbell LG, Albert LP, King MS. 2010. Patterns of hybridization in plants. Plant Ecol Evol Syst. 12:175–182.

Whittall JB et al. 2010. Finding a (pine) needle in a haystack: Chloroplast genome sequence divergence in rare and widespread pines. Mol Ecol. 19:100–114.

Willyard A, Syring J, Gernandt DS, Liston A, Cronn R. 2007. Fossil calibration of molecular divergence infers a moderate mutation rate and recent radiations for *Pinus*. Mol Biol Evol. 24:90–101.

Wolfe AD, Randle CP. 2004. Recombination, heteroplasmy, haplotype polymorphism, and paralogy in plastid genes: Implications for plant molecular systematics. Syst Bot. 29:1011–1020.

Wolfe KH, Li W-H, Sharp PM. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci U. S. A. 84:9054–9058.

Wong WSW, Yang Z, Goldman N, Nielsen R. 2004. Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. Genetics. 168:1041–1051.

bioRxiv preprint doi: https://doi.org/10.1101/097519; this version posted December 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Wright JW. 1955. Species crossability in spruce in relation to distribution and taxonomy. For Sci. 1:319–349.

Wu CS, Lin CP, Hsu CY, Wang RJ, Chaw SM. 2011. Comparative chloroplast genomes of Pinaceae: Insights into the mechanism of diversified genomic organizations. Genome Biol Evol. 3:309–319.

Yang JC et al. 2016. The complete plastid genome sequence of *Picea jezoensis* (Pinaceae: Piceoideae). Mitochondrial DNA Part A. 27:3761–3763.

Yang Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24:1586–1591.
Yang Z, Wong WSW, Nielsen R. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. Mol Biol Evol. 22:1107–1118.

Zhang Q, Liu Y, Sodmergen. 2003. Examination of the cytoplasmic DNA in male reproductive cells to determine the potential for cytoplasmic inheritance in 295 angiosperm species. Plant Cell Physiol. 44:941–951.

Zou J et al. 2016. DNA barcoding of recently diversified tree species: a case study on spruces based on 20 DNA fragments from three different genomes. Trees. 30:959–969.