Interspecific Plastome Recombination Reflects Ancient Reticulate Evolution in Picea (Pinaceae)

Alexis R. Sullivan, Bastian Schiffthaler, Stacey Lee Thompson, Nathaniel R. Street, and Xiao-Ru Wang

1Department of Ecology and Environmental Science, Umeå Plant Science Center, Umeå University, Umeå, Sweden
2Department of Plant Physiology, Umeå Plant Science Center, Umeå University, Umeå, Sweden
3Department of Biology, Dalhousie University, Halifax, NS, Canada

*Corresponding author: E-mail: lxslivn@gmail.com.
Associate editor: Song Ge

Abstract

Plastid sequences are a cornerstone in plant systematic studies and 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 tree species, to test for deviations from canonical plastome evolution. Using complementary hypothesis and data-driven tests, we found evidence for chimeric plastomes generated by interspecific hybridization and recombination in the clade comprising Norway spruce (P. abies) and 10 other species. Support for interspecific recombination remained after controlling for sequence saturation, positive selection, and potential alignment artifacts. These results reconcile previous conflicting plastid-based phylogenies and strengthen the mounting evidence of reticulate evolution in Picea. Given the relatively high frequency of hybridization and biparental plastid inheritance 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.

Introduction

Technical feasibility coupled with an apparent simple model of evolution have made plastid genomes (plastomes) a workhorse of plant systematics. While nuclear genomes vary in size by orders of magnitude, a typical land plant plastome contains around 115 unique genes distributed over 120–160 kb (Bock 2007). Plastid genes are generally conserved and encode the two photosystems, the large subunit of the carbon fixing enzyme RubisCO, and ~30 other housekeeping proteins. Compared with the nuclear genome, plastome mutation rates are characteristically low and coding sequences may be subject to stronger functional constraints (Wolfe et al. 1987). In addition, the uniparental plastid inheritance found in most species is expected to result in clonal lineages, a smaller effective population size ($N_e$) than the nuclear genome, and the absence of sexual recombination.

Exceptions to nearly every assumption of this simple model of plastome evolution are known (reviewed in Wolfe and Randle 2004). Of these, biparental plastid inheritance is particularly widespread phenomenon. Plastids are most often transmitted by the seed parent in angiosperms (Zhang et al. 2003). In coniferous gymnosperms, where plastid inheritance is typically paternal, realized biparental inheritance has been documented in Pinus (White 1990), Picea (Sutton et al. 1991), and Larix (Szmidt et al. 1987). Heteroplasmy resulting from biparental inheritance can be expected to be transient due to the vegetative sorting of organelles at every cell division (Wolfe and Randle 2004). However, a recent study of whole plastomes in Citrus revealed heteroplasmy consistent with hybridization in germplasm from mature trees (Carbonell-Caballero et al. 2015). Biparental plastid inheritance and stable heteroplasmy allow for violation of another axiom of plastome evolution: sexual recombination.

Robust evidence for recombination between biparentally inherited plastomes in natural populations is lacking. However, population studies of Pinus (Marshall et al. 2001) and Cycas (Huang et al. 2001) have invoked recombination to explain unusual patterns of plastome polymorphism and linkage disequilibrium. Interspecific recombination has been hypothesized as the source of the markedly different phylogenies inferred from different plastid loci in Silene (Erixon and Oxelman 2008), Picea (Bouillé et al. 2011), and the diatom genus Pseudo-nitzschia (D’Alelio and Ruggiero 2015). Given that all sites in the plastome should share the same topology under clonal inheritance, recombination may more broadly underlie cases of discordance among plastid loci in studies.

© The Author 2017. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Mol. Biol. Evol. 34(7):1689–1701 doi:10.1093/molbev/msx111 Advance Access publication April 5, 2017 1689
where its presence has not been inferred (e.g. polygrammoid ferns, Schneider et al. 2004; Asteraeaceae, Doorduin et al. 2011). As recombination can severely bias population and phylogenetic analyses (e.g. Schierup and Hein 2000), new detection methods are warranted. In particular, approaches using complete chromosomes or organelles can exploit the expected spatial mosaic of evolutionary histories generated by recombination (e.g. Maynard Smith 1992) and use structural features to identify probable breakpoints or hotspots.

Repetitive sequences contribute to intra- and intermolecular recombination during plastome replication and repair (Day and Madesis 2007; Marečal and Brison 2010). During intermolecular recombination, repeats can act as recognition or cleavage sites through primary and secondary structures (Kawata et al. 1997). In addition, repeats may help maintain plastome size and content during recombinational DNA repair. For example, double-stranded breaks in Chlamydomonas plastomes are repaired through homologous recombination, which can result in an unstable plastome dimer that is reduced through intramolecular recombination between repeats (Odom et al. 2008). This process could explain why sexual recombination hotspots in the Chlamydomonas plastome are correlated with dispersed and inverted repeats (Lemieux et al. 1990; Odom et al. 2008).

Based on this evidence, we reasoned rare sexual recombination in land plant plastomes should also be associated with repeats.

We developed and applied phylogenetic tests for recombination to 65 whole plastomes representing all ~35 extant species of Picea, a genus of coniferous trees broadly distributed in the northern hemisphere (Farjon 1991; Eckenwalder 2009). In contrast to the canonical quadripartite plastome structure found in most plants, the large inverted repeat (IR) region is highly reduced in Picea and other conifers (Strauss et al. 1988; Wu et al. 2011). Instead, the Picea plastome contains three single copy regions (the F1, F2, and F3 regions), which are intervened by three large repeats (Wu et al. 2011, see fig. 1). Large structural rearrangements between these repeats have been reported (Wu et al. 2011; Nystedt et al. 2013), which supports their role as recombination substrates during intra- and intermolecular recombination. We predicted sexual recombination in Picea would generate different evolutionary histories among the F1, F2, and F3 regions—regardless of genomic context, mutation rates, or natural selection—but not within them. Combining this hypothesis-based approach with purely data-driven tests, we identified at least one interspecific recombination event that produced a chimeric plastome in one-third of Picea species. This result remained after controlling for process that may generate patterns similar to recombination, such as process heterogeneity, saturation, and positive selection. Given that hybridization occurs in ~15% of plant genera (e.g. Ellstrand et al. 1996; Whitney et al. 2010), such recombinant plastomes may not be unique to Picea.

Results

Plastome Assembly and Divergence

We used whole-genome shotgun sequencing to generate 61 draft plastomes, which together with existing assemblies 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” section), but all assemblies had high coverage and large scaffolds (table 1). De novo plastomes averaged 121,885 bp (± 2,861) or ~98% of reported reference-quality assemblies (Wu et al. 2011; Nystedt et al. 2013; Jackman et al. 2015; Yang et al. 2016). Assemblies generally comprised five large nonoverlapping scaffolds, which correspond to the intervening regions between the two copies of the highly reduced inverted repeat (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 (Prijibelski et al. 2014) or differences in Illumina chemistry. Assembly statistics, detailed locality information, and Genbank and Short Read Archive (SRA) accession numbers for each individual are reported in supplementary table S1, Supplementary Material online.

Subsequent analyses included four additional published Picea plastomes (Wu et al. 2011; Nystedt et al. 2013; Jackman et al. 2015; 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 region between plastomes (Wu et al. 2011; Nystedt et al. 2013). Indels were found in only three genes: accD, chlN, and ycf1.

Nucleotide diversity (π) across the plastome was relatively low (π = 0.0030; table 2). Sliding-scale estimates of π in nonoverlapping 350-bp windows revealed three “hotspots” of diversity occurring near the boundaries of the structural regions (fig. 1). Five genes (psbH, psbT, rpl23, ycf1, ycf2) were more polymorphic than intergenic spacers on average. Notably, ycf1 and ycf2 together contain 57% of all coding sequence polymorphisms (supplementary table S2, Supplementary Material online). Mean synonymous and nonsynonymous divergence was low (dN = 0.017 ± 0.024, dS = 0.040 ± 0.031) but varied markedly among genes (fig. 2). Mean substitution rates were significantly elevated in the ycf genes (GLM, t = 2.828, P < 0.001)—a pattern driven entirely by ycf1 and ycf2—and in the photosystem II complex psb (GLM, t = 2.103, P < 0.05; fig. 2). Six genes comprised solely nonsynonymous substitutions while eight contained only synonymous substitutions (fig. 2 and supplementary table S2, Supplementary Material online).

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 ≤ 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, Supplementary Material online). Most genes had one or two positively selected codons, but the massive ycf1 gene contained 73 sites (3.5% of the predicted protein) potentially under selection. However, many of these sites were located within repetitive sequences and poor alignment may have

![Consensus map of the Picea plastome.](image)

**Fig. 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 structural regions (F1, F2, and F3) of the plastome, which are intervened by large repeats represented by white arrows (not to scale). The innermost circle depicts average pairwise nucleotide diversity ($\pi$) estimated in 350 bp moving windows.

**Table 1.** Sequencing and Assembly Metrics for 61 *Picea* Draft Plastomes.

| Sequencing system | Mean number of scaffolds | N50 scaffold length (bp) | Average largest scaffold (bp) | Mean read depth |
|-------------------|--------------------------|--------------------------|-------------------------------|-----------------|
| Illumina HiSeq 2000 and 2500 2 × 100 bp, 2 × 125 bp | 5.77 (1.76) | 56,475 (2,528) | 66,112 (13,562) | 46.83 (23.32) |
| Illumina HiSeq 4000 2 × 150 bp | 6.75 (1.35) | 22,939 (1,162) | 55,847 (6,949) | 413.51 (239.18) |
| Illumina MiSeq 2 × 100 bp | 7 | 23,241 | 55,799 | 47.17 |

Note.—Standard errors are reported in parentheses.
of this clade during the last two million years, the uncertain topology is likely due to the rapid and recent radiation of the "abies" clade, in which several nodes had weakly supported backbone nodes. Out of 1,000 replicates, supporting a most recent common ancestor of the "glauca", "abies", and "mariana" clades.

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

Saturation, selection, or recombination could explain the near-equal occurrence of conflicting bipartitions in the whole-plastome data set and the disagreement among earlier studies. To assess the causes of phylogenetic discord 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 regions (F1, F2, and F3) of Pinaceae plastomes.

**Table 2.** Alignment Length, Number of Polymorphic Sites, and Nucleotide Diversity (\(\pi\)) by Genomic Context Estimated from 65 Picea Plastomes.

| Context            | Alignment length (bp) | Bi-allelic sites | \(\pi\)          |
|--------------------|-----------------------|------------------|------------------|
| Protein-coding exons | 61,256                | 430              | 0.0020 (1.6 \(\times\) 10\(^{-4}\)) |
| Four-fold synonymous sites | 7,720                | 180              | 0.0025 (2.9 \(\times\) 10\(^{-4}\)) |
| tRNAs              | 2,874                 | 12               | 0.0015 (1.6 \(\times\) 10\(^{-4}\)) |
| rRNAs              | 4,522                 | 13               | 0.0002 (7.6 \(\times\) 10\(^{-5}\)) |
| Introns            | 9,929                 | 197              | 0.0023 (3.2 \(\times\) 10\(^{-4}\)) |
| Intergenic spacers | 49,373                | 1,681            | 0.0048 (1.1 \(\times\) 10\(^{-4}\)) |
| Whole plastome     | 127,954               | 2,333            | 0.0030 (6.6 \(\times\) 10\(^{-5}\)) |

Note.—Standard errors are reported in parentheses.

**Phylogenetic Discordance in the Picea Plastome**

Phylogenetic analysis of the 65 whole plastomes using maximum likelihood resolved five major clades comprising: (1) the North American P. engelmannii, P. glauca, and P. mexicana (hereinafter “glauca”), (2) P. abies and predominately northeast Asian taxa (“abies”), (3) P. mariana and four geographically disparate species (“mariana”), (4) taxa from the Qinghai-Tibetan Plateau and surrounding regions (“likiangensis”), and (5) a diverse clade containing species representing the entire geographic range of genus (“chihuahuana”; fig. 3). Relationships within most clades were strongly supported with the notable exception of the “abies” clade, in which several nodes had < 50% support. The uncertain topology is likely due to the rapid and recent radiation of this clade during the last two million years (Lockwood et al. 2013; Ran et al. 2015). In contrast, the backbone topology of the genus was poorly resolved, with 52% of trees supporting a “glauca”-“abies” sister relationship and 39% supporting a most recent common ancestor of the “glauca”, “abies”, and “mariana” clades.

We analyzed bipartition frequencies in bootstrapped trees to assess the presence of alternate topologies at these two weakly supported backbone nodes. Out of 1,000 replicates, the “glauca” clade occurred in a bipartition with P. stictensis and P. breweriana at even greater frequency than the relationship included on the consensus tree in figure 3 (46% vs. 39% of bootstrap replicates). Both the “mariana” and “glauca” clades were grouped with the “abies” clade to the exclusion of all other species with almost equal frequency (52% vs. 48%). This result is consistent with prior plastid-based studies, which discovered conflicting sister clades for “abies” taxa depending on the locus analyzed (cf. Boüillé et al. 2011; Ran et al. 2015). These results suggest the presence of highly supported yet conflicting topologies in the whole-plastome alignment.
which are subtended by large repeats that likely serve as recombination substrates (Wu et al. 2011; see fig. 1). We reason these repeats could facilitate recombination between divergent plastome copies if biparental inheritance and plastid fusion occurred. Discordance among structural fragments would be consistent with our hypothesis of recombination, but discordance between genic and intergenic regions would be more consistent with parallel mutations from 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 spacer 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 indicates 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 F3 region versus the F1 and F2 regions (table 4 and fig. 4). While the placement of “mariana” is not concordant between the F3 region versus the F1 and F2 regions, concatenation of the three tests strongly supported incongruence between the F3 region versus the F1 and F2 regions (table 4 and supplementary fig. S2, Supplementary Material online). Therefore, phylogenetic discordance between the F3 versus the F1 and F2 regions cannot be adequately explained by saturation or selection.

We tested if phylogenetic discordance between regions could be explained by differences in sequence evolution by comparing only the following: (1) intergenic sites, (2) protein-coding exons, and (3) 4-fold synonymous sites (4D) on each region. Intergenic sites in the F1 and F2 regions yielded statistically indistinguishable topologies but protein-coding exons were significantly discordant (table 4 and supplementary fig. S2, Supplementary Material online). Given the paucity of polymorphic 4D sites on these two regions, we instead compared protein-coding exons without significant evidence of selection. A disproportionate number of codons on the F1 region are putatively under positive selection (see table 3), and their removal yielded topologies no more different than expected by chance (table 4). As discordance between the F1 and F2 regions can be explained by just eight polymorphisms, we concatenated these two regions for tests against the larger F3 region because using a similar number of sites each phylogeny should improve the accuracy of the permutation tests.

Protein-coding exons, intergenic sites, and 4D sites on the concatenated F1 and F2 regions were strongly discordant with those on the F3 region according to all three tests, 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., 4D sites on the F3 region) yielded a similar topology as the entire region (fig. 4), albeit with poorer resolution (supplementary fig. S2, Supplementary Material online). Therefore, phylogenetic discordance between the F3 versus the F1 and F2 regions 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 structural regions 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 on the whole plastome alignment. This two-step approach helps safeguard against potentially spurious results from exploratory tests yet allows the data to refute our hypothesis that recombination should involve the repeats subtending the three structural regions.

Table 3. Positive Selection Inferred from the M1a and M2a Random Site Models Implemented in PAML (P ≤ 0.001).

| 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 | 165(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)                                                             |
|                          | rpsC2 | -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(AS), 1083(L,F), 1139(N,K), 1149(SA), 1223(F,L), F1306(F,L), 1378(LF), 1383(SL), 1507(SL), 1633(LF), 1708(D,E), 1709(K,Q), 1790(V,T), 1912(RE) |
|                          | ycf4 | -815.00 | -799.87 | 118(LP)                                                              |

Note.—Positively selected sites have ω > 1 with posterior probability ≥ 0.95 according to the Bayes empirical Bayes method. Selected sites are listed in parenthesis. Site number refers to the position in the Picea alignment for that protein.
The seven tests implemented in RDP4 identified two putative recombination events, both with breakpoints estimated near region junctions (fig. 5 and table 5). All tests supported a recombinant plastome shared by the “abies” clade originating from parents from the “glauc” and “mariana” clades (table 5). While the majority of the “abies”-type plastome suggested a close relationship with the “glauc” clade, a ~30 kbp segment was highly similar to P. jezoensis, a species from the “mariana” clade distributed from central Japan to northeastern Siberia. This relationship between the “abies” clade and P. jezoensis is also evident in our a priori phylogenetic analysis, which supports a shared ancestor of the F2 region (fig. 4). Maximum likelihood recombination breakpoints were placed near the repeats bounding the F2 region, but broad confidence intervals also encompassed most of the F1 region (fig. 5). Although the phylogenetic discordance and recombination tests differ somewhat in the extent of the exchange, both support a recombinant plastome in the “abies” clade with at least the F2 region inherited from a “mariana”-like parent.

A second putative recombination event spanning the F1 and F2 fragments involved the “likiangensis” clade of Qinghai-Tibetan Plateau taxa and an unknown species similar to P. pungens, a species from the “mariana” clade (table 5 and fig. 5). The F3 region topology supported a sister relationship between the “likiangensis” and “mariana” clades with 68% bootstrap support (fig. 4), whereas the F1 and F2 regions intercalated taxa from the “likiangensis” and “chihuahuana” clades (fig. 4 and supplementary fig. 2, Supplementary Material online). As the RF, QD, and ILD randomization tests were global measures of discordance, this more equivocal recombination event likely contributed to the highly significant results in table 4.

**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.

|      | ILD  | QD   | RF   |
|------|------|------|------|
| F1 vs. F2 PCE | 0.005 | 0.40  | 0.66  |
| F1 vs. F3 region | 0.01 < 0.001 | 0.11  | 0.001  |
| F2 vs. F3 region | 0.01 < 0.001 | 0.01  | 0.001  |
| F2 vs. F1 region | 0.20  | 0.17  | 0.23  |

**Comparisons by genomic context and structural region**

- **F1 vs. F2 PCE:** 0.01 < 0.001 < 0.0001
- **F1 vs. F2:** 0.16 0.26 0.11
- **F2 vs. F3 intergenic:** 0.85 1.00 1.00
- **F3 vs. F1/F2 intergenic:** 0.01 < 0.001 < 0.0001
- **F3 vs. F1/F2 PCE:** 0.01 < 0.001 < 0.0001
- **F3 vs. F1/F2 4Dc:** 0.86 < 0.001 < 0.0001

**Table:** The highly significant results in table 4.

**Note:** F1, F2, and F3 regions refer to the three major single-copy units of the Picea plastome (see fig. 1). The F1 and F2 regions are concatenated for comparisons by genomic context against the F3 region.

- **F1:** Protein-coding exons.
- **F2:** Protein-coding exons with no evidence of positive selection.
- **F3:** Four-fold synonymous sites.

**Fig. 3.** Maximum likelihood analysis of the whole plastome alignment of 65 Picea accessions provides strong support for many shallow nodes but the relationships between the “glauc”, “abies”, and “mariana” clades 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, Supplementary Material online.

**Discussion**

Uniparental inheritance, strong functional constraints, and absent recombination are key but often untested
assumptions in plant evolutionary studies (Wolfe and Randle 2004). Using whole genome assemblies and comprehensive taxon sampling, we found that *Picea* plastomes are highly conserved in sequence and structure but shaped by interspecific recombination and punctuated by putative positive selection. Our results add to the growing body of evidence of complex plastid evolution and demonstrate how overly simple models can lead to misleading inferences.

**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, however, substitution rates in *Picea* are similar to those reported from more limited sets of loci in *Pinus* (Willyard et al. 2007; Wang and Wang 2014). Both genera have ancient origins in the Cretaceous (Wang et al. 2000; Klymiuk and Stockey 2012), but the diversification of extant pine and spruce species occurred within the last 20 My, perhaps facilitated by a cooling climate after the Miocene optimum (Willyard et al. 2007; Lockwood et al. 2013; Wang and Wang 2014; Ran et al. 2015). Although we did not estimate absolute substitution rates explicitly, the similarity of the *Pinus* and *Picea* relative rates and divergence times together suggests a similar per-generation mutation rate in these two Pinaceae genera.

**Selection on 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 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, *psbH* and *ycf1*, were inferred to have codons with $\omega > 1$ in both *Pinus* and *Picea*. To our knowledge, *psbH* has not been widely implicated in positive evolution in angiosperms, and high $\omega$ at this gene may represent a unique attribute of Pinaceae. In contrast, *ycf1* is exceptionally divergent across land plants (Dong et al. 2015) and shows elevated $\omega$ in some angiosperm lineages (e.g. Hu et al. 2015). Surprisingly, *ycf1* encodes an essential component of plastid protein import apparatus (Kikuchi et al. 2013), the hypervariable repeat motifs of which may serve an adaptive function (de Vries et al. 2015).

Consistent with an adaptive explanation, positively selected sites in *Picea* were disproportionally inferred in genes involved in photosynthesis (table 3). However, nonsynonymous mutations may become fixed through nonadaptive processes. Small effective population sizes, rare recombination, and
strong mutational bias are all characteristic of plastomes and can facilitate the fixation of slightly deleterious alleles (Felsenstein 1974; Kimura 1983; Moran 1996). On the other hand, plastids have highly efficient DNA repair systems, which could be expected to purge deleterious mutations (Day and Madesis 2007; Mare´chal and Brisson 2010). Site-directed mutagenesis and in vitro characterization of mutant proteins (e.g. Lan et al. 2013) could help clarify the role of adaptive and nonadaptive processes in shaping plastid substitution rates.

**Table 5.** Bonferonni Corrected P Values for the Two Recombination Events Detected in the “abies” and “likiangensis” Clades by Unguided Tests as Implemented in RDP4.

| Recombination test | “abies” | “likiangensis” |
|--------------------|---------|----------------|
| RDP                | $6.71 \times 10^{-9}$ | $2.36 \times 10^{-5}$ |
| GENECOV            | $1.89 \times 10^{-11}$ | n.s. |
| BootScan           | $6.71 \times 10^{-5}$ | n.s. |
| MaxChi             | $1.81 \times 10^{-7}$ | $6.11 \times 10^{-6}$ |
| Chimaera           | $1.83 \times 10^{-4}$ | $6.87 \times 10^{-8}$ |
| SiScan             | $9.46 \times 10^{-11}$ | $1.55 \times 10^{-3}$ |
| 3Seq               | $4.67 \times 10^{-4}$ | $2.57 \times 10^{-9}$ |

n.s., not significant.

**Reticulate Evolution and Chimeric Plastomes**
Recombination between divergent haplotypes has been invoked to explain phylogenetic discordance among plastid loci (Erixon and Oxelman 2008; Bouillé et al. 2011; D’Alelio and Ruggiero 2015). 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 (Huang et al. 2001; Marshall et al. 2001; Erixon and Oxelman 2008; Bouillé et al. 2011; D’Alelio and 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 for *Picea*, topologies inferred for the F3 plastome region closely match those from mitochondrial loci (Ran et al. 2015) and nuclear loci (Zou et al. 2016; Liu J-Q, unpublished data), which suggests the “glauca” and “abies” clades may be related by vertical descent. In contrast, the F2 and potentially the F1 region of the “abies” plastome was inherited horizontally from a member of the “mariana” clade, 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 “abies” species successfully hybridize with the “mariana” clade in controlled crosses (Wright 1955). In natural populations, many parapatric spruces hybridize readily (e.g. Jaramillo-Correa and Bousquet 2005; Du et al. 2011), including distantly related species (Hamilton and Aitken 2013). Comparison of phylogenies inferred from mitochondrial and plastid loci, which are differentially inherited in conifers, also support widespread introgression throughout the genus (Bouillé et al. 2011; Ran et al. 2015). In particular, P. jezoensis has a “mariana”-type plastome but more closely resembles the “abies” clade at mitochondrial and some nuclear loci (Ran et al. 2015). Fossil and genetic evidence indicate the continuous presence of P. jezoensis and its extinct relative P. hondoensis in northeastern Russia throughout the Pleistocene (Aizawa et al. 2007). The most recent common ancestor of the “abies” clade likely emerged in eastern Asia during this same time period (Ran et al. 2015), and introgression between a P. jezoensis-like species and the “abies” ancestor may have occurred here, prior to the clade’s diversification and colonization of Eurasia.

Mechanisms maintaining uniparental inheritance are varied (Bock 2007), but hybridization or interpopulation crosses have been linked to a breakdown of these processes (Hansen et al. 2007; Ellis et al. 2008; Barnard-Kubow et al. 2017). In Helianthus and Passiflora, rates of plastid leakage vary by family and species but realized biparental inheritance occurs in ~2–4% of offspring overall, respectively (Hansen et al. 2007; Ellis et al. 2008). As biparental inheritance is not uncommon on an evolutionary timescale, fusion of the two parental plastids is probably the limiting step preventing more widespread recombination (Day and 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 could occur between highly similar genomes through normal DNA replication and repair mechanisms (Day and Madesis 2007; Maréchal and Brisson 2010). Spread of the chimeric plastome through the ancestor of the “abies” clade may have then proceeded through entirely neutral processes, such as allele surfing during range expansion (Excoffier and Ray 2008).

The prevalence of sexual recombination in plastomes is unknown. Taxa with higher rates of biparental plastid inheritance, including groups prone to hybridization (Hansen et al. 2007; Ellis et al. 2008; Barnard-Kubow et al. 2017), are presumably the most susceptible. In addition, three genera with predominately paternal plastid inheritance, including Picea, have been suggested to form chimeric plastomes (Huang et al. 2001; Marshall et al. 2001), which may indicate a link between this inheritance mode and a propensity for recombination. However, until further research clarifies the factors promoting sexual plastome recombination, adding recombination tests to phylo- and population genetic pipelines is prudent.

A Word of Caution for “Chloroplast Phylogenomics”

Plastome phylogenies effectively estimate the history of a single gene and do not necessarily represent species histories. Incomplete lineage sorting can cause the plastome to vary markedly from the species tree, and introgression can further exacerbate this discordance (Doyle 1992). While whole Picea plastomes yielded high resolution within most clades, they also revealed conflicting histories generated by vertical descent and introgression. Furthermore, the three genomes in Picea suggest incongruent histories (Bouillé et al. 2011; Ran et al. 2015), and substantial gene-tree discordance is also evidence among nuclear loci (Ran et al. 2015). Reconstructing a useful approximation of the species tree for Picea—and potentially many other plant genera—will likely necessitate methods explicitly allowing for gene tree heterogeneity arising from both incomplete lineage sorting and introgression.

Plastome haplotypes for many species were not reciprocally monophyletic, an issue which further confounds the use of single-gene phylogenies as species tree proxies. Haplotype sharing can arise from some combination of incomplete lineage sorting, hybridization, cryptic and poorly delimited species, or misidentified material. For example, Norway spruce (P. abies) was polyphyletic in our data set, with the two Scandinavian plastomes more closely related to P. obovata than to a Norway spruce originating from Serbia. A polyphyletic Norway spruce was noted by Lockwood et al. (2013) in their predominantly mitochondrial data set, which indicates a need to clarify species boundaries in this taxon. In the absence of a robust species tree estimate, we urge caution when interpreting plastome phylogenies and note results from exemplar accessions may not generalize to an entire species.

Conclusions

We used whole plastome sequences and comprehensive taxon sampling to characterize sequence divergence, positive selection, and evolutionary histories 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 repeats, to develop phylogenetic hypotheses to test for recombination. Given that hybridization (Elkstrand et al. 1996; Whitney et al. 2010) and the potential for biparental plastid are common in plants (Corriveau and Coleman 1988; Zhang et al. 2003), interspecific plastome recombination may not be unique to Picea.

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, Supplementary Material online 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, Supplementary Material online).

Genome Sequencing and Assembly
Total genomic DNA was isolated from leaf material using the Macherey-Nagel NucleoSpin® Plant II kit according to the manufacturer’s 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 × 100 bp and 2 × 125 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, TX) using 2 × 150 bp chemistry. Finally, an entire lane of 2 × 100 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; Prijibelski et al. 2014) using either total genomic reads or reads pre-filtered for similarity to the plastome. For read filtering, we used local alignment in bowtie2 v. 2.2.7 (Langmead and 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 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, default error correction and kmer selection, and a minimum contig size of 500 bp. Sequencing methods, assembly metrics, and Genbank accession numbers for each accession reported in supplementary table S1, Supplementary Material online.

Annotation and Alignment
We tested for structural rearrangements in de novo scaffolds of each accession versus the P. abies reference genome using nuccmer, 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 and 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 (dS) and nonsynonymous (dN) substitution rates were estimated per-gene using the maximum likelihood method (Goldman and 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 dN to dS (ω) using nested random site models (M1a and 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 ycf12 and psbl were excluded because they are located within large repeats (white arrows in fig. 1) and were not uniformly recovered across all accessions.

Phylogenetic Analyses
Sequence alignments were prepared for phylogenetic analysis by stripping sites with >20% missing data and removing tandem repeats using Phobos v. 3.3.12 (http://www.ruhr-uni-bochum.de/spezzoo/cm/cm_phobos.htm; last accessed March 25, 2017). 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 to 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 regions (F1, F2, and F3) of the Picea plastome. Phylogenies were inferred from the partitioned data sets 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. Permutation tests based on tree QD (Bryant et al. 2000) and RF (Robinson and Foulds 1981)
distances tested if phylogenies estimated from different \textit{a priori} subsets are more different than expected by chance. To simulate 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 \textit{a priori} subsets. We implemented the permutation test using the \textit{R} package “\texttt{perm}” (Fay and Shaw 2010) using 2,000 replications and estimated \( P \) values using the default algorithm. QDs were calculated using the program \texttt{rqtdist v. 1.0 (Sand et al. 2013)} and RF distances in the \textit{R} packages “\texttt{phangorn}” (Schliep 2011) and “\texttt{ape}” (Paradis et al. 2004).

Recombination Tests

Unguided tests for recombination were conducted in the \texttt{RDP4} package (Martin et al. 2015) using the \texttt{RDP} (Martin and Rybicki 2000), \texttt{GENECOV}, \texttt{BootScan} (Salminen et al. 1995), \texttt{MaxChi} (Maynard Smith 1992), \texttt{Chimaera} (Posada and Crandall 2001), \texttt{SiScan} (Padidam et al. 1999), and \texttt{3seq} (Boni et al. 2007) methods. After identifying a recombination event, \texttt{RDP4} implements a hidden Markov model to estimate breakpoint positions. Then, \texttt{RDP4} identifies the probable recombinant sequence using the \texttt{PHYLPRO} (Weiller 1998), \texttt{VISRD} (Lemey et al. 2009), and \texttt{EEEP} (Beiko and Hamilton 2006) methods. Bonferroni 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.

Supplementary Material

Supplementary data are available at Molecular Biology and Evolution online.

Acknowledgments

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 Csíba and Rhinaixa Duque-Thiús (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 and Nicolas Delhomme at the Umeå Plant Science Center (UPSC) Bioinformatics Facility provided bioinformatics support. This study was supported by the Trees and Crops for the Future (TC4F) program. ARS was supported in part by Bergvik Skog AB. Constructive criticism from anonymous reviewers improved the manuscript.

References

Aberer AI, Pattengale ND, Stamatakis A. 2010. Parallelized phylogenetic post-analysis on multi-core architectures. J Comput Sci. 1:107–114.
Aizawa M, Yoshimaru H, Saito H, Katsuki T, Kawahara T, Kitamura K, Shi F, Kaji M. 2007. Phylogeography of a northeast Asian spruce, \textit{Picea jezoensis}, inferred from genetic variation observed in organelle DNA markers. Mol Ecol. 16:3393–3405.
Bankovich A, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 19:455–477.
Barnard-Kubow KB, McCoy MA, Galloway LF. 2017. Biparental chloroplast inheritance leads to rescue from cytonuclear incompatibility. New Phytol. 213:1466–1476.
Beiko RG, Hamilton N. 2006. Phylogenetic identification of lateral genetic transfer events. BMC Evol Biol. 6:15–17.
Bock R. 2007. Structure, function, and inheritance of plastid genomes. In: Bock, R, editor. Cell and molecular biology of plasts. Berlin, Heidelberg: Springer Berlin Heidelberg. p. 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 \textit{Picea}. Tree Gene Genomes 7:469–484.
Bryant D, Tsang J, Jarne PE, Li M. 2000. Computing the quartet distance between evolutionary trees. In: Proceedings of the Eleventh Annual ACM-SIAM Symposium on Discrete Algorithms. New York City: ACM Press. p. 285–286.
Carbonell-Caballerò J, Alonso R, Ibañez Y, Terol J, Talon M, Dospazo J. 2015. A phylogenetic analysis of 34 chloroplast genomes elucidates the relationships between wild and domestic species within the genus \textit{Citrus}. Mol Biol Evol. 32:2015–2035.
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.
D’Alessio D, Ruggiero MV. 2015. Interspecific plastidial recombination in the diatom genus \textit{Pseudo-nitzschia}. J Physcol. 51:1024–1028.
Day A, Madesis P. 2007. DNA replication, recombination, and repair in plastids. In: Bock, R, editor. Cell and molecular biology of plasts. Berlin, Heidelberg: Springer Berlin Heidelberg. p. 65–119.
de Vries J, Sousa FL, Böltcr B, Soll J, Gould SB. 2015. YCF1: a green TIC?. Plant Cell 27:1827–1833.
Dong W, Xu C, Li C, Sun J, Zuo Y, Shi S, Cheng T, Guo J, Zhou S. 2015. \texttt{ycf1}, the most promising plastid DNA barcode of land plants. Sci Rep. 5:8348.
Doorduin L, Gravendeel B, Lamers T, Aiyuk A, Hoog T, Vrielink K. 2011. The complete chloroplast genome of 17 individuals of pest species \textit{Jacoba caea vulgaris}: SNPs, microsatellites and barcoding markers for population and phylogenetic studies. DNA Res. 28:93–105.
Doyle JJ. 1992. Gene trees and species trees: molecular systematics as one-character taxonomy. Syst Bot. 17:144–163.
Du FK, Peng XL, Liu JQ, Lacouze M, Hu FS, Petit RJ. 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. Portland (OR): Timber Press, Inc.
Ellis JR, Bentley KE, McCauley DE. 2008. Detection of rare paternal chloroplast inheritance in controlled crosses of the endangered sunflower \textit{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 Silenae (Caryophyllaceae)? Mol Phylogenet Evol. 48:313–315.

Excoffier L, Ray N. 2002. Reconstructing species trees from gene trees. In Tree Hierarchy as an explicit strategy for reconstructing species trees. 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. Königstein: Koeltz Scientific Books.

Fay MP, Shaw PA. 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. Genetcs 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, Attrik SN. 2013. Genetic and morphological structure of a spruce hybrid (Picea sitchensis P. glauca) zone along a climatic gradient. Am J Bot. 100:1651–1662.

Hansen AK, Escobar LZ, 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.

Hsu S, Sablok G, Wang B, Qu D, Barbaro E, Viola R, Li M, Varotto C. 2015. Plastome organization and evolution of chloroplast genes in Cardamine species adapted to contrasting habitats. BMC Genomics 16:354.

Huang S, Chiang YC, Schaal BA, Chou CH, Chiang TY. 2001. Organelle genetic revolutions and structuration. Trends Ecol Evol. 18:299–317.

Kawata M, Harada T, Shimamoto Y, Oono K, Takaiwa F. 1997. Short gradient. Mol Biol Evol. 14:930–940.

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

Kawata M, Harada T, Shimamoto Y, Oono K, Takaiwa F. 1997. Short inverted repeats function as hotspots of intermolecular recombination giving rise to oligomers of deleted plastid DNAs (ptDNAs). Curr Genet. 31:179–184.

Kikuchi S, Bédard J, Hirano M, Hirabayashi Y, Oishi M, Imai M, Takase M, Ide T, Nakai M. 2013. Uncovering the protein translocase at the chloroplast inner envelope membrane. Science 339:571–574.

Kimura M. 1983. The neutral theory of molecular evolution. Cambridge: Cambridge University Press.

Klymiuk AA, Stockey RA. 2012. A lower Cretaceous (Valanginian) seed cone provides the earliest fossil record for coniferous families and homologies with extant taxa. Am J Bot. 99:1069–1082.

Krzywinski MI, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: an information aesthetic for comparative genomics. Genome Res. 19:1639–1645.

Kurtz S, Phillippy A, Delcher AL, Shmoo M, Shumway M, Antonescu C, Salzberg SL. 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 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 phylogenetic 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.
Polypodiaceae and Grammitidaceae: exploring aspects of the diversification of epiphytic plants. Mol Phylogenet Evol. 31:1041–1063.

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 AH. 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. Sunderland (MA): Sinauer Associates.

Szmidt AE, Aldén T, Häggren 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.

White EE. 1990. Chloroplast DNA in Pinus monticola. 2. Survey of within-species variability and detection of heteroplasmic individuals. Theor Appl Genet. 79:251–255.

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

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.

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, Joo M, So S, Yi DK, Shin CH, Lee YM, Choi K. 2016. The complete plastid genome sequence of Picea jezoensis (Pinaceae: Piceoideae). Mitochondrial DNA 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, Sodmergen Liu Y. 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, Yue W, Li L, Wang X, Lu J, Duan B, Liu J. 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.