GBE

Genome Biology and Evolution

SMBE

Predominant and Substoichiometric Isomers of the Plastid Genome Coexist within *Juniperus* Plants and Have Shifted Multiple Times during Cupressophyte Evolution

Wenhu Guo[1,2], Felix Grewe[1,3], Amie Cobo-Clark[4], Weishu Fan[1,3], Zelin Duan[1], Robert P. Adams[5], Andrea E. Schwarzbach[6], and Jeffrey P. Mower[1,3]

[1]Center for Plant Science Innovation, University of Nebraska-Lincoln

[2]School of Biological Sciences, University of Nebraska-Lincoln

[3]Department of Agronomy and Horticulture, University of Nebraska-Lincoln

[4]Department of Biological Sciences, University of Texas at Brownsville

[5]Biology Department, Baylor University

[6]Department of Biomedicine, University of Texas at Brownsville

Corresponding author: E-mail: jpmower@unl.edu.

Accepted: February 23, 2014

Data deposition: This project has been deposited at the EMBL/GenBank data libraries under the accession KF866297–KF866300.

Abstract

Most land plant plastomes contain two copies of a large inverted repeat (IR) that promote high-frequency homologous recombination to generate isomeric genomic forms. Among conifer plastomes, this canonical IR is highly reduced in Pinaceae and completely lost from cupressophytes. However, both lineages have acquired short, novel IRs, some of which also exhibit recombinational activity to generate genomic structural diversity. This diversity has been shown to exist between, and occasionally within, cupressophyte species, but it is not known whether multiple genomic forms coexist within individual plants. To examine the recombinational potential of the novel cupressophyte IRs within individuals and between species, we sequenced the plastomes of four closely related species of *Juniperus*. The four plastomes have identical gene content and genome organization except for a large 36 kb inversion between approximately 250 bp IR containing *trnQ-UUG*. Southern blotting showed that different isomeric versions of the plastome predominate among individual junipers, whereas polymerase chain reaction and high-throughput read-pair mapping revealed the substoichiometric presence of the alternative isomeric form within each individual plant. Furthermore, our comparative genomic studies demonstrate that the predominant and substoichiometric arrangements of this IR have changed several times in other cupressophytes as well. These results provide compelling evidence for substoichiometric shifting of plastomic forms during cupressophyte evolution and suggest that substoichiometric shifting activity in plastid genomes may be adaptive.

Key words: plastid genome, cupressophytes, *Juniperus*, inverted repeat, substoichiometric shifting.

Introduction

The plastid genome (plastome) of photosynthetic land plants is generally 120–160 kb in size with a quadripartite structure involving two inverted repeat (IR) sequences that separate the rest of the genome into large and small single-copy regions (Wicke et al. 2011; Jansen and Ruhlman 2012). The IR in land plants is typically 10–30 kb in length and duplicates the ribosomal RNA cluster and other genes, although some species have much larger or smaller IRs and others have no IR at all (Wicke et al. 2011; Jansen and Ruhlman 2012). When the IR is present, homologous recombination occurs between the two copies causing frequent "flip-flop" inversion of the intervening single copy regions, such that the two isomeric genomic orientations coexist at roughly equal frequency within a single plant (Palmer 1983; Stein et al. 1986).

Among gymnosperms, the IR ranges from large to absent. Species in Gnetales, Cycadales, and Ginkgoales have retained a canonical IR ranging from 17.3 to 25.1 kb (Wu et al. 2007, 2009; Lin et al. 2012). In Pinaceae, the IR is highly reduced to include only the full *trnI-CAU* gene and, in most species,

a portion of the *psbA* gene (Lin et al. 2010; Wu et al. 2011). This reduced Pinaceae IR has lost the ability to generate inversions by homologous recombination (Wu et al. 2011). However, several novel, Pinaceae-specific repeats are recombinationally active, resulting in rearrangement events that have generated distinct interspecific and intraspecific genomic configurations in Pinaceae (Tsumura et al. 2000; Wu et al. 2011). Although these results clearly demonstrated that recombinant forms exist among individuals within a species, they did not establish whether both genomic forms coexist within an individual plant. Southern blot analyses of individuals failed to detect the presence of the isomeric form for any species (Tsumura et al. 2000), indicating that the genomes exist predominantly or fully in one arrangement within a single plant. However, polymerase chain reaction (PCR)-based approaches using DNA extracted from single plants were sometimes (depending on PCR parameters) able to detect the isomeric form (Tsumura et al. 2000; Wu et al. 2011), suggesting that it may be present within an individual, albeit at a substoichiometric level. Alternatively, the rearranged forms detected by PCR may be artifacts arising from PCR-mediated recombination events (Lahr and Katz 2009; Alverson et al. 2011).

Unlike other gymnosperms, cupressophytes have lost one copy of the ancestral land plant IR during evolution, but, like Pinaceae, they have acquired one or more short, novel IRs (Hirao et al. 2008; Wu and Chaw 2013; Yi et al. 2013). In *Cephalotaxus oliveri,* a 544-bp IR that duplicates the *trnQ-UUG* gene was inferred to be recombinationally active because both isomeric forms were detected within a single individual by PCR (Yi et al. 2013). In *Cryptomeria japonica* and *Taiwania cryptomerioides*, however, this *trnQ*-containing IR is only approximately 280 bp and the rearranged form was undetectable by PCR (Yi et al. 2013). Because these results were obtained using a PCR-based approach, it is unclear whether the positive results were due to PCR-mediated recombination and whether the negative results were due to nonoptimal PCR conditions. Thus, as in Pinaceae, it is uncertain whether both genomic isomers coexist in individual cupressophyte plants due to homologous recombination at intermediately sized IRs.

The repeat-mediated rearrangement activity in conifer plastomes, coupled with the possible presence of substoichiometric isomeric forms within individuals, is reminiscent of the well-known process of substoichiometric shifting (SSS) that affects angiosperm mitochondrial genomes (Small et al. 1987). SSS changes the relative ratio of predominant and substoichiometric forms of a plant mitochondrial genome through recombination at intermediately sized repeats (~50–1,000 bp), and possibly also by selective amplification of the substoichiometric form (Woloszynska 2010; Arrieta-Montiel and Mackenzie 2011). Repeat-mediated recombinational activity can be increased by mutations in nuclear factors such as *MSH1*, *OSB1*, and *RECA3*, suggesting that SSS is under nuclear control (Zaegel et al. 2006; Shedge et al. 2007). Mitochondrial SSS activity can create or amplify open reading frame copy numbers, alter gene expression patterns, and lead to phenotypic effects, at least some of which may be evolutionarily adaptive (Arrieta-Montiel and Mackenzie 2011).

To date, SSS has not been reported for the plastome. If plastomic SSS exists, repeat-mediated recombination would have the potential to cause functional genomic changes with phenotypic consequences. To examine the issue of repeat-mediated rearrangement and the presence of substoichiometric isomeric arrangements in conifers, we sequenced the plastomes from four closely related species in genus *Juniperus*. For this work, we collected tissue from single individuals to examine whether different plastomic forms coexist within individual plants. In addition, while we used PCR-based approaches to examine this issue, we also examined mapping information from paired-end reads generated by next-generation sequencing techniques, which should avoid potential issues with PCR-mediated recombination artifacts (Alverson et al. 2011). Comparative genomic analysis provided compelling evidence for the existence of SSS in the plastome of junipers and other cupressophytes.

Materials and Methods

DNA Extraction and Sequencing

Fresh leaf tissue was collected from a single individual of *Juniperus bermudiana* (voucher Adams 11080), *J. monosperma* (voucher Adams 13595), *J. scopulorum* (voucher Adams 13594), and *J. virginiana* (voucher Adams 13549). Vouchers were deposited in the herbarium at Baylor University (BAYLU). Total genomic DNA was extracted from 20 mg of silica-dried *J. bermudiana* material using the Qiagen DNeasy Mini Kit (QIAGEN, Valencia, CA) and from 2 g tissue for the other three species by use of a modified CTAB-based protocol (Doyle JJ and Doyle JL 1987). Quality and quantity of the extracted DNAs were examined by spectrophotometry and agarose gel electrophoresis.

For *J. bermudiana*, long-range PCR products were amplified, pooled, and then sequenced at the University of Georgia Sequencing Core Facility on the Illumina GAII platform. The PCR products covered about 70% of the genome. The remaining 30% of the genome was filled by PCR and Sanger sequencing. The other three DNAs were sequenced at the Indiana University Bloomington Center for Genomics and Bioinformatics Core Facility on the Illumina MiSeq platform, generating approximately 10 million paired-end 250 bp reads from an approximately 800 bp sequencing library for each species.

Genome Assembly and Annotation

The *C. japonica* plastid genome (GenBank Acc. No. NC_010548) was used as a reference to order the

J. bermudiana sequenced genome fragments in Geneious version R6-1 (http://www.geneious.com/, last accessed March 11, 2014). All potential rearrangements found in comparison with *Cryptomeria* were confirmed by additional Sanger sequencing. The other three plastid genomes were assembled by running Velvet version 1.2.03 (Zerbino and Birney 2008) using different pairwise combinations of Kmer values (51, 61, 71, 81, and 91) and expected coverage values (50, 100, 200, 500, 1000, and 2000). Scaffolding was turned off and the minimum coverage value was set to 10% of expected coverage. For each species, a single full-length contig was recovered in at least three independent runs, and the alignment consensus of three independent runs was taken as the final consensus sequence. To validate the genome assemblies, Illumina reads were mapped onto the consensus sequences with Bowtie version 2.0.0 beta 5 (Langmead and Salzberg 2012) as described previously (Grewe et al. 2013). Discrepancies were corrected using the sequence present in the majority of mapped read sequences.

Protein-coding, ribosomal RNA, and transfer RNA genes in the four juniper plastomes were initially annotated by use of the DOGMA webserver (Wyman et al. 2004). DOGMA annotations were manually checked by blast searches with orthologous sequences from other Cupressaceae plastomes, and, in some cases, by sequence alignment using MUSCLE version 3.8.31 (Edgar 2004).

Genome Structural Analyses

To compare plastome organization within cupressophytes, genomes from 8 representative species were aligned using Mauve version 2.3.1 (Darling et al. 2010). For this analysis, the start point of each genome was arbitrarily set as the *ycf2* start codon.

For Southern blotting, approximately 1 μg of *J. monosperma, J. virginiana,* and *J. scopulorum* DNA was digested with restriction enzymes EcoRI and HindIII, separated on a 0.5% agarose gel, and transferred to nylon membrane following established procedures (Sambrook and Russell 2001). Approximately 800 ng of PCR-derived probes were labeled with Digoxigenin (DIG) using the DIG High Prime DNA Labeling and Detection Starter Kit II following the manufacturer's protocol (Roche, Mannheim, Germany). The membrane was prehybridized in ULTRAhyb hybridization solution (Life Technologies, Carlsbad, CA) for 4 h at 42 °C and then hybridized in ULTRAhyb solution containing the DIG-labeled probe overnight at 42 °C. The membrane was washed twice in 2× saline-sodium citrate (SSC) + 0.1% sodium dodecyl sulphate (SDS) for 5 min at room temperature and then twice in 0.5× SSC + 0.1% SDS for 15 min at 65 °C. Hybridized probes were detected with chemiluminescent substrate (CSPD) ready-to-use according to the DIG High Prime DNA Labeling and Detection Starter Kit II protocol. Subsequently, the membrane was exposed to a photo film for 10 min prior to development.

Primers for the variable cycle PCR analysis were designed in genes flanking the *Juniperus* IR: *rps4* (5′-CCTGGTAAAGTTTTG ABACG-3′), *psbK* (5′-CAAATGAAAAGCGGCATCG-3′), *chlB* (5′-GTTCCAATATGAGCAGGACCAG-3′), and *trnL*-UAA (5′-G TTTCCATACCAAGGCTC-3′). PCR was performed with a C1000 thermal cycler (Bio-Rad) using the following primer combinations (*rps4* + *chlB*, *rps4* + *trnL*-UAA, *psbK*+*chlB*, *psbK*+*trnL*-UAA) and GoTaq Flexi DNA Polymerase with supplied reagents (Promega). Each reaction was 10 μl in volume and included 20 ng DNA. Reactions were amplified for 5, 10, 15, 20, 25, 30, or 35 cycles of denaturation (95 °C for 30 s), annealing (55 °C for 1 min), and elongation (72 °C for 2 min). All reactions also included an initial denaturation step (95 °C for 2 min) and a final elongation step (72 °C for 5 min).

To quantify the relative frequency of the two isomeric genomic forms, Illumina paired-end reads were mapped to the genome using Bowtie 2 with default parameters. To avoid any mapping ambiguity, reads were required to unambiguously map to the nonrepetitive flanking sequences on either side of the repeats. This was possible because the average insert size of the sequencing libraries was approximately 800 bp, which easily spanned the approximately 250 bp repeats. A custom Perl script was used to count repeat-spanning read pairs, enabling us to quantify the frequency of the repeat in each possible genomic arrangement. Isomer frequencies were calculated by dividing the number of read pairs that support the alternative conformation by the total number of read pairs that support either conformation. To ensure that the results of the read-pair mapping analysis were not the result of cross-contamination of the Illumina data sets, raw Illumina sequence reads were aligned to three genomic regions that exhibited variability among the *Juniperus* species. The variable genomic regions were identified by manual inspection of a genomic alignment generated by MAFFT (Katoh and Toh 2008) with default parameters. All Illumina sequence reads that could be mapped to these genomic regions were extracted and then aligned to the variable regions by MAFFT with default parameters.

Phylogenetic Analysis

Plastomes from 12 cupressophyte and 9 Pinaceae species (supplementary table S1, Supplementary Material online) were downloaded from GenBank or generated in this study. To ensure annotation consistency among genomes, we performed an all-against-all BlastN search of all protein-coding genes from all species to identify missing genes or genes with incorrect start or stop codon annotations. Genomes with potentially missing or misannotated genes were manually checked, and their annotations were corrected if the gene could be identified or reannotated to improve consistency among species. For *petL* of *Picea morrisonicola* and *ndhB* of *Cryptomeria*, *T. flousiana* and *Taxus*, we used an upstream

start codon to improve sequence similarity to orthologous genes. For *petD* of *Keteleeria*, we used an upstream stop codon. We identified numerous unannotated genes in *Pinus thunbergii* (*ccsA*, *cemA*, *petL*, *petN*, *psbZ*, *ycf1*, *ycf2*, *ycf3,* and *ycf4*), *Podocarpus* (*psaI*, *psaJ*, *psaM*, *rpl20*, *rpl33,* and *rps18*), *Pseudotsuga* (*psbA*), and *Taxus* (*psbZ*, *rps2,* and *rps12*).

All 83 cupressophyte protein-coding genes were extracted from the corrected annotations and then individually aligned with MUSCLE version 3.8.31 (Edgar 2004) using default settings. Alignments were filtered using Gblocks version 0.91b (Castresana 2000) in DNA mode with relaxed parameters (b4=5 b5=h). Filtered alignments were concatenated in SequenceMatrix version 1.7.8 (Vaidya et al. 2011), producing a final alignment of 68,497 bp. Maximum likelihood phylogenetic trees were constructed with the GTR+G substitution model in RAxML version 7.2.8 (Stamatakis 2006). Tree robustness was assessed by nonparametric bootstrapping with 1,000 replicates.

Results

Juniper Plastid Genome Size and Content

The plastomes of *J. bermudiana, J. monosperma*, *J. scopulorum*, and *J. virginiana* are 127,659, 127,744, 127,774, and 127,770 bp in length, respectively (fig. 1; supplementary fig. S1, Supplementary Material online), which are similar to the 127–146 kb sizes of other sequenced cupressophytes (Wu and Chaw 2013). The four junipers have an identical set of 82 protein-coding genes, 4 ribosomal RNAs, 33 transfer RNAs, and 18 introns. Notably, the *rps16* gene contains frameshift mutations indicating nonfunctionality, whereas the *clpP* gene probably requires RNA editing to create a stop codon but is otherwise intact and presumably functional. Overall, protein-coding gene content is relatively stable among cupressophytes (table 1). Gene content variation includes 1) loss of *rpl32* from *Cephalotaxus*; 2) pseudogenization of *rps16* in *Juniperus*, *Taxus,* and *Podocarpus*; 3) pseudogenization of *clpP* in *Cryptomeria*, *Cunninghamia*, *Taxus,* and *Podocarpus*; 4) pseudogenization of *infA* in *Cunninghamia*; and 5) pseudogenization of *ycf1* in *Taxus*. All sequenced cupressophytes lack the *ycf12* gene that is present in other gymnosperms.

Different Isomeric Plastome Arrangements Predominate among Junipers

The four juniper genome assemblies have an identical organization except for a 36 kb inversion (fig. 1; supplementary fig. S1, Supplementary Material online), such that the *J. virginiana* and *J. monosperma* plastomes exist in one arrangement (designated "A"), whereas *J. scopulorum* and *J. bermudiana* exist in the inverted isomeric form (designated "B"). This inversion segment is precisely flanked by the largest IR in the four juniper plastomes (244 bp in *J. monosperma* and 257 bp in the other three species) that fully duplicates the *trnQ-UUG* gene. The presence of different isomeric forms among junipers indicates that this approximately 250 bp *trnQ*-IR has promoted homologous recombination activity in *Juniperus*, as previously observed for the short IRs in *Cephalotaxus* and some Pinaceae species (Wu et al. 2011; Yi et al. 2013).

To verify the interspecific variation of plastome arrangements and to test for the presence of both arrangements within an individual plant, we performed southern blot analysis using the IR sequence as a probe. Restriction sites surrounding the IRs for the three examined juniper plastome sequences are shown in figure 2*A* (*J. bermudiana* was not examined because we lacked sufficient quantity of DNA). If the assembled genomic arrangement is the only genomic form within the single individual from each of the four species, we would expect two restriction fragments (consistent in size to the restriction maps in fig. 2*A*) to hybridize with the IR probe. However, if both isomeric forms are present at a detectable frequency within an individual plant, we would expect hybridization to four different fragments (two that correspond to the restriction maps in fig. 2*A* and two that correspond to the inverted isomeric form). The southern blot results show two hybridization signals for each plant (fig. 2*B*), which corresponds to the restriction pattern expected from figure 2*A*. These results indicate that the plastomes exist predominantly or completely in the "A" arrangement within the two *J. monosperma* and *J. virginiana* plants, whereas the plastome of the *J. scopulorum* plant exists predominantly or completely in the isomeric "B" arrangement, in agreement with the genome assembly results.

The southern blot failed to detect both genomic forms within the individual plants, but it is possible that the alternative forms are at a substoichiometric level that is too low to be detectable by southern blotting. As an alternative approach, we performed semiquantitative PCR with a variable number of cycles to compare the relative frequency of the "A" and "B" arrangements in the four *Juniperus* species (fig. 2*C*). From these results, it is clear that all four reactions generate products for all four species. Second, it is also apparent that there are minor differences in amplification efficiency between the two PCR reactions for each arrangement (note for *J. bermudiana*, *J. scopulorum*, and *J. virginiana* the earlier appearance of *psbK/chlB* products compared with *rps4/trnL* products for the "A" arrangement, and the earlier appearance of *rps4/chlB* products compared with *psbK/trnL* products for the "B" arrangement), After accounting for the minor differences in amplification efficiency, the results indicate that "A" products appear earlier than "B" products for *J. monosperma* and *J. virginiana*, while "B" products appear earlier than "A" products for *J. bermudiana* and *J. scopulorum*. Taken together, these PCR results confirm the predominant genomic configuration for each species ("A" for *J. monosperma* and *J. virginiana*; "B" for *J. bermudiana* and *J. scopulorum*). Additionally, they suggest that the isomeric form is present at a substoichiometric level within each individual plant from

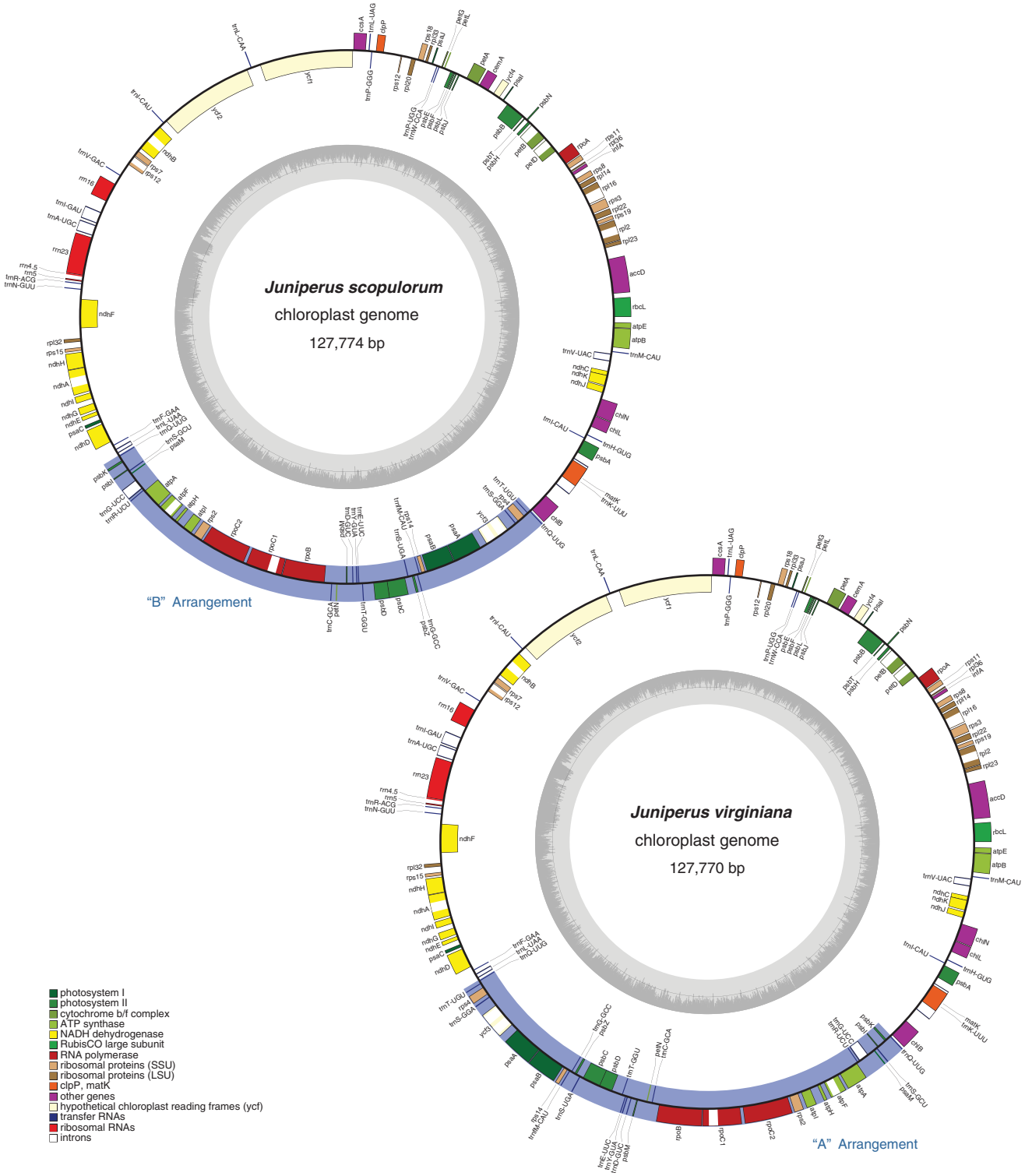

Fig. 1.—*Juniperus virginiana* and *J. scopulorum* plastome maps. Genes transcribed clockwise are depicted on the inside of the circle, genes transcribed anti-clockwise on the outside. The location of the IR-mediated inversion is highlighted on the outer circle by blue bars. GC content is represented on the inner circle by dark gray bars.

Table 1
Variation in Protein-Coding Gene Content among Cupressophytes

	Juniperus	*Cryptomeria*	*Taiwania*	*Cunninghamia*	*Cephalotaxus*	*Taxus*	*Podocarpus*
78 genes[a]	+	+	+	+	+	+	+
clpP	+	ψ	+	ψ	+	ψ	ψ
infA	+	+	+	ψ	+	+	+
rpl32	+	+	+	+	–	+	+
rps16	ψ	+	+	+	+	ψ	ψ
ycf1	+	+	+	+	+	ψ	+
ycf12	–	–	–	–	–	–	–
Total	82	82	83	81	82	80	81

[a]Seventy-eight genes include *accD, atpA, atpB, atpE, atpF, atpH, atpI, ccsA, cemA, chlB, chlL, chlN, matK, ndhA, ndhB, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK, petA, petB, petD, petG, petL, petN, psaA, psaB, psaC, psaI, psaJ, psaM, psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ, rbcL, rpl2, rpl14, rpl16, rpl20, rpl22, rpl23, rpl33, rpl36, rpoA, rpoB, rpoC1, rpoC2, rps2, rps3, rps4, rps7, rps8, rps11, rps12, rps14, rps15, rps18, rps19, ycf2, ycf3,* and *ycf4.*

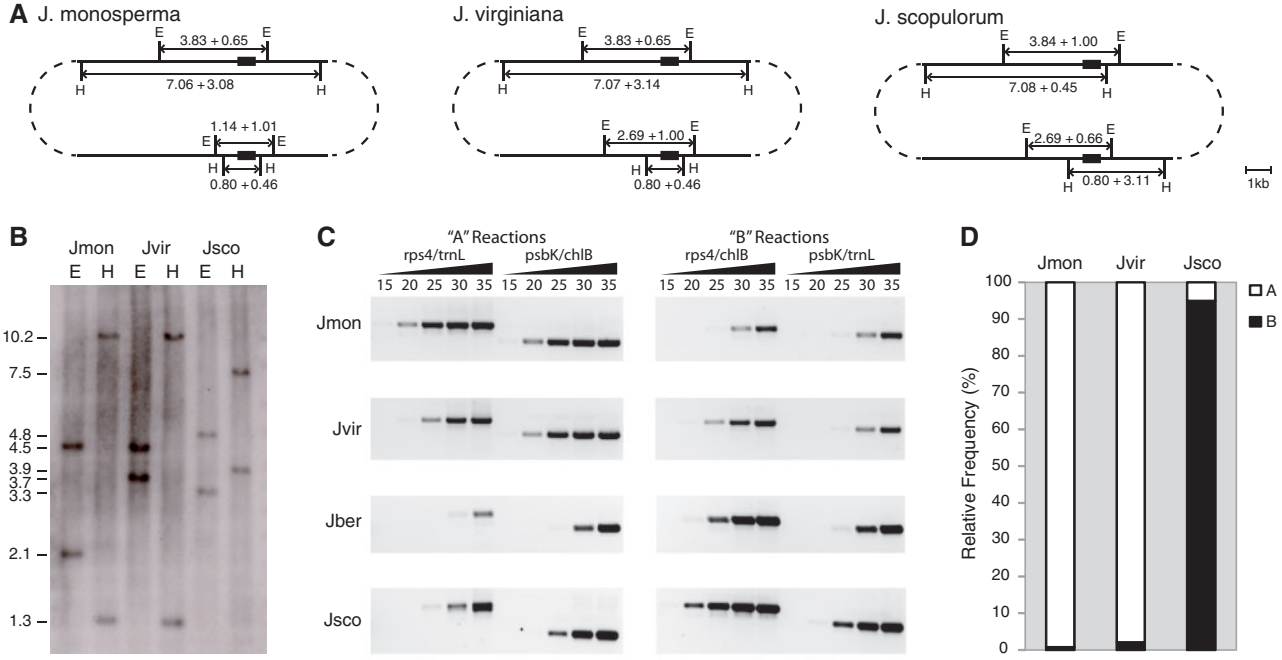

Fig. 2.—Analysis of IR arrangements in junipers. (*A*) Map of restriction enzyme digestion sites around the IR. EcoRI and HindIII digestion sites are indicated by E and H. Solid lines were drawn to scale. Genomic distances between digestion sites and center of the repeats are shown. Repeats are depicted by the solid boxes (not in scale). (*B*) Southern blot analysis of repeat arrangements. Sizes of each restriction fragment are shown on the left (in kilobases). (*C*) PCR analysis of repeat arrangements at different cycle numbers. The "A reactions" and "B reactions" label the primer combinations designed to amplify products from "A" and "B" plastomic arrangements. Numbers above the panels indicate PCR cycle number. (*D*) Relative frequencies of the isomeric genomic arrangements based on read-pair mapping. Heights of open and solid bars indicate relative frequencies of "A" and "B" arrangements, respectively.

each species. However, we cannot rule out the possibility that PCR-mediated recombination is generating artifactual positive results in the semiquantitative PCR results. It is also possible that PCR-mediated recombination or in vivo asymmetric recombination is contributing to the observed differences in amplification efficiency. Because quantitative PCR would be expected to be equally prone to such recombinational issues, we did not use this technique.

As another approach, studies of plant mitochondrial genomes have shown that substoichiometric recombinant forms can be detected and quantified using high-throughput read-pair mapping data (Alverson et al. 2011; Davila et al. 2011; Mower et al. 2012; Sloan et al. 2012). To test for the substoichiometric presence of the isomeric plastomic arrangement within an individual plant, we searched for Illumina paired-end reads that mapped inconsistently to the predominant genomic arrangement in *J. monosperma*, *J. scopulorum* and *J. virginiana* (*J. bermudiana* was not tested because we lacked paired-end sequencing data). For all three examined junipers, we found a small number of read pairs

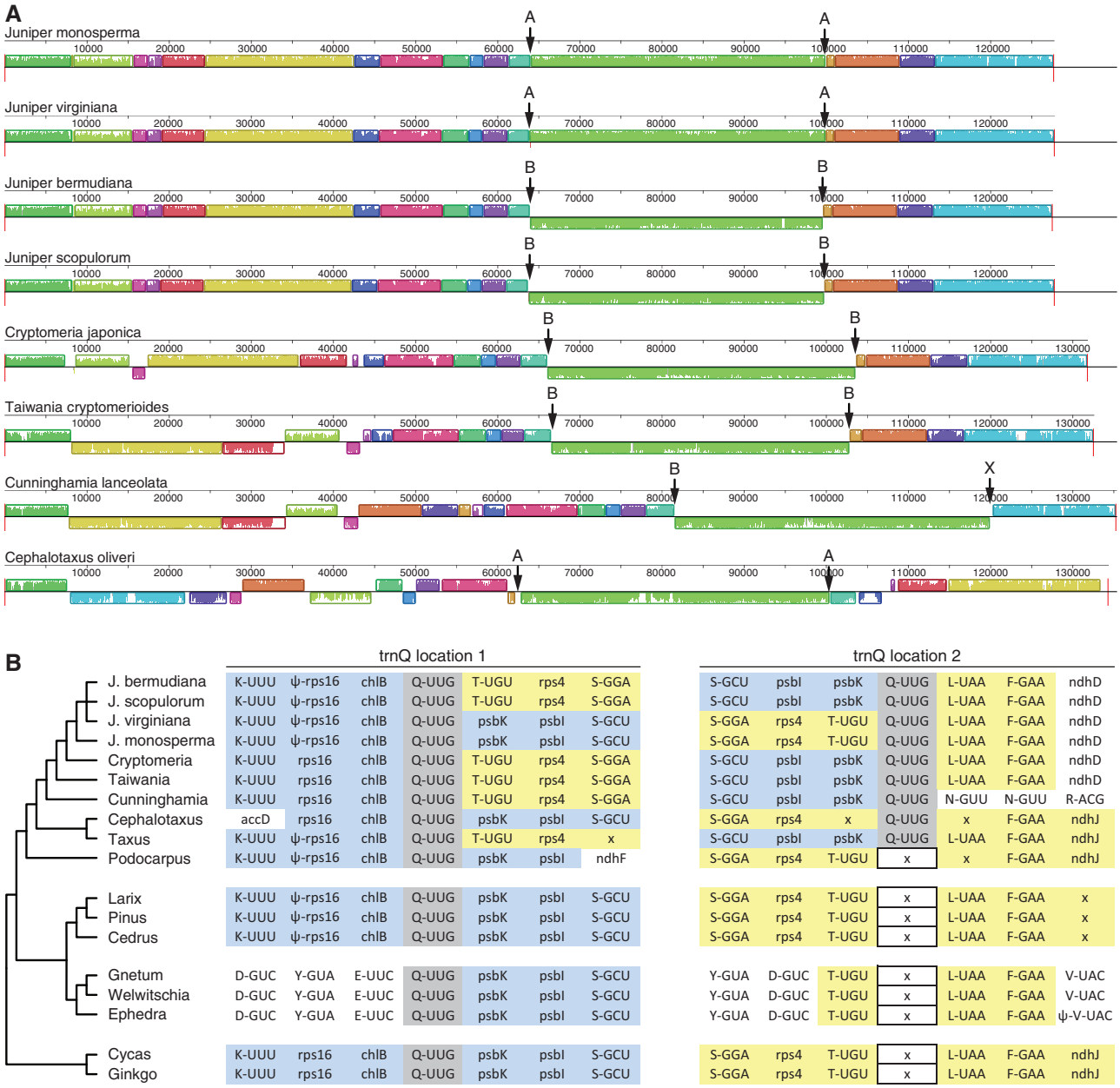

Fig. 3.—Cupressophyte plastome structural alignments. (*A*) Mauve alignments. The colored blocks represent collinear sequence blocks shared by all plastomes. Blocks drawn below the horizontal line indicate sequences found in inverted orientation. The height of the colored bars within each block reflects the level of sequence similarity among plastomes. Arrows indicate boundaries of the inversion segment and their associated letter ("A" or "B") indicate the orientation of the inversion relative to the flanking sequences, while an "X" indicates a novel flanking sequence. (*B*) Gene context of *trnQ*-IR. Light blue shading indicates the inferred ancestral gene synteny of the original *trnQ-UUG* gene, while light yellow shading indicates the ancestral synteny of the location of the second *trnQ-UUG* gene copy. Unshaded genes indicate a loss of synteny. An "x" indicates a missing gene. The phylogeny was based on currently accepted relationships (Zhong et al. 2011; Mao et al. 2012; Adams and Schwarzbach 2013).

that supported the isomeric genomic arrangement (fig. 2*D*). For *J. virginiana*, there were 190 read pairs that spanned the *trnQ*-containing IR copies, of which 186 pairs (97.9%) supported the predominant "A" arrangement while 4 pairs (2.1%) supported the isomeric "B" arrangement. Similar calculations showed that the isomeric plastomic arrangements were present at 5.0% (6/119) frequency in *J. scopulorum* and 0.8% (1/132) frequency in *J. monosperma*. To ensure that these low-frequency read pairs supporting the recombinant conformation are not due to cross-contamination, we mapped the raw reads from all three species to three variable regions in the genomes (supplementary fig. S2,

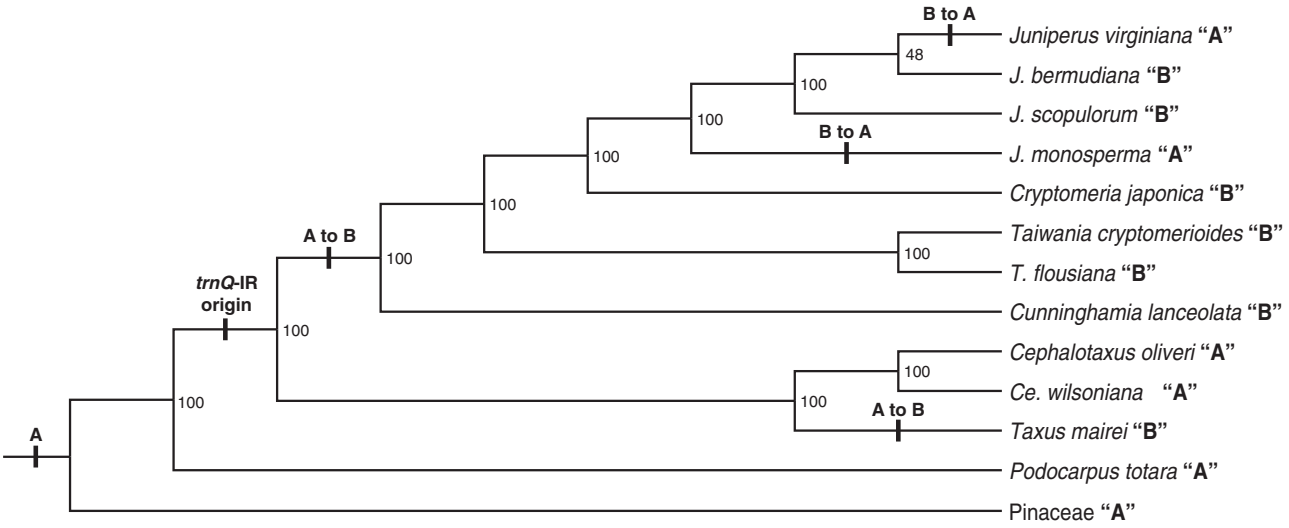

Fig. 4.—Phylogenetic analysis of cupressophyte plastid genes. The tree was generated by maximum likelihood method of a data set containing 83 plastid protein genes from 21 conifers and was rooted on Pinaceae. Bootstrap values are given for each branch, including the <50% value for the relationship among *J. bermudiana*, *J. scopulorum*, and *J. virginiana*. The "A" or "B" after each species indicates its plastome is completely or predominantly oriented in the "A" or "B" genomic arrangement. The origin of the *trnQ*-IR and "A" type of genomic arrangement, as well as evolutionarily shifts between "A" and "B" genomic arrangements, are mapped on branches.

Supplementary Material online). Of the 600 reads that were mapped, all but one of the reads were matched to the expected species rather than to the other species, indicating that there is very little cross-contamination in the data sets (~0.17%).

Frequent Shifting of Isomeric Genomic Forms in Cupressophytes

A homologous *trnQ*-containing IR is also present in other sequenced cupressophyte species from Cupressaceae, Cephalotaxaceae, and Taxaceae. To examine the genomic arrangement of the inversion segment among cupressophyte plastomes, we aligned a representative set of genomes with Mauve (fig. 3*A*). Most cupressophytes have the inversion segment in the "B" arrangement, except for the two *Juniperus* species reported here and for *Ce. oliveri*, which exists predominantly in the "A" arrangement based on previous PCR results (Yi et al. 2013). The discontinuous presence of "A" and "B" forms among cupressophytes indicates that the intervening segment has inverted multiple times during cupressophyte evolution. Although the "B" arrangement predominates in cuppressophyte plastomes, inspection of the *Podocarpus*, Gnetales, Pinaceae, *Ginkgo* and *Cycas* plastomes shows that the ancestral gene orders of the two *trnQ-UUG* locations are collinear with the "A" arrangement (fig. 3*B*).

To understand the evolution of plastid genome structure in cupressophytes, we first determined the phylogenetic relationships among the four juniper species and other sequenced cupressophytes based on plastomic data. The resulting phylogeny (fig. 4) shows a monophyletic *Juniperus* clade in which *J. monosperma* split earliest from other three species, as expected (Adams and Schwarzbach 2013), but relationships among *J. bermudiana*, *J. scopulorum*, and *J. virginiana* were not resolved. A closer relationship between *J. bermudiana* and *J. virginiana* has weak bootstrap support (BS = 48%) in our analysis, although previous analyses suggested a closer affinity between *J. virginiana* and *J. scopulorum* (Adams and Schwarzbach 2013). For other cupressophytes, the strongly supported topology within Cupressaceae and among the four families (Cephalotaxaceae, Cupressaceae, Podocarpaceae, and Taxaceae) is fully congruent with currently accepted organismal relationships (Mao et al. 2012; http://www.mobot.org/MOBOT/research/APweb/, last accessed March 11, 2014).

We used parsimony to map the origin of the *trnQ*-IR and the changes in orientation of the inversion segment onto the cupressophyte phylogeny (fig. 4). Based on the absence of this IR in *Podocarpus totara* and its presence in all other examined cupressophytes, we infer that the IR originated in the common ancestor of Cupressaceae, Cephalotaxaceae, and Taxaceae after they split from Podocarpaceae. Once the IR was established, it is clear that the intervening segment has indeed inverted multiple times during cupressophyte evolution. Based on the phylogenetic relationship of the four juniper species, at least two inversion events must have occurred to generate the distribution of arrangements. A minimum of two additional inversions are further required to explain the variable inversion arrangement in other cupressophytes. Given the number of inversions already identified, it is conceivable that this parsimony approach has underestimated the true number of inversion events that occurred during cupressophyte

diversification. It is also possible that rearrangements have occurred within species, which would increase the inferred number of events further. Finally, it should be noted that the inversion orientation has not been verified in some previously published species (*T. flousiana*, *Cunninghamia lanceolata*, and *Taxus mairei*), which would alter the inferred number and timing of rearrangement events. Importantly, however, these uncertainties would not alter (and in most cases would strengthen) our major conclusion of multiple shifts between "A" and "B" arrangements during cupressophyte evolution. Additional sampling of plastid genome structures and verification of previously published genomes will be necessary to more fully examine this issue.

Notably, there are now plastomes from five species (*J. bermudiana*, *J. monosperma*, *J. scopulorum*, *J. virginiana*, and *Ce. oliveri*) with evidence for the substoichiometric presence of the alternative isomeric arrangement within a single individual. Furthermore, by close inspection of the PCR results for the *C. japonica* and *T. cryptomerioides trnQ*-IR arrangements, we observe a very faint band for one of the alternative genomic arrangements (supplementary fig. S4 in Yi et al. 2013), suggesting that these species may also contain the isomeric form at a substoichiometric level. Thus, there is weak to compelling evidence for substoichiometric isomeric arrangements of the plastome in all seven cupressophyte species whose structures have been examined. Importantly, the predominant and substoichiometric plastomic arrangements have changed several times during cupressophyte evolution, which provides the first evidence for the process of SSS activity in a plastid genome.

Discussion

Southern mapping studies of large repeats (>1,000 bp) in plant organellar genomes have shown that all of the possible genomic arrangements generated by repeat-mediated homologous recombination are present at roughly equal stoichiometry (Palmer 1983; Palmer and Shields 1984; Stein et al. 1986; Klein et al. 1994). These findings imply that high-frequency homologous recombination maintains the various repeat arrangements in stoichiometric equilibrium (Lonsdale et al. 1988; Woloszynska 2010). Low-frequency recombination at intermediately sized repeats (~50–1,000 bp) is well documented in plant mitochondrial genomes (Woloszynska 2010; Arrieta-Montiel and Mackenzie 2011) and is becoming increasingly reported in plastomes (Tsumura et al. 2000; Gray et al. 2009; Wu et al. 2011; Xu et al. 2011; Yi et al. 2013). Because recombination at these intermediate repeats is infrequent, the rearranged genomic forms are present at substoichiometric levels relative to the predominant genomic form, and the substoichiometric forms are often undetectable by southern blotting (fig. 2*B*; Small et al. 1987; Janska et al. 1998; Tsumura et al. 2000; Shedge et al. 2007).

In plant mitochondria, the process of SSS enables a substoichiometric genomic form to become the predominant form through increased recombination activity at intermediate repeats, and possibly also by selective amplification (Woloszynska 2010; Arrieta-Montiel and Mackenzie 2011). Mitochondrial SSS activity appears to be the major driver of genomic structural diversification within and between angiosperms (Small et al. 1987; Arrieta-Montiel et al. 2009; Davila et al. 2011), and it can also result in phenotypic changes, some of which may be adaptive (Arrieta-Montiel and Mackenzie 2011). The most common phenotypic effect of SSS is cytoplasmic male sterility (CMS) (Sandhu et al. 2007), which causes normally hermaphroditic plants to fail to produce viable pollen. The regulatory control of CMS and fertility restoration is considered an evolutionarily adaptive strategy to modulate the relative numbers of hermaphrodites and females in a gynodioecious population (Delph et al. 2007; Caruso et al. 2012). More recently, mitochondrial SSS activity has been associated with increased height and delayed flowering (Albert et al. 2003) and with leaf variegation and thermotolerance (Shedge et al. 2010), although it is not clear whether these phenotypes are directly caused by the mitochondrial rearrangements.

In plastomes, although low-frequency recombination at intermediate repeats is becoming apparent, there have been no previous reports of SSS activity that has switched the predominant and substoichiometric forms between individuals or species, nor has it been established convincingly that different genomic forms exist within individual plants. Here, we showed the substoichiometric presence of isomeric genomic forms of the plastid genome within four individual junipers, and, more importantly, we showed that the predominant genomic form has changed multiple times during the evolutionary diversification of cupressophytes. Thus, our results indicate that SSS activity can drive plastid genomic diversity among species. As SSS in mitochondria is known to act quickly (within one or two generations), it will be necessary to explore whether isomeric variation exists between individuals within a species to firmly establish that SSS is active in cupressophyte plastids. Given our findings of variation within individuals and between species, we fully expect that plastomic SSS activity can also promote genomic diversity among individuals of the same species.

As we know that mitochondrial SSS can have important phenotypic consequences on a plant, our discovery of plastomic SSS raises the intriguing possibility that it may also cause phenotypic changes of evolutionary significance. Recent studies of *Arabidopsis thaliana msh1* or *why1/why3* mutants have shown that the plastid genome undergoes infrequent genomic rearrangement and the plants exhibit slow growth, delayed flowering, altered leaf morphology, and leaf variegation (Marechal et al. 2009; Cappadocia et al. 2010; Xu et al. 2012). Importantly, leaf variegation may provide thermotolerance in high light conditions. Thus, it appears that plastomic

rearrangements can indeed elicit phenotypic changes of adaptive significance. In junipers, the genomic inversion we identified would seem to have little effect on plastid gene function, and so it would be unlikely to influence phenotype. Nevertheless, it is possible that the combination of polycistronic transcriptional units created by the arrangement of the inversion may provide some selective benefit in particular environments. Alternatively, it may be that the act of genomic rearrangement generates retrograde signals which are acted on by the nucleus to promote a physiological response.

Supplementary Material

Supplementary figures S1 and S2 and table S1 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

The authors thank Yashitola Wamboldt and Kempton Horken for providing technical help with the Southern blot and PCR experiments. This work was supported in part by the Baylor University grant BU-044-4512 to R.P.A. and the National Science Foundation grants DEB-0629402 to A.E.S., and IOS-1027529 and MCB-1125386 to J.P.M.

Literature Cited

Adams RP, Schwarzbach AE. 2013. Phylogeny of *Juniperus* using nrDNA and four cpDNA regions. Phytologia 95:179–187.

Albert B, et al. 2003. Amplification of *Nicotiana sylvestris* mitochondrial subgenomes is under nuclear control and is associated with phenotypic changes. Genetica 117:17–25.

Alverson AJ, Zhuo S, Rice DW, Sloan DB, Palmer JD. 2011. The mitochondrial genome of the legume *Vigna radiata* and the analysis of recombination across short mitochondrial repeats. PLoS One 6:e16404.

Arrieta-Montiel MP, Mackenzie SA. 2011. Plant mitochondrial genomes and recombination. In: Kempken F, editor. Plant mitochondria. New York: Springer. p. 65–82.

Arrieta-Montiel MP, Shedge V, Davila J, Christensen AC, Mackenzie SA. 2009. Diversity of the *Arabidopsis* mitochondrial genome occurs via nuclear-controlled recombination activity. Genetics 183:1261–1268.

Cappadocia L, et al. 2010. Crystal structures of DNA-Whirly complexes and their role in *Arabidopsis* organelle genome repair. Plant Cell 22: 1849–1867.

Caruso CM, Case AL, Bailey MF. 2012. The evolutionary ecology of cytonuclear interactions in angiosperms. Trends Plant Sci. 17:638–643.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 17: 540–552.

Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One 5: e11147.

Davila JI, et al. 2011. Double-strand break repair processes drive evolution of the mitochondrial genome in *Arabidopsis*. BMC Biol. 9:64.

Delph LF, Touzet P, Bailey MF. 2007. Merging theory and mechanism in studies of gynodioecy. Trends Ecol Evol. 22:17–24.

Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull. 19:11–15.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.

Gray BN, Ahner BA, Hanson MR. 2009. Extensive homologous recombination between introduced and native regulatory plastid DNA elements in transplastomic plants. Transgenic Res. 18:559–572.

Grewe F, Guo W, Gubbels EA, Hansen AK, Mower JP. 2013. Complete plastid genomes from *Ophioglossum californicum*, *Psilotum nudum*, and *Equisetum hyemale* reveal an ancestral land plant genome structure and resolve the position of Equisetales among monilophytes. BMC Evol Biol. 13:8.

Hirao T, Watanabe A, Kurita M, Kondo T, Takata K. 2008. Complete nucleotide sequence of the *Cryptomeria japonica* D. Don. chloroplast genome and comparative chloroplast genomics: diversified genomic structure of coniferous species. BMC Plant Biol. 8:70.

Jansen RK, Ruhlman TA. 2012. Plastid genomes of seed plants. In: Bock R, Knoop V, editors. Genomics of chloroplasts and mitochondria. Dordrecht (The Netherlands): Springer. p. 103–126.

Janska H, Sarria R, Woloszynska M, Arrieta-Montiel M, Mackenzie SA. 1998. Stoichiometric shifts in the common bean mitochondrial genome leading to male sterility and spontaneous reversion to fertility. Plant Cell 10:1163–1180.

Katoh K, Toh H. 2008. Recent developments in the MAFFT multiple sequence alignment program. Brief Bioinform. 9:286–298.

Klein M, et al. 1994. Physical mapping of the mitochondrial genome of *Arabidopsis thaliana* by cosmid and YAC clones. Plant J. 6:447–455.

Lahr DJ, Katz LA. 2009. Reducing the impact of PCR-mediated recombination in molecular evolution and environmental studies using a new-generation high-fidelity DNA polymerase. Biotechniques 47:857–866.

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

Lin CP, Huang JP, Wu CS, Hsu CY, Chaw SM. 2010. Comparative chloroplast genomics reveals the evolution of Pinaceae genera and subfamilies. Genome Biol Evol. 2:504–517.

Lin CP, Wu CS, Huang YY, Chaw SM. 2012. The complete chloroplast genome of *Ginkgo biloba* reveals the mechanism of inverted repeat contraction. Genome Biol Evol. 4:374–381.

Lonsdale DM, Brears T, Hodge TP, Melville SE, Rottmann WH. 1988. The plant mitochondrial genome: homologous recombination as a mechanism for generating heterogeneity. Philos Trans R Soc Lond B Biol Sci. 319:149–163.

Mao K, et al. 2012. Distribution of living Cupressaceae reflects the breakup of Pangea. Proc Natl Acad Sci U S A. 109:7793–7798.

Marechal A, et al. 2009. Whirly proteins maintain plastid genome stability in *Arabidopsis*. Proc Natl Acad Sci U S A. 106:14693–14698.

Mower JP, Case AL, Floro ER, Willis JH. 2012. Evidence against equimolarity of large repeat arrangements and a predominant master circle structure of the mitochondrial genome from a monkeyflower (*Mimulus guttatus*) lineage with cryptic CMS. Genome Biol Evol. 4: 670–686.

Palmer JD. 1983. Chloroplast DNA exists in two orientations. Nature 301: 92–93.

Palmer JD, Shields CR. 1984. Tripartite structure of the *Brassica campestris* mitochondrial genome. Nature 307:437–440.

Sambrook J, Russell DW 2001. Molecular cloning: a laboratory manual, 3rd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory.

Sandhu AP, Abdelnoor RV, Mackenzie SA. 2007. Transgenic induction of mitochondrial rearrangements for cytoplasmic male sterility in crop plants. Proc Natl Acad Sci U S A. 104:1766–1770.

Shedge V, Arrieta-Montiel M, Christensen AC, Mackenzie SA. 2007. Plant mitochondrial recombination surveillance requires unusual *RecA* and *MutS* homologs. Plant Cell 19:1251–1264.

Shedge V, Davila J, Arrieta-Montiel MP, Mohammed S, Mackenzie SA. 2010. Extensive rearrangement of the *Arabidopsis* mitochondrial genome elicits cellular conditions for thermotolerance. Plant Physiol. 152:1960–1970.

Sloan DB, et al. 2012. Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. PLoS Biol. 10:e1001241.

Small ID, Isaac PG, Leaver CJ. 1987. Stoichiometric differences in DNA molecules containing the *atpA* gene suggest mechanisms for the generation of mitochondrial genome diversity in maize. EMBO J. 6: 865–869.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690.

Stein DB, Palmer JD, Thompson WF. 1986. Structural evolution and flip-flop recombination of chloroplast DNA in the fern genus *Osmunda*. Curr Genet. 10:835–841.

Tsumura Y, Suyama Y, Yoshimura K. 2000. Chloroplast DNA inversion polymorphism in populations of *Abies* and *Tsuga*. Mol Biol Evol. 17: 1302–1312.

Vaidya G, Lohman DJ, Meier R. 2011. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. Cladistics 27:171–180.

Wicke S, Schneeweiss GM, dePamphilis CW, Muller KF, Quandt D. 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol Biol. 76:273–297.

Woloszynska M. 2010. Heteroplasmy and stoichiometric complexity of plant mitochondrial genomes—though this be madness, yet there's method in't. J Exp Bot. 61:657–671.

Wu CS, Chaw SM. 2013. Highly rearranged and size-variable chloroplast genomes in conifers II clade (cupressophytes): evolution towards shorter intergenic spacers. Plant Biotechnol J. Advance Access published November 28, 2013, doi: 10.1111/pbi.12141.

Wu CS, Lai YT, Lin CP, Wang YN, Chaw SM. 2009. Evolution of reduced and compact chloroplast genomes (cpDNAs) in gnetophytes: selection toward a lower-cost strategy. Mol Phylogenet Evol. 52:115–124.

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.

Wu CS, Wang YN, Liu SM, Chaw SM. 2007. Chloroplast genome (cpDNA) of *Cycas taitungensis* and 56 cp protein-coding genes of *Gnetum parvifolium*: insights into cpDNA evolution and phylogeny of extant seed plants. Mol Biol Evol. 24:1366–1379.

Wyman SK, Jansen RK, Boore JL. 2004. Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20:3252–3255.

Xu YZ, et al. 2011. MutS HOMOLOG1 is a nucleoid protein that alters mitochondrial and plastid properties and plant response to high light. Plant Cell 23:3428–3441.

Xu YZ, et al. 2012. The chloroplast triggers developmental reprogramming when mutS HOMOLOG1 is suppressed in plants. Plant Physiol. 159: 710–720.

Yi X, Gao L, Wang B, Su YJ, Wang T. 2013. The complete chloroplast genome sequence of *Cephalotaxus oliveri* (Cephalotaxaceae): evolutionary comparison of *Cephalotaxus* chloroplast DNAs and insights into the loss of inverted repeat copies in gymnosperms. Genome Biol Evol. 5:688–698.

Zaegel V, et al. 2006. The plant-specific ssDNA binding protein OSB1 is involved in the stoichiometric transmission of mitochondrial DNA in *Arabidopsis*. Plant Cell 18:3548–3563.

Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18: 821–829.

Zhong B, et al. 2011. Systematic error in seed plant phylogenomics. Genome Biol Evol. 3:1340–1348.

Associate editor: Shu-Miaw Chaw