

The plastome sequence of the endemic Amazonian conifer, *Retrophyllum piresii* (Silba) C.N.Page, reveals different recombination events and plastome isoforms

Leila do Nascimento Vieira¹ · Marcelo Rogalski² · Helisson Faoro^{3,4} · Hugo Pacheco de Freitas Fraga¹ · Karina Goulart dos Anjos¹ · Gisele Fernanda Assine Picchi³ · Rubens Onofre Nodari¹ · Fábio de Oliveira Pedrosa⁴ · Emanuel Maltempi de Souza⁴ · Miguel Pedro Guerra¹

Received: 17 August 2015 / Revised: 4 January 2016 / Accepted: 8 January 2016 / Published online: 20 January 2016
© Springer-Verlag Berlin Heidelberg 2016

Abstract *Retrophyllum piresii* (Podocarpaceae) is an endemic conifer species from the Brazilian Amazonian Region, and very few data related to ecological and genetic characteristics of this species are available. Plastome sequencing is an efficient tool to understand enigmatic and basal phylogenetic relationships at different taxonomic levels, as well as to probe the structural and functional evolution of plants. Usually, the plastome of photosynthetic land plants is quadripartite, with two copies of the inverted repeats (IRs) separating the small and large single-copy regions. However, in gymnosperms, IR

can vary from large in size to completely absent, being constituted principally by transfer RNA (tRNA) genes, or a part of sequence of other genes. Here, we sequenced and characterized the complete plastome of *R. piresii*. This plastome was determined to be 133,291 bp (~480-fold coverage), presenting a total of 120 identified genes, of which 118 were single copy and two genes, *trnN-GUU* and *trnD-GUC*, were found to be duplicated and occurring as inverted and directed repeat (DR) sequences, respectively. These repeated regions presented recombinationally active sites, resulting in an IR-mediated inversion and a DR-mediated deletion. However, the isoform resulted from DR-mediated deletion may result in unviable plastome, with deletion of photosynthetic and expression machinery-related genes.

Communicated by Y. Tsumura

Leila do Nascimento Vieira and Marcelo Rogalski contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s11295-016-0968-0) contains supplementary material, which is available to authorized users.

✉ Miguel Pedro Guerra
miguel.guerra@ufsc.br

¹ Laboratório de Fisiologia do Desenvolvimento e Genética Vegetal, Programa de Pós-graduação em Recursos Genéticos Vegetais, Universidade Federal de Santa Catarina, Rodovia Admar Gonzaga, 1346, ZC: 88040-900 Florianópolis, Santa Catarina, Brazil

² Departamento de Biologia Vegetal, Univesidade Federal de Viçosa, Avenida Peter Henry Rolfs s/n, ZC: 36570-000 Viçosa, Minas Gerais, Brazil

³ Instituto Carlos Chagas, Fundação Oswaldo Cruz, Rua Professor Algacyr Munhoz Mader, 3775, bloco C 81350010 Curitiba, Paraná, Brazil

⁴ Departamento de Bioquímica e Biologia Molecular, Núcleo de Fixação Biológica de Nitrogênio, Universidade Federal do Paraná, Rua Coronel Francisco H. dos Santos s/n, ZC: 81531-980 Curitiba, Paraná, Brazil

Keywords Chloroplast genome · Evolution · Homologous recombination · Gymnosperms · Podocarpaceae

Introduction

Extant gymnosperms are considered the most ancient group of seed-bearing plants that first appeared, approximately 300 million years ago (Murray 2013). They consist of four major groups, including gnetophytes, conifers, cycads, and ginkgo. The Podocarpaceae family is considered the most diverse family of conifers, comprising 173 species in 18 genera, which are mainly distributed in the Southern Hemisphere, extending also to the north in subtropical China, Japan, Mexico, and the Caribbean (Farjon 1998; Biffin et al. 2011). The *Retrophyllum* genus comprises five species: *Retrophyllum comptonii*, *R. minor*, *Retrophyllum piresii*, *R. rospigliosii*, and *R. vitiense*. The endemic species from highlands of Pacaás Novos

National Park, Brazil, *R. piresii* was classified in 1976 by João Murça Pires, who collected seeds, which were germinated and the plants maintained in the Botanical Garden Museu Paraense Emílio Goeldi, Belém, Pará, Brazil. Nowadays, very few data related to physiological, ecological, and genetic characteristics of this species are known.

Plastid genome (plastome) sequencing is an efficient tool for increasing phylogenetic resolution at lower taxonomic levels in plant phylogenetic and genetic population analyses (Besnard et al. 2011; Dexter et al. 2012; López et al. 2012; Rogalski et al. 2015). It has been used to understand enigmatic and basal phylogenetic relationships at different taxonomic levels, being sources of structural and functional information about the evolution of the different groups of plants (Jansen et al. 2007; Moore et al. 2007; Parks et al. 2009; Moore et al. 2010; Wu et al. 2011; Yi et al. 2013; Vieira et al. 2014a). Plastome sequences are available for all families of conifers: Cephalotaxaceae (Yi et al. 2013), Cupressaceae (Hirao et al. 2008), Pinaceae (Wakasugi et al. 1994; Cronn et al. 2008; Lin et al. 2010), Taxaceae (Zhang et al. 2014), Araucariaceae (Wu and Chaw 2014; Ruhsam et al. 2015), and Podocarpaceae (Wu and Chaw 2014; Vieira et al. 2014a). For Podocarpaceae family, the plastome sequence has recently been revealed for three species: the endemic New Zealand *Podocarpus totara* G. Benn. ex Don (NC_020361.1), *Podocarpus lambertii*, a species from the biodiversity hotspot of South America, the Araucaria forest (Vieira et al. 2014a), and the Asiatic species *Nageia nagi* (Wu and Chaw 2014).

Usually, the plastome of photosynthetic land plants is 120–220 kb in size, with two copies of the inverted repeats (IRs) separating the small and large single-copy (SSC and LSC) regions (Palmer 1983; Knox 2014). The size of IRs in plastids of land plants is highly variable and it is dependent on plant group, genus, family, or species (Wicke et al. 2011; Jansen and Ruhlman 2012; Guo et al. 2014; Gurdon and Maliga 2014; Vieira et al. 2014a). The IR copies recombine themselves and are intended to maintain or confer stability to the remaining plastome (Palmer 1983; Stein et al. 1986; Knox 2014).

In gymnosperms, IR size ranges from large to completely absent. Different taxonomic orders as Gnetales, Cycadales, and Ginkgoales have retained the classical IRs, which can range from 17.3 to 25.1 kb (Wu et al. 2007, 2009; Lin et al. 2012; Guo et al. 2014). In conifers, there are short IR regions, containing different genes, but principally transfer RNA (tRNA) genes or a part of other gene sequence. Recently, in species of the *Juniperus* genus, the presence of short IRs containing two copies of full *trnQ*-UUG (Guo et al. 2014) was observed. These short IR sequences (~250 bp) showed to be able to recombine and create different isoforms of plastome, which have been proven to happen in different individual plants and in different tissues of the same plant (Guo et al. 2014). Gurdon and Maliga (2014) reported an unprecedented presence of two plastome configurations, with ~45 kb

inversion, produced by recombination of short imperfect inverted sequences containing 20–24 bp in different *Medicago truncatula* ecotypes.

In transgenic plastids, the presence of inverted or direct repeats produced by using short endogenous plastid 5'- or 3'-UTRs as signals for expression cassettes was demonstrated to generate different plastome isoforms (Rogalski et al. 2006; 2008a; 2008b; Gray et al. 2009; Alkatib et al. 2012). The genome rearrangement is dependent on the direction of the two repeated sequences (i.e., directed repeats or inverted repeats). Whether the sequences are presented as directed repeats, the sequence between them and one of them are deleted from the plastome (Rogalski et al. 2008b; Alkatib et al. 2012), whereas if the sequences are found as inverted repeats, they recombine and induce an inversion of the sequence between them (Rogalski et al. 2006; 2008a).

Here, we demonstrated the presence of recombinationally active repeated sequences, consisting of different copies of tRNA genes, one as inverted and the other as directed repeat in the same plastome. These repeated sequences produce an IR-mediated inversion and a directed repeat (DR)-mediated deletion, resulting in different plastome arrangements. However, the isoform created by DR-mediated deletion may produce an unviable plastome, with deletion of photosynthetic genes and other genes involved in plastid gene expression machinery.

Results and discussion

Retrophyllum piresii plastome size and gene content

R. piresii plastome size was determined to be 133,291 bp, only 443 bp smaller than *P. lambertii* (133,734 bp; NC_023805) and 431 bp smaller than *N. nagi* (133,722; NC_023120). The plastome size of Podocarpaceae species is consistent with other non-Pinaceae species (conifer clade II), which present a plastome ranging from 127,311 bp in *Calocedrus formosana* (NC_023121) to 145,625 bp in *Agathis dammara* (NC_023119). Otherwise, they are larger than the sequenced plastomes of Pinaceae species (conifer clade I), which range from 107,122 bp in *Cathaya argyrophylla* (NC_014589) to 124,168 bp in *Picea morrisonicola* (NC_016069), and smaller than the cycads *Cycas taitungensis* (163,403 bp; NC_009618) and *Cycas Revoluta* (162,489 bp; NC_020319). The GC content determined for *R. piresii* plastome is 37.25 %, which is very similar to other Podocarpaceae species *P. lambertii* (37.10 %) and *N. nagi* (37.26 %).

A total of 120 genes were identified in the *R. piresii* plastome, of which 118 were single copy and two genes, *trnN-GUU* and *trnD-GUC*, were found to be duplicated and occurring as inverted and directed repeat sequences, respectively. The following genes were identified and are listed in

Fig. 1 and Table 1: 4 ribosomal RNA genes, 31 unique transfer RNA genes, 20 genes encoding large and small ribosomal subunits, 1 translational initiation factor, 4 genes encoding DNA-dependent RNA polymerases, 50 genes encoding photosynthesis-related proteins, 8 genes encoding other proteins, including the unknown function gene *ycf2*, and 1 pseudogene, *ycf68*

Among these 118 single copy genes, 13 were genes containing introns (Table 1). Even though the *R. piresii* and *N. nagi* gene content is strictly similar to *P. lambertii*, they lost the *rpoC1* intron (Wu and Chaw 2014; Vieira et al. 2014a). In addition, a double copy of *trnD-GUC* is an exception only present in *R. piresii* plastome. The *rps16* is also absent in the *R. piresii* plastome, indicating that Podocarpaceae and Araucariaceae families have lost this gene during the evolution process, while other non-Pinaceae species did not (Wu et al. 2007; Hirao et al. 2008; Wu et al. 2011; Yi et al. 2013; Wu and Chaw 2014; Vieira et al. 2014a). Although this gene was shown to be essential for cell survival in tobacco, an angiosperm species (Fleischmann et al. 2011), it is also absent or nonfunctional in other gymnosperms, such as Pinaceae and Gnetophyte species (Tsudzuki et al. 1992; Wu

et al. 2007; 2009), and also in some angiosperms, such as species from Fabaceae (Guo et al. 2007; Tangphatsomruang et al. 2009), Dioscoreaceae (Hansen et al. 2007), and Melanthiaceae (Do et al. 2014) families.

Repeat sequence analysis

The plant population genetic studies may be greatly facilitated by the use of chloroplast DNA (cpDNA) markers due to its nonrecombinant, uniparentally inherited nature in most plant species, and low rates of mutation perceived in plastome (Powell et al. 1995; Provan et al. 2001). Plastome presents a conserved gene set and a general lack of heteroplasmy and recombination, which made it an attractive tool for plant phylogenetic studies. Furthermore, cpDNA may be applied to studies involving genetic structure of natural populations due to its mode of inheritance in comparison to nuclear markers (Provan et al. 2001).

Hence, chloroplast simple sequence repeat (SSR) has been widely used for high-resolution phylogeographic studies (Ahmed et al. 2013; Tomar et al. 2014). Other applications include the characterization of alloplasmic

Fig. 1 Gene map of *Retrophyllum piresii* plastome. Genes drawn inside the circle are transcribed clockwise, and genes drawn outside are counterclockwise. Genes belonging to different functional groups are color-coded. The darker gray in the inner circle corresponds to GC content, and the lighter gray to AT content. The location of the IR-mediated inversion and the DR-mediated deletion are highlighted on the outer circle by blue and red bars, respectively

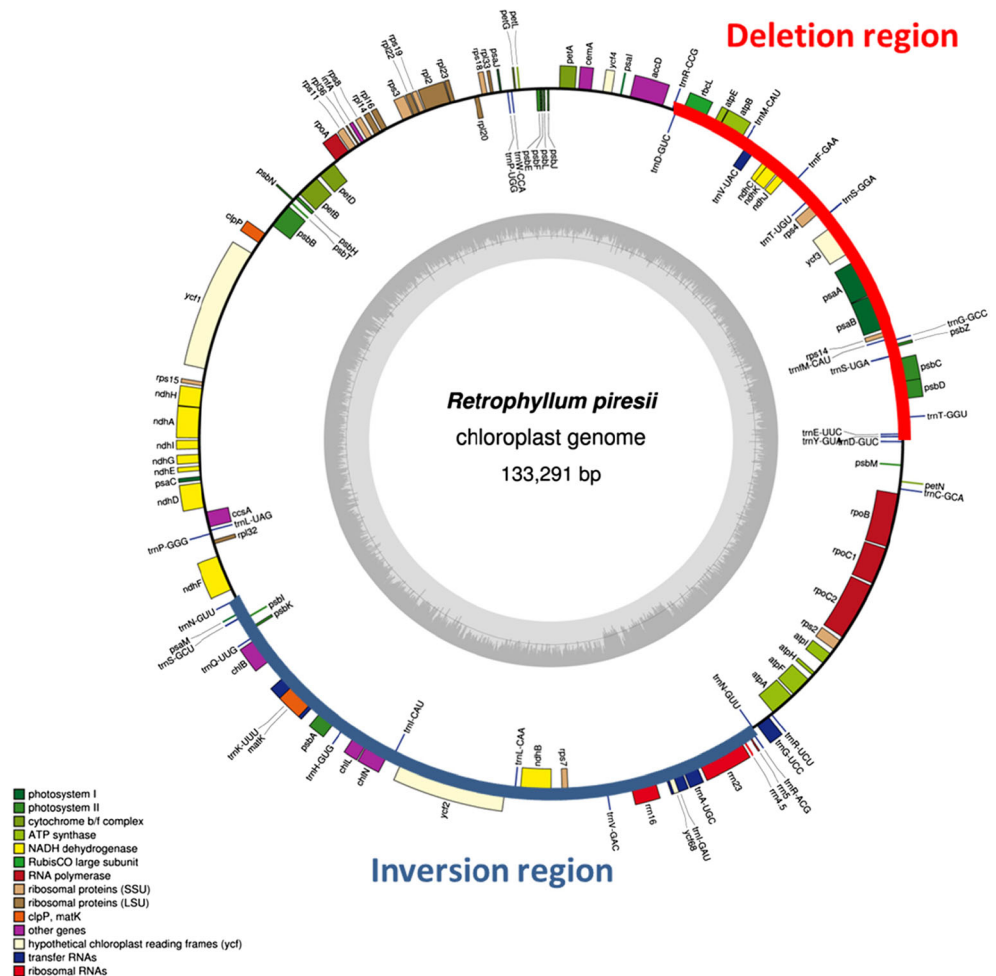


Table 1 List of genes identified in *Retrophyllum piresii* plastome

Category of genes	Group of gene	Name of gene				
	Ribosomal RNA genes	<i>rrn23</i>	<i>rrn4.5</i>	<i>rrn5</i>	<i>rrn16</i>	
	Transfer RNA genes	<i>trnA</i> -UGC ^a	<i>trnC</i> -GCA	<i>trnD</i> -GUC ^b	<i>trnE</i> -UUC	<i>trnF</i> -GAA
		<i>trnG</i> -GCC	<i>trnG</i> -UCC ^a	<i>trnH</i> -GUG	<i>trnI</i> -GAU ^a	<i>trnL</i> -CAU
		<i>trnL</i> -UAG	<i>trnL</i> -CAA	<i>trnK</i> -UUU ^a	<i>trnM</i> -CAU	<i>trnJ</i> M-CAU
		<i>trnN</i> -GUU ^b	<i>trnP</i> -UGG	<i>trnP</i> -GGG	<i>trnQ</i> -UUG	<i>trnR</i> -ACG
		<i>trnR</i> -UCU	<i>trnR</i> -CCG	<i>trnS</i> -GCU	<i>trnS</i> -UGA	<i>trnS</i> -GGA
		<i>trnT</i> -UGU	<i>trnT</i> -GGU	<i>trnV</i> -GAC	<i>trnV</i> -UAC ^a	<i>trnW</i> -CCA
		<i>trnY</i> -GUA				
	Small subunit of ribosome	<i>rps2</i>	<i>rps3</i>	<i>rps4</i>	<i>rps7</i>	<i>rps8</i>
		<i>rps11</i>	<i>rps12</i> ^a	<i>rps14</i>	<i>rps15</i>	<i>rps18</i>
		<i>rps19</i>				
	Large subunit of ribosome	<i>rpl2</i> ^a	<i>rpl14</i>	<i>rpl16</i>	<i>rpl20</i>	<i>rpl22</i>
		<i>rpl23</i>	<i>rpl32</i>	<i>rpl33</i>	<i>rpl36</i>	
	DNA-dependent RNA polymerase	<i>rpoA</i>	<i>rpoB</i>	<i>rpoC1</i>	<i>rpoC2</i>	
	Translational initiation factor	<i>infA</i>				
Genes for photosynthesis	Subunits of photosystem I	<i>psaA</i>	<i>psaB</i>	<i>psaC</i>	<i>psaI</i>	<i>psaJ</i>
		<i>psaM</i>	<i>ycf3</i> ^a	<i>ycf4</i>		
	Subunits of photosystem II	<i>psbA</i>	<i>psbB</i>	<i>psbC</i>	<i>psbD</i>	<i>psbE</i>
		<i>psbF</i>	<i>psbH</i>	<i>psbI</i>	<i>psbJ</i>	<i>psbK</i>
		<i>psbL</i>	<i>psbM</i>	<i>psbN</i>	<i>psbT</i>	<i>psbZ</i>
	Subunits of cytochrome	<i>petA</i>	<i>petB</i> ^a	<i>petD</i> ^a	<i>petL</i>	<i>petN</i>
		<i>petG</i>				
	Subunits of ATP synthase	<i>atpA</i>	<i>atpB</i>	<i>atpE</i>	<i>atpF</i> ^a	<i>atpH</i>
		<i>atpI</i>				
	Large subunit of Rubisco	<i>rbcL</i>				
Chlorophyll biosynthesis	<i>chlB</i>	<i>chlL</i>	<i>chlN</i>			
Subunits of NADH dehydrogenase		<i>ndhA</i> ^a	<i>ndhB</i> ^a	<i>ndhC</i>	<i>ndhD</i>	<i>ndhE</i>
		<i>ndhF</i>	<i>ndhG</i>	<i>ndhH</i>	<i>ndhI</i>	<i>ndhJ</i>
		<i>ndhK</i>				
Other genes	Maturase	<i>matK</i>				
	Envelope membrane protein	<i>cemA</i>				
	Subunit of acetyl-CoA	<i>accD</i>				
	C-type cytochrome synthesis gene	<i>ccsA</i>				
	Protease	<i>clpP</i>				
	Component of TIC complex	<i>ycf1</i>				
Genes of unknown function	Conserved open reading frames	<i>ycf2</i>				
	Pseudogene	<i>ycf68</i>				

^a Genes containing introns^b Duplicated gene

lines in wheat (Tomar et al. 2014), the support of sweet potato domestications theory (Roullier et al. 2011), the distribution of genetic diversity in *Pinus pinaster* (Vendramin et al. 1998), the gene flow and hybridization among almond tree species (Delplancke et al. 2012), and the studies involving population genetic structure in different species (Kato et al. 2011; 2013; Roullier et al. 2013; Baskauf et al. 2014).

In the present study, we analyzed the occurrence and type of SSRs, consisting of tandemly repeated motifs of 6 bp or less in *R. piresii* plastome. In total, 168 SSRs were identified. Among them, homo- and dipolymers were the most common with, respectively, 96 and 62 occurrences, whereas tri- (2) and tetrapolymers (8) occurred with lower frequency (Table 2). Among the mono- and dipolymers identified, only 4 mono- and 1 dipolymer presented more than 15 repeats (Table 2),

Table 2 List of simple sequence repeats identified in *Retrophyllum piresii* plastome

SSR Sequence	Number of repeats																				Total	
	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22		23
A/T	–	–	–	–	–	43	18	10	5	3	4	1	1	–	2	–	–	–	–	–	1	88
C/G	–	–	–	–	–	2	4	–	1	1	–	–	–	–	–	–	–	–	–	–	–	8
AC/GT	–	3	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	3
AG/CT	–	23	–	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	24
AT/AT	–	18	4	1	3	2	1	2	2	1	–	–	–	–	–	–	–	–	–	–	1	35
AAG/CTT	–	2	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	2
AAAG/CTTT	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	1
AACT/AGTT	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	1
AATC/ATTG	2	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	2
ACCT/AGGT	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	1
AGAT/ATCT	3	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	3

which is in accordance to the nature of chloroplast microsatellites of generally <15 mononucleotide repeats (Provan et al. 2001). Penta- and hexapolymers were not identified in *R. piresii*, what differs from *P. lambertii* plastome, in which one penta- and one hexapolymer were identified (Vieira et al. 2014a), and from *N. nagi*, in which one pentapolymer was identified using the same parameters described in the “Material and methods” section (data not shown).

The homopolymers were mostly constituted by A/T sequences (91.66 %), but for dipolymers, only 56.45 % was constituted by multiple A and T bases. In *Colocasia* spp., the complete plastome sequence was used to identify polymorphic microsatellites suitable for high-resolution phylogeographic studies (Ahmed et al. 2013). The intraspecific sequence alignments revealed that polymorphic microsatellites were mostly mononucleotide A/T, and only one polymorphic, dinucleotide microsatellite AT/TA (Ahmed et al. 2013). Similarly, in wheat, 24 cpSSRs of the 25 polymorphic SSRs were mononucleotide A/T repeats, and only one was C/G repeat (Tomar et al. 2014).

In this study, we identified 158 repeats with one or two nucleotide repeat, totaling almost 94.5 % of all SSRs identified, most of them consisting of A/T sequences. These results reveal the presence of several SSR sites in *R. piresii* plastome that can be assessed for the intraspecific level of polymorphism, leading to innovative highly sensitive phylogeographic and population genetics studies for this species. This study may help to describe the conservation status of this species in its endemic region, Pácaás Novos National Park, Brazil.

Plastome structure

In land plants, most plastomes consist of large single-copy region (LSC), small single-copy region (SSC) and two inverted repeat regions (IR) (Palmer 1983; Shinozaki et al. 1986, Knox 2014). This plastome organization is highly

conserved in angiosperms, with very few exceptions (Guo et al. 2007; Hansen et al. 2007; Tangphatsornruang et al. 2009; Do et al. 2014; Gurdon and Maliga 2014). In gymnosperms, the loss of the large IR has been reported in several species, mainly in conifers (Hirao et al. 2008; Wu and Chaw 2014; Yi et al. 2013). Also, many rearrangements may be observed in the plastome, and such rearrangements appear to play an important role in their evolution (Wu and Chaw 2014; Yi et al. 2013; Vieira et al. 2014a). As in other species of Podocarpaceae family (Wu and Chaw 2014; Vieira et al. 2014a), the plastome of *R. piresii* lacks one of the IRs (Fig. 1).

Comparing the plastome of *R. piresii* with *P. lambertii* and *N. nagi* by dot-plot analyses (Fig. 2), we noted that the structure of the *R. piresii* plastome differs from the other two species by one large inversion (~56 kb) flanked by a short IR region containing the *trnN-GUU* gene. Thus, we investigated if these short IR sequences were a recombinationally active site, leading to an IR-mediated inversion. The presence of these arrangements occurring between the short IR containing the *trnN-GUU* gene was confirmed by specific PCR primers suitable to amplify all recombination products (Fig. 3). PCR amplification with several primer combinations confirmed that indeed, this IR-mediated inversion produced two different isoforms of the *R. piresii* plastome (Fig. 3). The plastid DNA used for the PCR amplification was isolated from the same plant and revealed that both isoforms coexist in a single *R. piresii* plant (Fig. 3). The presence of the different plastome isoforms in needle tissues of the same plant was also confirmed by mapping of paired-end reads (Electronic Supplementary Material 1).

Although, the two isoforms differ in the orientation of a 56-kb segment of the plastome (Fig. 1), they are functionally equivalent, considering that they both carry the same gene content and do not affect the integrity of other chloroplast genes (Fig. 1). The different isoforms were readily detectable by PCR, and it is highly unlikely that PCR artifacts are

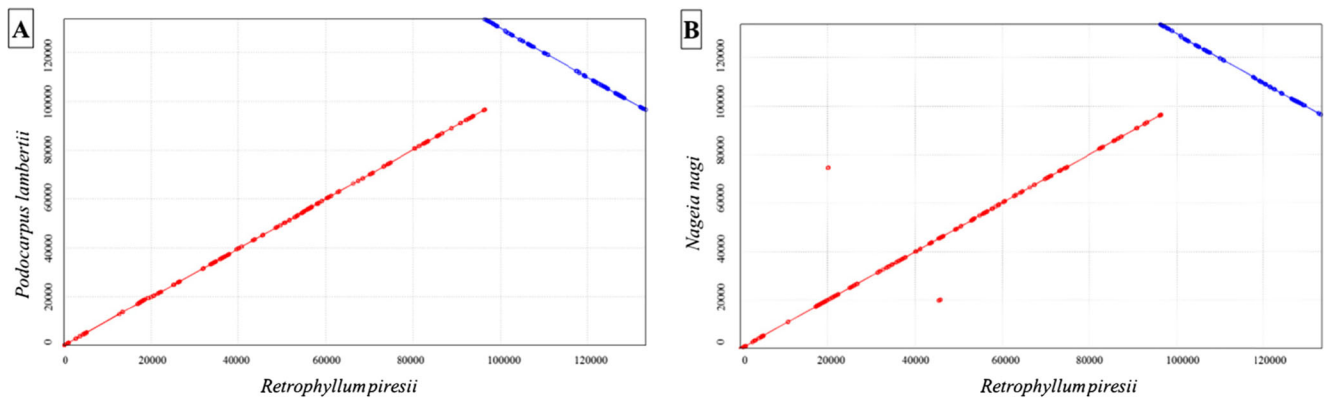


Fig. 2 Dot-plot analyses of *Podocarpus lambertii* and *Nageia Nagi* plastome sequence against *Retrophyllum piresii*. A positive slope denotes that the compared two sequences are in the same orientations, whereas a negative one indicates that the compared sequences can be

aligned but their orientations are opposite. *Graphs* represent comparisons between *R. piresii* (axis X) and *P. lambertii* (axis Y) (a) and *R. piresii* (axis X) and *N. Nagi* (axis Y) (b)

involved here since the recombination was also observed by sequencing data (Electronic Supplementary Material 1).

In *M. truncatula*, an unprecedented presence of two stable alternative plastomes configuration was reported (Gurdon and Maliga 2014). These two configurations were a ~45 kb inversion between a short (20–24 nt) imperfect repeat in different

ecotypes. Shortly after, Guo et al. (2014) described these multiple genomic isoforms coexisting within individual plants. In *Juniperus*, two plastome configurations with a large ~36-kb inversion between inverted repeats of 250 bp containing two copies of *trnQ-UUG* genes and coexist also in the same plant. Different isoforms are not always present in similar amounts

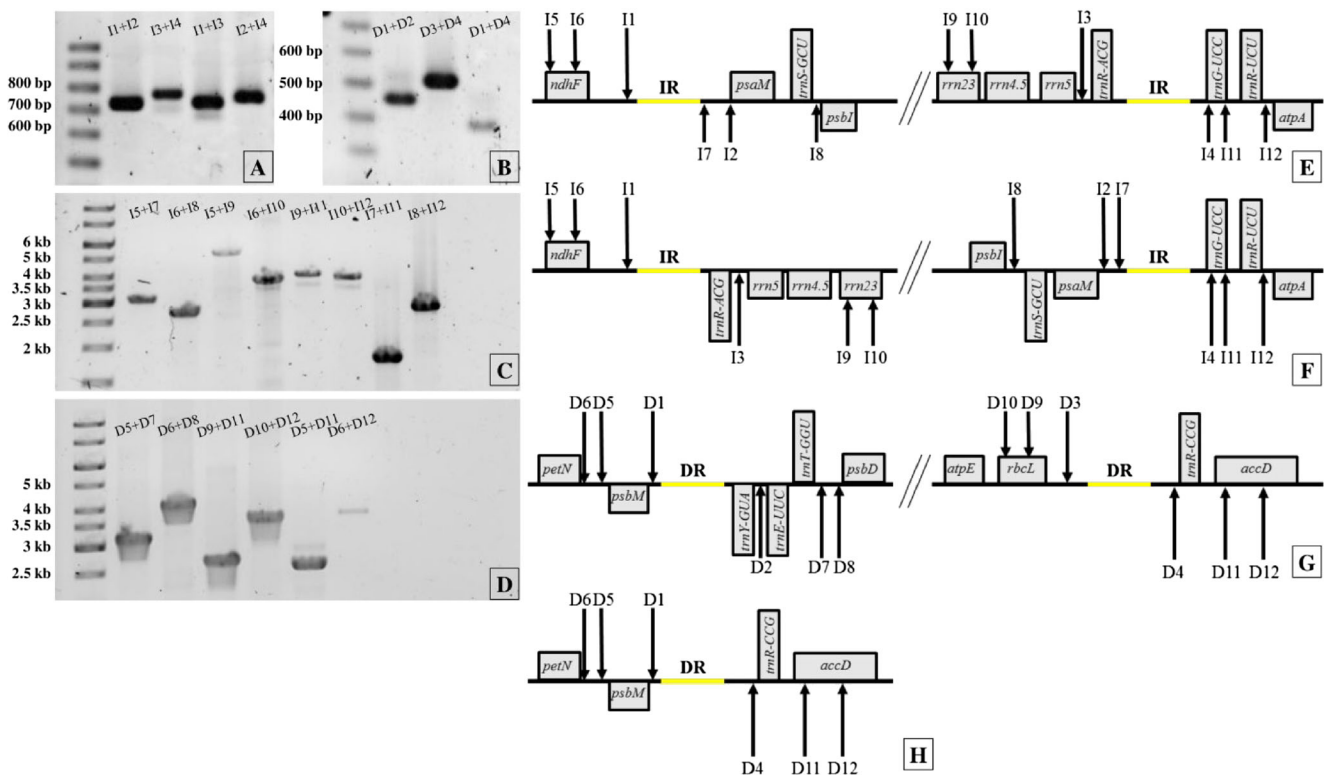


Fig. 3 PCR analysis of recombinant genomes. **a** PCR amplification products for IR-mediated inversion with 100 bp ladder; **b** PCR amplification products for DR-mediated deletion with 100 bp ladder; **c** PCR amplification products for IR-mediated inversion with 1 kb ladder; **d** PCR amplification products for DR-mediated deletion with 1 kb ladder; **e** PCR primer combination designed to amplify genome IR-mediated inversion, isoform 1; **f**, **e** PCR primer combination designed to amplify

genome IR-mediated inversion, isoform 2; **g** PCR primer combination designed to amplify genome DR-mediated deletion, isoform 1; **h** PCR primer combination designed to amplify genome DR-mediated deletion, isoform 3. *IR* indicates the short inverted repeat located in the position of *trnN-GUU*. *DR* indicates the short directed repeat formed in the position of *trnD-GUC*. In **h**, one copy of the directed repeat is deleted and only one copy remains

Table 3 List of repeated tRNA in sequenced gymnosperms plastomes

Taxon	Species	Repeated tRNA												
		trnI- CAU	trnQ- UUG	trnN- GUU	trnV- GAC	trnH- GUG	trnL- CAA	trnR- ACG	trnI- GAU	trnA- UCG	trnR- CCG	trnD- GUC	trnS- GCU	trnT- GGU
UGA conifer clade I														
IPinaceae	<i>Cathaya argyrophylla</i>	I												I
	<i>Cedrus deodara</i>	I												II
	<i>Keteleeria davidiana</i>	I												I
	<i>Larix decidua</i>	I												I
	<i>Picea abies</i>	I				I								ID
	<i>Picea morrisonicola</i>	I				D								
	<i>Pinus koraiensis</i>	I				I								ID
	<i>Pinus krempfii</i>	I				I								I
	<i>Pinus lambertiana</i>	I				I								I
	<i>Pinus massoniana</i>	I				I								ID
	<i>Pinus monophylla</i>	I				I								ID
	<i>Pinus nelsonii</i>	I												ID
	<i>Pinus taeda</i>	I				I								ID
	<i>Pinus thunbergii</i>	I				I								ID
	<i>Pseudotsuga sinensis</i>	I												
Conifer clade II														
Araucariaceae	<i>Agathis dammara</i>	I												D
Cupressaceae	<i>Calocedrus formosana</i>	I												
	<i>Cryptomeria japonica</i>	II												
	<i>Cunninghamia lanceolata</i>	II												
	<i>Juniperus bermudiana</i>	I												
	<i>Juniperus monosperma</i>	II												
	<i>Juniperus scopulorum</i>	II												
	<i>Juniperus virginiana</i>	II												
	<i>Taiwania cryptomerioides</i>	II	I											
	<i>Taiwania flousiana</i>	II	I											
	Cephalotaxaceae	<i>Cephalotaxus oliveri</i>	I											
<i>Cephalotaxus wilsoniana</i>		D I												
Podocarpaceae	<i>Nageia nagi</i>	I												
	<i>Podocarpus lambertii</i>	I												
	<i>Podocarpus totara</i>													
	<i>Retrophyllum piresii</i>	I												
Taxaceae	<i>Taxus mairei</i>	II												
Cycadidae														
Cycadaceae	<i>Cycas revoluta</i>	I	I	I	I	II	I							
	<i>Cycas taitungensis</i>	I	I	I	I	II	I							
Ginkgooidae														
Ginkgoaceae	<i>Ginkgo biloba</i>	I	I	I	I	II	D							
Gnetidae														
Ephedraceae	<i>Ephedra equisetina</i>	II	I	I	I	II	I							
Gnetaceae	<i>Gnetum montanum</i>	II	I	I	I	II	I							
	<i>Gnetum parvifolium</i>	II	I	I	I	II	I							
Welwitschiaceae	<i>Welwitschia mirabilis</i>	II	I		I	II	D							

I inverted repeat, D direct repeat

because homologous recombination is a randomly physical mechanism and is distributed due to random segregation of the isoforms during cell and organelle division (Rogalski et al. 2006, 2008b; Guo et al. 2014).

Analyzing different gymnosperm plastome sequences available in GenBank, it is possible to detect the presence of different tRNA genes repeated in direct or inverted copies (Table 3). In general, conifer clade I present *trnI-CAU*, *trnS-GCU*, and *trnH-GUG* in inverted repeat, and *trnT-GGU* in direct repeat. The conifer clade II, families Cupressaceae and Taxaceae, present *trnI-CAU* and *trnQ-UUG* in inverted repeat, while Cephalotaxaceae presents the *trnQ-UUG*, and Podocarpaceae presents the *trnN-GUU*. Cycadidae, Gnetidae, and Ginkgoidae did not lose the large IRs; therefore, they have several tRNAs in inverted repeats.

The *trnI-CAU* gene in conifers clade I was not reported to show ability to recombine and generate inversion between them (Lin et al. 2010; Wu et al. 2011). However, in conifer clade II species, the recombinational activity of the short IR containing *trnQ-UUG* (544 bp) was found to occur in *C. oliveri* (Cephalotaxaceae) but not in *C. japonica* (Cupressaceae) and *T. cryptomerioides* (Cupressaceae). The last two species have *trnQ-UUG*-containing short IRs of approximately 280 bp (Yi et al. 2013). Two species of Podocarpaceae family showed short IRs composed of two copies of *trnN-GUU* (Vieira et al. 2014a; Wu and Chaw 2014), but it remains to be assayed if they are recombinationally active. In the conifer clade II species, *Juniperus* genus (Cupressaceae), short IRs (~250 bp) containing *trnQ-UUG* were shown to recombine and created a large 36 kb inversion (Guo et al. 2014). More recently, a triplication of *trnI-CAU* was observed in an angiosperm species, *Paris verticillata*, although, at first analyses, no rearrangements were observed (Do et al. 2014).

We also identified in *R. piresii* plastome a short DR of 173 bp containing the *trnD-GUC* gene. This DR is separated by several tRNA genes and genes encoding proteins related to photosynthesis, chlororespiration, and translation (~25 kb) (Fig. 1). We investigated whether this DR could recombine and cause the deletion of its internal content. PCR data containing amplified products with suitable primers (Fig. 3) confirmed the presence of the two plastome isoforms, one containing the DR and the other one with a single copy of *trnD-GUC* gene and the deletion of the previous internal gene content (Fig. 3). This hypothesis was confirmed by mapping the paired-end reads with both plastome isoforms (Electronic Supplementary Material 2).

In transgenic plastids, the appearance of unexpected plastome conformations was observed when endogenous regulatory sequences were used (Rogalski et al. 2006; 2008a; 2008b; Fleischmann et al. 2011; Alkatib et al. 2012). The use of endogenous sequences (promoters, 5'- and 3'-UTRs)

to control transgene expression duplicated relatively short sequences in the plastome and these recombinationally active duplicated sequences can be distributed by chance as IR or DR. If they were positioned as DR, they induced deletion of the sequence between them (Rogalski et al. 2008a) and, otherwise, if they were found as IR, they can work as flip-flop recombination (Rogalski et al. 2006; 2008b).

Deletion of plastome sequences via genetic engineering of directly repeated sequences is a precise method already used successfully for elimination of the selectable marker gene (Iamtham and Day 2000; Day et al. 2005) and targeted disruption of a plastid gene (Kode et al. 2006). The two mechanisms in transgenic plastids, deletion or inversion, mediated by repeated sequences were demonstrated to be a totally random process considering that the different isoforms were found in the same plastids, cells, and/or tissues with different predominance (Rogalski et al. 2006; 2008a; 2008b; Fleischmann et al. 2011; Alkatib et al. 2012).

The results found in the present work comprise the first report in nature of a DR-mediated deletion in plastome of untransformed plants. Similarly to the previous analysis, DNA from only one plant was used, confirming that these isoforms co-exist within a single plant. Given that no abnormal or variegated needles were observed in the *R. piresii* plant used for plastome sequencing, there are several interesting and remaining questions: Is this a peculiarity of *R. piresii* plastome or a more common phenomenon present in other plastomes that has been overlooked before? What is the evolutionary advantage of this recombination since photosynthetic and housekeeping genes are deleted? Considering that plastomes have a high ploidy level, is there a mix of viable and unviable plastome isoforms which suffice for gene expression, providing sufficient amount of tRNAs and proteins related to plastid gene expression and photosynthesis? If there is a selection pressure exerted by plastid gene expression and photosynthesis on plastome to eliminate the unviable plastome isoforms and prevent aberrant growth in conifers, how does it work? Deepen on these and others questions can help unravel important aspects of the adaptive evolution of conifers.

Material and methods

Plant material and cpDNA purification

Chloroplast isolation of *R. piresii* was performed from a single individual fresh leaf gently provided by Goeldi Museum (Museu Paraense Emilio Goeldi), Brazil. Chloroplasts and plastid DNA from young needles were obtained according to Vieira et al. (2014b).

Plastome sequencing, assembling, and annotation

Approximately 50 ng of cpDNA was used to prepare sequencing libraries with Nextera DNA Sample Prep Kit (Illumina Inc., San Diego, CA) according to the manufacturer's instructions. The obtained library was sequenced using Illumina MiSeq (Illumina Inc., San Diego, CA). The paired-end reads (2 × 300 bp) were applied on a de novo assembly performed using Newbler 2.6v and CLC Genomics Workbench 6.5v. The plastome coverage was estimated using the CLC Genomics Workbench 6.5v software. By using this approach, a total of 165,080 paired-end reads were mapped resulting in ~480-fold plastome coverage. Initial annotation of the *R. piresii* plastome was performed using Dual Organellar GenoMe Annotator (DOGMA) (Wyman et al. 2004). From this initial annotation, putative starts, stops, and intron positions were determined based on comparisons to homologous genes in other plastomes. The tRNA genes were further verified by using tRNAscan-SE (Schattner et al. 2005). The physical map of the circular plastome was drawn using OrganellarGenomeDRAW (OGDRAW) (Lohse et al. 2013). The physical map of the circular plastome isoforms 2 and 3 are supplied in Electronic Supplementary Material 3.

Repeat sequence analysis and IR identification

Simple sequence repeats (SSRs) were detected using MISA perl script, available at <http://pgrc.ipk-gatersleben.de/misa/>, with thresholds of eight repeat units for mononucleotide SSRs, four repeat units for di- and trinucleotide SSRs, and three repeat units for tetra-, penta-, and hexanucleotide SSRs. REPuter (Kurtz et al. 2001) was used to visualize the remaining IRs in *R. piresii* by forward vs. reverse complement (palindromic) alignment. The minimal repeat size was set to 30 bp and the identity of repeats ≥90 %.

Comparative analysis of plastome structure

We used the PROtein MUMmer (PROmer) Perl script in MUMmer 3.0 (Kurtz et al. 2004), available at <http://mummer.sourceforge.net/>, to visualize gene order conservation (dot-plot analyses) between *R. piresii* and the Podocarpaceae conifer representatives *P. lambertii* and *N. nagi*.

Plastome recombination analysis

The presence of the three plastome isoforms produced by DR- and IR-mediated recombinations was confirmed by PCR amplification using the combinations of primers indicated in Table 4. In 25 µl reactions, 25 ng of

Table 4 Set of primers used in PCR amplification

Primer	Sequence (5'-3')
I1	GGAATTCACCTGGGACAGAT
I2	GGCTACAGACTGTCTCTCCTTG
I3	GAATAGCTGAGAGCGGATCAAG
I4	CCTTGAGAACATTCTGTTTCG
I5	GCCCAATCTGGTACAAATC
I6	GGGACAAGAATCCAGCACAT
I7	AATTACAGCTGCTCTGGGCTA
I8	CCCGGACGTAAAGGATAGTG
I9	GCTAAGGCCCTAAATGACC
I10	AAAGCTCGAACGACCGTTTA
I11	TTCATTTTCATTCGGCTCCTC
I12	CCTCATTGAGCGCAAAAGAT
D1	ATTTCTCTGGGGAGGAGCAT
D2	TTATGGTTCTGGGTCGATG
D3	ATGGCGGCATAATTAACCAA
D4	AATAGGGCTCGGGCTAGAAG
D5	CCGCCTCATGATCCATATTC
D6	TTGCGTTGATCGGTCTTAGTC
D7	TTGTCTCGGGATTCTGCTTT
D8	TGATGTGAGCGAATAGGAACA
D9	CCAAAACCTTCCAAGGTCCA
D10	CTGGGTTCAAAGCTGGTGT
D11	CGGTTTCATCTGCCAAGAAAT
D12	AGAGGATCCCTTGCAGTTCA

cpDNA was amplified in a reaction mixture containing 200 mM of each dNTP (Sigma-Aldrich), 2.0 mM MgCl₂ (Sigma-Aldrich), 5 pmol of each primer (Sigma-Aldrich), and 1 U Taq DNA polymerase (Sigma-Aldrich). The standard PCR program was 40 cycles of 30 s at 94 °C, 30 s at 63 °C, and 60 s at 72 °C with a 3-min extension of the first cycle at 94 °C and a 10-min final extension at 72 °C. PCR products were analyzed by electrophoretic separation in 1 % agarose gel.

The different plastome presence was also confirmed by mapping of Illumina paired-end reads, which indicated the presence of the three plastome products of recombination. The products were analyzed by using the software CLC Genomics Workbench 6.5v.

Data archiving statement

The complete nucleotide sequence of *R. piresii* plastome sequenced in this study is available in GenBank database under accession number KJ617081.

Acknowledgments This work was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and Fundação de Amparo à Pesquisa e Inovação do Estado de Santa Catarina (FAPESC 14848/2011-2, 3770/2012, and 2780/2012-4). The authors thank Prof. Dr. Willian Antônio Rodrigues for the information related to the species, Angelo Schuabb Heringer and Ramon Felipe Scherer for collecting the plant material, and Museu Paraense Emílio Goeldi for kindly providing the plant material.

References

- Ahmed I, Matthews PJ, Biggs PJ, Naeem M, McLenachan PA, Lockhart PJ (2013) Identification of chloroplast genome loci suitable for high-resolution phylogeographic studies of *Colocasia esculenta* (L.) Schott (Araceae) and closely related taxa. *Mol Ecol Resour* 13: 929–937
- Alkatib S, Scharff LB, Rogalski M, Fleischmann TT, Matthes A et al (2012) The contributions of wobbling and superwobbling to the reading of the genetic code. *PLoS Genet* 8(11), e1003076
- Baskauf CJ, Jinks NC, Mandel JR, McCauley DE (2014) Population genetics of Braun's Rockcress (*Boecheira perstellata*, Brassicaceae), an endangered plant with a disjunct distribution. *J Hered* 105:265–275
- Besnard G, Hernández P, Khadari B, Dorado G, Savolainen V (2011) Genomic profiling of plastid DNA variation in the Mediterranean olive tree. *BMC Plant Biol* 11:80
- Biffin E, Conran J, Lowe A (2011) Podocarp evolution: a molecular phylogenetic perspective. In: Turner BL, Cernusak LA (eds) *Ecology of the Podocarpaceae in tropical forests*. Smithsonian Institution Scholarly Press, Washington, pp 1–20
- Cronn R, Liston A, Parks M, Gernandt DS, Shen R et al (2008) Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. *Nucleic Acids Res* 36, e122
- Day A, Kode V, Madesis P, Iamtham S (2005) Simple and efficient removal of marker genes from plastids by homologous recombination. *Methods Mol Biol* 286:255–270
- Delplancke M, Alvarez N, Espíndola A, Joly H, Benoit L, Brouck E, Arrigo N (2012) Gene flow among wild and domesticated almond species: insights from chloroplast and nuclear markers. *Evol Appl* 5: 317–329
- Dexter KG, Terborgh JW, Cunningham CW (2012) Historical effects on beta diversity and community assembly in Amazonian trees. *Proc Natl Acad Sci U S A* 109:7787–7792
- Do HDK, Kim JS, Kim J (2014) A *trnI-CAU* triplication event in the complete chloroplast genome of *Paris verticillata* M.Bieb. (Melanthiaceae, Liliales). *Genome Biol Evol* 6:1699–1706
- Farjon A (1998) World checklist and bibliography of conifers. The Royal Botanical Gardens, Kew
- Fleischmann TT, Scharff LB, Alkatib S, Hasdorf S, Schottler MA et al (2011) Nonessential plastid-encoded ribosomal proteins in tobacco: a developmental role for plastid translation and implications for reductive genome evolution. *Plant Cell* 23:3137–3155
- 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
- Guo X, Castillo-Ramírez S, González V, Bustos P, Fernández-Vázquez JL, Santamaría RI, Arellano J, Cevallos MA, Dávila G (2007) Rapid evolutionary change of common bean (*Phaseolus vulgaris* L.) plastome, and the genomic diversification of legume chloroplasts. *BMC Genomics* 8:228
- Guo W, Grewe F, Cobo-Clark A, Fan W, Duan Z, Adams RP, Schwarzbach AE, Mower JP (2014) Predominant and substoichiometric isomers of the plastid genome coexist within *Juniperus* plants and have shifted multiple times during cupressophyte evolution. *Genome Biol Evol* 6:580–590
- Gurdon C, Maliga P (2014) Two distinct plastid genome configurations and unprecedented intraspecific length variation in the *accD* coding region in *Medicago truncatula*. *DNA Res* 21:417–427
- Hansen DR, Dastidar SG, Cai Z, Penafior C, Kuehl JV, Boore JL, Jansen RK (2007) Phylogenetic and evolutionary implications of complete chloroplast genome sequences of four early-diverging angiosperms: *Buxus* (Buxaceae), *Chloranthus* (Chloranthaceae), *Dioscorea* (Dioscoreaceae), and *Illicium* (Schisandraceae). *Mol Phylogenet Evol* 45:547–563
- 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
- Iamtham S, Day A (2000) Removal of antibiotic resistance genes from transgenic tobacco plastids. *Nat Biotechnol* 18:1172–1176
- Jansen RK, Ruhlman TA (2012) Plastid genomes of seed plants. In: Bock R, Knoop V (eds) *Genomics of chloroplasts and mitochondria*. Springer, Netherlands, pp 103–126
- Jansen RK, Cai Z, Raubeson LA, Daniell H, dePamphilis CW et al (2007) Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc Natl Acad Sci U S A* 104:19369–19374
- Kato S, Iwata H, Tsumura Y, Mukai Y (2011) Genetic structure of island populations of *Prunus lannesiana* var. *speciosa* revealed by chloroplast DNA, AFLP and nuclear SSR loci analyses. *J Plant Res* 124: 11–23
- Kato S, Imai A, Rie N, Mukai Y (2013) Population genetic structure in a threatened tree, *Pyrus calleryana* var. *dimorphophylla* revealed by chloroplast DNA and nuclear SSR locus polymorphisms. *Conserv Genet* 14:983–996
- Knox EB (2014) The dynamic history of plastid genomes in the Campanulaceae *sensu lato* is unique among angiosperms. *Proc Natl Acad Sci U S A* 111:11097–11102
- Kode V, Mudd EA, Iamtham S, Day A (2006) Isolation of precise plastid deletion mutants by homology-based excision: a resource for site-directed mutagenesis, multi-gene changes and high-throughput plastid transformation. *Plant J* 46:901–909
- Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R (2001) REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res* 29:4633–4642
- Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL (2004) Versatile and open software for comparing large genomes. *Genome Biol* 5:R12
- Lin C, Huang J, Wu C, Hsu C, Chaw S (2010) Comparative chloroplast genomics reveals the evolution of Pinaceae genera and subfamilies. *Genome Biol Evol* 2:504–517
- Lin CP, Wu C, Huang Y, Chaw S (2012) The complete chloroplast genome of *Ginkgo biloba* reveals the mechanism of inverted repeat contraction. *Genome Biol Evol* 4:374–381
- Lohse M, Drechsel O, Kahlau S, Bock R (2013) OrganellarGenomeDRAW: a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucl Acids Res*. doi:10.1093/nar/gkt289
- López P, Tremetsberger K, Kohl G, Stuessy T (2012) Progenitor-derivative speciation in *Pozoa* (Apiaceae, Azorelloideae) of the southern Andes. *Ann Bot* 109:351–363
- Moore MJ, Bell CD, Soltis PS, Soltis DE (2007) Using plastid genomic-scale data to resolve enigmatic relationships among basal angiosperms. *Proc Natl Acad Sci U S A* 104:19363–19368
- Moore MJ, Soltis PS, Bell CD, Burleigh JG, Soltis DE (2010) Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proc Natl Acad Sci U S A* 107:4623–4628

- Murray BG (2013) Karyotype variation and evolution in gymnosperms. In: Leitch IJ, Greilhuber J, Dolezel J, Wendel JF (eds) Plant genome diversity 2: physical structure, behaviour and evolution of plant genomes. Springer, Vienna, p 231–243
- Palmer JD (1983) Chloroplast DNA exists in two orientations. *Nature* 301:92–93
- Parks M, Cronn R, Liston A (2009) Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biol* 7:84
- Powell W, Morgante M, Andre C, McNicol JW, Machray GC, Doyle JJ, Tingey SV, Rafalski JA (1995) Hypervariable microsatellites provide a general source of polymorphic DNA markers for the chloroplast genome. *Curr Biol* 5:1023–1029
- Provan J, Powell W, Hollingsworth PM (2001) Chloroplast microsatellites: new tools for studies in plant ecology and evolution. *Trends Ecol Evol* 16:142–147
- Rogalski M, Ruf S, Bock R (2006) Tobacco plastid ribosomal protein S18 is essential for cell survival. *Nucleic Acids Res* 34:4537–4545
- Rogalski M, Schottler MA, Thiele W, Schulze WX, Bock R (2008a) *Rpl33*, a nonessential plastid-encoded ribosomal protein in tobacco, is required under cold stress conditions. *The Plant Cell* 20:2221–2237
- Rogalski M, Karcher D, Bock R (2008b) Superwobbling facilitates translation with reduced tRNA sets. *Nat Struct Mol Biol* 15:192–198
- Rogalski M, Vieira LN, Fraga HPF, Guerra MP (2015) Plastid genomics in horticultural species: importance and applications for plant population genetics, evolution, and biotechnology. *Front Plant Sci* 6:586
- Roullier C, Rossel G, Tay D, McKey D, Lebot V (2011) Combining chloroplast and nuclear microsatellites to investigate origin and dispersal of New World sweet potato landraces. *Mol Ecol* 20:3963–3977
- Roullier C, Duputié A, Benoit L, Fernández-Bringas VM et al (2013) Correction: disentangling the origins of cultivated sweet potato (*Ipomoea batatas* (L.) Lam). *PLoS One* 8(10):10.1371
- Ruhsam M, Rai HS, Mathews S, Ross TG, Graham SW et al (2015) Does complete plastid genome sequencing improve species discrimination and phylogenetic resolution in *Araucaria*? *Mol Ecol Resour* 15:1067–1078
- Schattner P, Brooks AN, Lowe TM (2005) The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res* 33:W686–W689
- Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N et al (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *EMBO J* 5:2043–2049
- 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
- Tangphatsomruang S, Sangsrakru D, Chanprasert J, Uthapaisanwong P, Yoocha T et al (2009) The chloroplast genome sequence of mungbean (*Vigna radiata*) determined by high-throughput pyrosequencing: structural organization and phylogenetic relationships. *DNA Res* 17:11–22
- Tomar RSS, Deshmukh RK, Naik BK, Tomar SMS, Vinod (2014) Development of chloroplast-specific microsatellite markers for molecular characterization of alloplasmic lines and phylogenetic analysis in wheat. *Plant Breed* 133:12–18
- Tsudzuki J, Nakashima K, Tsudzuki T, Hiratsuka J, Shibata M et al (1992) Chloroplast DNA of black pine retains a residual inverted repeat lacking rRNA genes: nucleotide sequences of *trnQ*, *trnK*, *psbA*, *trnI* and *trnH* and the absence of *rps16*. *Mol Gen Genet* 232:206–214
- Vendramin GG, Anzidei M, Madaghiele A, Bucci G (1998) Distribution of genetic diversity in *Pinus pinaster* Ait. as revealed by chloroplast microsatellites. *Theor Appl Genet* 97:456–463
- Vieira LN, Faoro H, Rogalski M, Fraga HPF, Cardoso RLA (2014a) The complete chloroplast genome sequence of *Podocarpus lambertii*: genome structure, evolutionary aspects, gene content and SSR detection. *PLoS One* 9(3), e90618
- Vieira LN, Faoro H, Fraga HPF, Rogalski M, de Souza EM et al (2014b) An improved protocol for intact chloroplasts and cpDNA isolation in conifers. *PLoS One* 9(1), e84792
- Wakasugi T, Tsudzuki J, Ito S, Nakashima K, Tsudzuki T et al (1994) Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. *Proc Natl Acad Sci U S A* 91:9794–9798
- Wicke S, Schneeweiss GM, dePamphilis CW, Müller 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
- Wu CS, Chaw SM (2014) Highly rearranged and size-variable chloroplast genomes in conifers II clade (cupressophytes): evolution towards shorter intergenic spacers. *Plant Biotechnol J* doi: doi:10.1111/pbi.12141
- 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
- Wu CS, Lai YT, Lin CP, Wang YN, Chaw SM (2009) Evolution of reduced and compact chloroplast genomes (cpDNAs) in gnetophytes: selection towards a lower cost strategy. *Mol Phylogent Evol* 52:115–124
- Wu CS, Wang YN, Hsu CY, Lin CP, Chaw SM (2011) Loss of different inverted repeat copies from the chloroplast genomes of Pinaceae and cupressophytes and influence of heterotachy on the evaluation of gymnosperm phylogeny. *Genome Biol Evol* 3:1284–1295
- Wyman SK, Jansen RK, Boore JL (2004) Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20:3252–3255
- Yi X, Gao L, Wang B, Su Y, 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
- Zhang Y, Ma J, Yang B, Li R, Zhu W, Sun L, Tian J, Zhang L (2014) The complete chloroplast genome sequence of *Taxus chinensis* var. *mairei* (Taxaceae): loss of an inverted repeat region and comparative analysis with related species. *Gene* 540:201–209