Leila do Nascimento Vieira

ESTRUTURA E EVOLUÇÃO DO GENOMA PLASTIDIAL EM ARAUCARIACEAE HENKEL & W.HOCHST. E PODOCARPACEAE ENDL.

Tese submetida ao Programa de Pós-graduação em Recursos Genéticos Vegetais da Universidade Federal de Santa Catarina para a obtenção do Grau de Doutor em Ciências.
Orientador: Prof. Dr. Miguel Pedro Guerra.
Coorientador: Prof. Dr. Marcelo Rogalski

Florianópolis
2017

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Vieira, Leila
ESTRUTURA E EVOLUÇÃO DO GENOMA PLASTIDIAL EM ARAUCARIACEAE HENKEL & W.HOCHST. E PODOCARPACEAE ENDL. / Leila Vieira ; orientador, Miguel Pedro Guerra ; coorientador, Marcelo Rogalski. - Florianópolis, SC, 2017.
112 p.

Tese (doutorado) - Universidade Federal de Santa Catarina, Centro de Ciências Agrárias. Programa de Pós Graduação em Recursos Genéticos Vegetais.

Inclui referências

1. Recursos Genéticos Vegetais. 2. Genoma plastidial. 3. Filogenômica. 4. Araucariaceae. 5. Podocarpaceae. I. Guerra, Miguel Pedro . II. Rogalski, Marcelo. III. Universidade Federal de Santa Catarina. Programa de Pós Graduação em Recursos Genéticos Vegetais. IV. Título.

Folha de assinaturas

Este trabalho é dedicado aos meus pais, Rogério Alano Vieira e Eloisa do Nascimento Vieira.

AGRADECIMENTOS

Aos meus pais Rogério Alano Vieira e Eloisa do Nascimento Vieira pelo apoio incondicional durante toda minha formação acadêmica.

À toda minha família, em especial ao meu irmão Felipe do Nascimento Vieira, Eloysa Caroline Oliveira e ao meu sobrinho Lucas Oliveira Vieira.

Ao Hugo P. de Freitas Fraga por toda a dedicação, carinho e compreensão essenciais para o desenvolvimento desse trabalho.

Ao meu orientador, Prof. Miguel Pedro Guerra, por todos os ensinamentos desde a minha iniciação científica, pela confiança depositada no meu trabalho e pelas diversas oportunidades proporcionadas durante todos os anos de orientação.

Aos Professores Rubens O. Nodari, Marcelo Rogalski, Maurício S. dos Reis e Robson F. de Souza por todas as conversas e ensinamentos que tornaram esse trabalho possível e pelo apoio de sempre.

À toda a equipe da UFPR, em especial ao Prof. Emanuel M. de Souza e ao Dr. Helisson Faoro, pela disponibilização da plataforma de sequenciamento e pelos infinitos ensinamentos desde o trabalho de mestrado que contribuíram para minha formação acadêmica.

À pequena/grande equipe de genômica plastidial, Karina Goulart dos Anjos, Karla Gasparini e Lilian de Oliveira Machado pela troca de experiências.

Aos amigos Morgana Lopes, Tiago Tomazetti, Márcia Rossarolla, Luíza Polesi e Catarina Puttkammer pelos momentos de descontração e apoio.

À todos os colegas do LFDGV (UFSC) pela agradável convivência.

Ao pessoal do Núcleo de Fixação de Nitrogênio (NFN) - UFPR pela ótima recepção e convivência, em especial ao pessoal das salas 271 e 272.

Ao ICC/Fiocruz pela disponibilização da estrutura para realização de análises, em especial à Dr. Gisele A. Picchi e ao Dr. Helisson Faoro.

Ao Museu Paraense Emílio Goeldi pela disponibilização das amostras de material vegetal de *Retrophyllum piresii* utilizadas nesse trabalho, em especial ao Eng. Agrônomo Amir Lima.

Ao “Instituto de Investigaciones Agropecuarias” (INIA) do Chile pela autorização de coleta da espécie *Araucaria araucana*.

À Bernadete Ribas e aos coordenadores do PPG/RGV pelo suporte administrativo, e à banca avaliadora pelas sugestões a esse trabalho.

À CAPES pelo apoio financeiro através da bolsa de estudos, à FAPESC e ao CNPq pelo financiamento das atividades e à UFSC por prover um ensino de qualidade.

RESUMO

As coníferas são um grupo composto por seis famílias e distribuído por todo o mundo, mas no Brasil apenas espécies das famílias Araucariaceae e Podocarpaceae apresentam ocorrência natural. A família Araucariaceae é representada pela *Araucaria angustifolia* e a Podocarpaceae por oito espécies do gênero *Podocarpus* e o *Retrophyllum piresii*. A *Araucaria araucana* e *A. angustifolia* são as duas únicas espécies da família Araucariaceae encontradas na América do Sul, ambas pertencentes à secção *Araucaria*. Esta secção é filogeneticamente irmã do grupo *Bunya + Intermedia*, cujas únicas representantes são a *Araucaria bidwillii* (nativa da Austrália) e a *Araucaria hunsteinii* (nativa da Nova Guiné), respectivamente. As espécies de *Araucaria* nativas da América do Sul sofreram uma grande exploração nas últimas décadas, resultando em uma alta fragmentação dos seus remanescentes florestais. Além disso, as famílias Podocarpaceae e Araucariaceae são consideradas umas das famílias mais antigas existentes, o que as torna interessante para estudos evolutivos. Comparações baseadas na sequência do genoma plastidial possibilitam a realização de uma série de análises estruturais e evolutivas. Além disso, há um grande potencial no estudo do genoma plastidial de coníferas para o desenvolvimento de marcadores moleculares espécie-específicos para análises de filogeografia, com implicações para o manejo e a conservação dessas espécies. Sendo assim, o presente trabalho visou realizar estudos de filogenômica, estrutura e evolução do genoma plastidial em Araucariaceae e Podocarpaceae a partir do sequenciamento completo do genoma plastidial de três espécies da família Araucariaceae (*A. angustifolia*, *A. araucana* e *A. bidwillii*) e uma espécie da família Podocarpaceae (*R. piresii*). O sequenciamento foi realizado utilizando DNA plastidial isolado para construção das bibliotecas, seguido por sequenciamento em plataforma Illumina MiSeq. A partir dos dados de sequenciamento foi realizada montagem, anotação e análise das sequências geradas, incluindo filogenia, estudos estruturais e de regiões de repetição no genoma plastidial destas coníferas. Com o sequenciamento do genoma plastidial da espécie *R. piresii* foi possível a identificação de sítios ativos de recombinação mediados pela presença de repetições diretas e palíndromas. Essas repetições resultam na presença de uma inversão e uma deleção no genoma plastidial. Além disso, o sequenciamento das três espécies do gênero *Araucaria* possibilitou a realização de análises filogenômicas dentro do gênero *Araucaria* de maior resolução até o

momento relatadas. Os resultados obtidos possibilitarão a identificação e seleção de regiões do genoma com alta taxa evolutiva, podendo aumentar a resolução de estudos de filogeografia das espécies alvo.

Palavras-chave: Genoma plastidial. Sequenciamento de nova geração. Filogenômica. Araucariaceae. Podocarpaceae

ABSTRACT

Conifers are a worldwide-distributed group divided into six families. However, the families Araucariaceae and Podocarpaceae are the only indigenous to Brazil. These families are represented by one species of Araucariaceae (*Araucaria angustifolia)* and nine species of the Podocarpaceae, eight of the genus *Podocarpus* and one of *Retrophyllum*. *Araucaria araucana* and *A. angustifolia* are the only two species from Araucariaceae indigenous to South America, both positioned in the section *Araucaria*. This section is sister group to *Bunya + Intermedia*, whose only extant representative is *Araucaria bidwillii* (native to Australia) and *Araucaria hunsteinii* (native to New Guinea), respectively. *Araucaria* species native to South America have undergone extensive exploration in recent decades, resulting in a high fragmentation of forest remnants. In addition, the families Podocarpaceae and Araucariaceae are considered one of the most ancient extant families, which is interesting for evolutionary studies. Comparisons based on the plastome sequence possibilities a series of comparative, structural and phylogenetic analyzes, and may also clarify factors related to the function of repeated inverted and intergenic sequences. In addition, there is great potential of studying conifers plastomes for the development of species-specific molecular markers for phylogeography analysis, with implications in the management and conservation of these species. In this sense, the present study aimed to perform phylogenomic, structural and evolutionary analysis of the plastomes of Araucariaceae and Podocarpaceae. Thus, we sequenced the complete plastome of three species of the family Araucariaceae (*A. angustifolia, A. araucana,* and *A. bidwillii*) and one species of the family Podocarpaceae (*R. piresii*). The DNA sequencing was performed using isolated plastidial DNA for library construction, followed by sequencing on the Illumina MiSeq platform. From the sequencing data, assembly, annotation and analysis of the generated sequences, including phylogeny, structural studies and of regions of repetition in the plastome of these conifers were carried out. With the sequencing of the plastome of the *R. piresii* it was possible to identify recombination sites mediated by the presence of direct and palindromic repetitions. These repetition sites result in the presence of an inversion, and a deletion in the plastome. In addition, the sequencing of the three *Araucaria* species enabled to perform phylogenetic analyzes within the genus *Araucaria* of the highest resolution so far reported. The results obtained here will

allow the identification and selection of genome regions with a high evolutionary rate, which may increase the resolution of phylogeography studies of the target species.
Keywords: Plastidial genome. Next-generation sequencing. Phylogenomics. Araucariaceae. Podocarpaceae

LISTA DE FIGURAS

ANTECEDENTES E JUSTIFICATIVA

Figura 1. Distribuição global das espécies de coníferas…………..…..26
Figura 2. Relações filogenéticas entre a família Araucariaceae e outras coníferas………………………...………………………………27
Figura 3. Representação global da ocorrência de espécies da família Araucariaceae…………………...………………………………28
Figura 4. Sumário comparativo entre as árvores filogenéticas mais parcimoniosas baseado em marcadores AFLP e sequências do gene *rbcL*…………………………....……………………………...…29
Figura 5. Organização estrutural típica do genoma plastidial, dividido em região grande de cópia simples (LSC), região pequena de cópia simples (SSC) e duas regiões invertidas repetidas (IRA e IRB)…...32

CAPÍTULO 1

Figure 1. Gene map of *Retrophyllum piresii* plastome………………...43
Figure 2. Dot-plot analyses of *Podocarpus lambertii* and *Nageia nagi* plastome sequence against *Retrophyllum piresii*………………….....48
Figure 3. PCR analysis of recombinant genomes……………………...51

CAPÍTULO 2

Figure 1. Gene map of *Araucaria angustifolia* plastid genome (plastome)……………………………………………………..67
Figure 2. Multiple plastome alignment with species from families Araucariaceae and Podocarpaceae…………………………………69
Figure 3. Phylogenetic relationship (Bayesian phylogeny) within *Araucaria* genus based on complete plastomes of 17 *Araucaria* species and the sister genus *Wollemia* as outgroup………………..70

LISTA DE QUADROS

CAPÍTULO 1

Table 1. List of genes identified in *Retrophyllum piresii* plastome.….45
Table 2. List of simple sequence repeats identified in *Retrophyllum piresii* plastome……………………………………………………………...47
Table 3. List of repeated tRNA in sequenced gymnosperms plastomes…………………………………………………………………50
Table 4. Set of primers used in PCR amplification…………………..55

CAPÍTULO 2

Table 1. Plastid genome comparison of Araucariaceae species analyzed in this study…………………………………………………………….62
Table 2. *Araucaria angustifolia*, *Araucaria bidwillii*, and *Araucaria araucana* plastid genome sequencing and assembly data. …………....63
Table 3. List of genes identified in *Araucaria angustifolia*, *Araucaria araucana*, and *Araucaria bidwillii* plastomes…….……….…………..64

LISTA DE ABREVIATURAS E SIGLAS

AFLP – do inglês “Amplified Fragment Length Polymorphisms”
BI – do inglês “Bayesian Inference”
CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico
cpDNA – do inglês “Chloroplast DNA”
DOGMA – do inglês “Dual Organellar GenoMe Annotator”
DR – do inglês “Direct Repeat”
FAPESC - Fundação de Apoio à Pesquisa Científica e Tecnológica do Estado de Santa Catarina
INIA – do espanhol “Instituto de Investigaciones Agropecuarias”
IR – do inglês “Inverted Repeat”
IUCN – do inglês “International Union for Conservation of Nature”
LCB – do inglês “Locally Collinear Blocks”
lnL – do inglês “Log-likelihoods”
LSC – do inglês “Large Single-copy”
MCMC – do inglês “Markov Chain Monte Carlo”
ML – do inglês “Maximum Likelihood”
OGDRAW – do inglês “Organellar Genome Draw”
PCR – Reação em Cadeia da Polimerase
Plastome – do inglês “Plastid Genome”
RFLP – do inglês “Restriction Fragment Length Polymorphism”
SSC – do inglês “Small single-copy”
SSR – do inglês “Simple sequence repeat”
TIC - do inglês “Translocon of the Inner Membrane of Chloroplast”
TOC – do inglês “Translocon of the Outer Membrane of Chloroplast”

SUMÁRIO

SUMÁRIO..**41**
ANTECEDENTES E JUSTIFICATIVA..**25**
CONÍFERAS DO HEMISFÉRIO SUL..26
ARAUCARIACEAE..27
PODOCARPACEAE..31
ESTRUTURA E EVOLUÇÃO DO GENOMA PLASTIDIAL.........................33
OBJETIVOS..**38**
OBJETIVO GERAL..38
OBJETIVOS ESPECÍFICOS..38
CAPÍTULO 1..**39**
ABSTRACT..40
INTRODUCTION..40
RESULTS AND DISCUSSION..42
***Retrophyllum piresii* Plastome Size And Gene Content**...............................**42**
Repeat Sequence Analysis..**45**
Plastome Structure..**47**
MATERIAL AND METHODS..53
Plant Material and cp DNA Purification..**53**
Plastome Sequencing, Assembling And Annotation.......................................**53**
Repeat Sequence Analysis And IR Identification...**54**
Comparative Analysis Of Plastome Structure..**54**
Plastome Recombination Analysis..**54**
DATA ARCHIVING STATEMENT..56
ACKNOWLEDGEMENTS..56
CAPÍTULO 2..**57**
ABSTRACT..58
ACKNOWLEDGEMENTS..58
INTRODUCTION..59
MATERIAL AND METHODS..60

Plastid Genome Sequencing, Assembling And Annotation............................60
Comparative Analysis Of Plastome Structure..61
Phylogenetic Inference...61
RESULTS AND DISCUSSION...62
Plastome Sequence And Structure..62
Phylogenetic Inference...68
AVAILABILITY OF SUPPORTING DATA...71
CONSIDERAÇÕES FINAIS E PERSPECTIVAS............................72
REFERÊNCIAS...75
ANEXO A..91

ANTECEDENTES E JUSTIFICATIVA

A presente tese faz parte do projeto “Análises genômicas e transcriptômicas nas coníferas brasileiras *Araucaria angustifolia*, *Podocarpus sellowii*, *Podocarpus lambertii* e *Retrophyllum piresii* visando o uso, conservação, estudos evolutivos, moleculares e biotecnológicos” com financiamento aprovado pela Fundação de Amparo à Pesquisa e Inovação do Estado de Santa Catarina (FAPESC, Proc. 14848/2011-2). As análises foram realizadas no Laboratório de Fisiologia do Desenvolvimento e Genética Vegetal (UFSC) sob a orientação do Prof. Dr. Miguel Pedro Guerra, co-orientação do Prof. Dr. Marcelo Rogalski da UFV e no Núcleo de Fixação de Nitrogênio (UFPR) sob a supervisão do Prof. Dr. Emanuel Maltempi de Souza. O trabalho contou com o apoio do Museu Paraense Emílio Goeldi - Brasil para autorização de coleta e envio de material vegetal da espécie *R. piresii* e do “Instituto de Investigaciones Agropecuarias” (INIA) - Chile para autorização de coleta da espécie *Araucaria araucana*.

As famílias de coníferas ocorrentes no Brasil, Araucariaceae e Podocarpaceae, são grupos irmãos. Da família Podocarpaceae, apresentam ocorrência natural no Brasil nove espécies do gênero *Podocarpus* e uma espécie do gênero *Retrophyllum*. Recentemente, a espécie *P. lambertii* teve seu genoma plastidial sequenciado (VIEIRA et al., 2014a). Dessa forma, a fim de aumentar a representatividade de gêneros da família Podocarpaceae, o *R. piresii* foi a espécie escolhida para a realização desse trabalho. Essa espécie é endêmica do Brasil e foi identificada no Parque Nacional dos Pacaás novos - Brasil. Atualmente não há dados sobre a diversidade ou dispersão dessa espécie.

A *A. angustifolia* é a única espécie da família Araucariaceae com ocorrência natural no Brasil. Essa espécie é nativa da Mata Atlântica e apresenta grande relevância ecológica, cultural e econômica. Além da *A. angustifolia*, a única espécie da família Araucariaceae que ocorre na América do Sul é a *A. araucana*, nativa da Argentina e Chile. Ambas as espécies são pertencentes à secção *Araucaria*. Esta secção é filogeneticamente irmã do grupo *Bunya* + *Intermedia*, cujas únicas representantes são a *Araucaria bidwillii* (nativa da Austrália) e a *Araucaria hunsteinii* (nativa da Nova Guiné), respectivamente. Até o momento, quatorze espécies do gênero *Araucaria* tiveram seu genoma plastidial sequenciado, todas da secção *Eutacta* (RUHSAM et al., 2015).

Um protocolo de isolamento de cloroplastos de coníferas foi desenvolvido por (VIEIRA et al., 2014b), possibilitando o rápido e eficiente sequenciamento do genoma plastidial. Esse protocolo foi

utilizado para o sequenciamento do genoma plastidial de *P. lambertii*, possibilitando análises comparativas da estrutura do genoma plastidial dessa espécie com os outras espécies coníferas não pináceas e pináceas, além da identificação de regiões repetidas e alterações no conteúdo gênico (VIEIRA et al., 2014a).

Dessa forma, o sequenciamento do genoma plastidial das espécies *A. angustifolia*, *A. bidwillii*, *A. araucana* e *R. piresii*, possibilitará a ampliação de estudos evolutivos, estruturais e filogenômicos em Araucariaceae e Podocarpaceae, as famílias de coníferas de ocorrência natural e também de grande importância para o Brasil.

CONÍFERAS DO HEMISFÉRIO SUL

As coníferas (subclasse Pinidae) formam um grupo monofilético de gimnospermas com ampla distribuição por todos os continentes (Fig. 1). Esse grupo apresenta registros fósseis abundantes desde o início da Era Mesozoica (ca. 250 Ma), sendo as únicas espécies arbóreas até o surgimento das angiospermas no início do período Cretáceo (ca. 120 Ma) (FARJON; FILER, 2013).

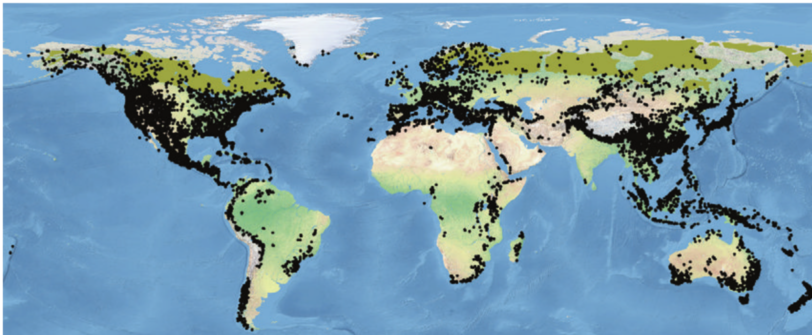

Figura 1. Distribuição global das espécies de coníferas (Extraído de FARJON; FILER, 2013).

Atualmente, as coníferas estão divididas em três ordens (Araucariales, Cupressales e Pinales) e seis famílias, as quais podem ser divididas em coníferas do clado I – pináceaes (Pinaceae) e coníferas do clado II – não pináceas (Araucariaceae, Podocarpaceae, Taxaceae, Cupressaceae e Sciadopityaceae). Apenas quatro famílias de coníferas têm ocorrência natural no Hemisfério Sul: Araucariaceae, Cupressaceae,

Podocarpaceae e Taxaceae (FARJON; FILER, 2013). Essas famílias apresentam uma ampla distribuição geográfica, desde florestas de várzea tropical à zonas áridas e charnecas alpinas (HILL, 1995).

No Brasil, apenas as famílias Araucariaceae e Podocarpaceae apresentam ocorrência natural (JARDIM BOTÂNICO DO RIO DE JANEIRO, 2016). Essas famílias são evolutivamente irmãs (Fig. 2) (QUINN; PRICE, 2003), formando a ordem Araucariales.

A família Araucariaceae é atualmente restrita ao Hemisfério Sul, mas grande parte de seus registros fósseis são do Hemisfério Norte (HILL, 1995). Estima-se que a vegetação de Araucariaceae tenha praticamente desaparecido do Hemisfério Norte no final do período Cretáceo (KUNZMANN, 2007). Essa família é considerada por paleobotânicos como uma das mais antigas entre as famílias de coníferas existentes (KUNZMANN, 2007; WILLIAMS, 2009). De acordo com os registros macrofósseis encontrados referentes ao primeiro período da Era Mesozoica (Triássico), a família Podocarpaceae também se configura como uma das mais antigas famílias existentes (BRODRIBB; HILL, 1999).

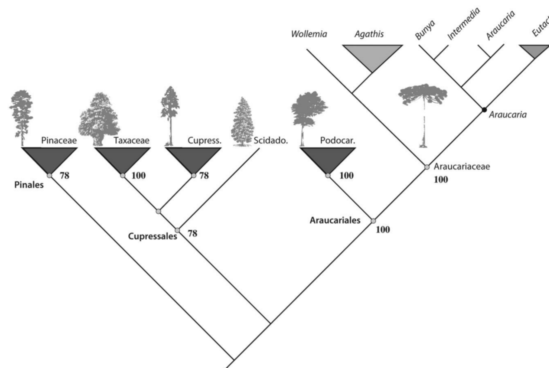

Figura 2. Relações filogenéticas entre a família Araucariaceae e outras coníferas. Os números sobre os nós indicam valores de suporte Jackknife. Cupress.= Cupressaceae; Podocar.= Podocarpaceae; Scidado.= Sciadopitys (Extraído de ESCAPA; CATALANO, 2013).

ARAUCARIACEAE

A família Araucariaceae está atualmente distribuída pela América do Sul e Oceania (Fig. 3) (SETOGUCHI et al., 1998) e

compreende três gêneros: *Araucaria*, *Agathis* e *Wollemia*. O gênero *Araucaria* está dividido em quatro seções: *Araucaria*, *Bunya*, *Eutacta* e *Intermedia* (Fig. 2), somando 19 espécies. Dentre elas, duas espécies oriundas da América do Sul: *A. angustifolia* (Brasil e Argentina) e *A. araucana* (Argentina e Chile) são pertencentes à seção *Araucaria* (BRODRIBB; HILL, 1999). A única espécie ainda existente pertencente à seção *Bunya* é *A. bidwillii*, a qual tem uma dispersão bem limitada, ocorrendo naturalmente apenas nas montanhas de Bunya, ao norte da Austrália (FARJON; FILER, 2013).

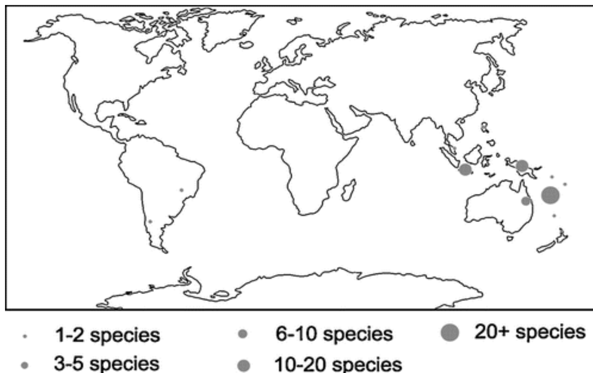

Figura 3. Representação global da ocorrência de espécies da família Araucariaceae (Extraído de KERSHAW; WAGSTAFF, 2001).

Estudos de filogenia realizados na família Araucariaceae associados ao gene plastidial *rbcL* (SETOGUCHI et al., 1998) e a marcadores AFLP (STEFENON; GAILING; FINKELDEY, 2006) mostraram que as espécies do gênero *Araucaria* apresentam seções bem definidas, suportando fortes seções monofiléticas. *A. angustifolia* e *A. araucana* estão agrupadas na mesma secção (Fig. 4), compartilhando importantes caracteres taxonômicos, tais como folhas grandes e planas, germinação hipógea e a presença de dois cotilédones, que ao longo da germinação são retidos no tegumento (STEFENON; GAILING; FINKELDEY, 2006). A *A. bidwillii* está classificada na secção *Bunya*, apresentando maior proximidade e similaridade com a secção *Intermedia* (da Nova Zelândia) e com as espécies da América do Sul do que com as espécies da secção *Eutacta*, de ocorrência na própria Oceania (Fig. 4) (SETOGUCHI et al., 1998; STEFENON; GAILING; FINKELDEY, 2006). A secção *Intermedia*, com um única espécies distribuída na Nova Zelândia, não foi amostrada no estudo de AFLP

(STEFENON; GAILING; FINKELDEY, 2006), mas tanto o estudo com *rbcL* quanto estudo mais recentes e de melhor amostragem indicam que esta é irmã da secção *Bunya*, e que ambas formam o clado irmão da secção *Araucaria*, restrita à América do Sul (ESCAPA; CATALANO, 2013).

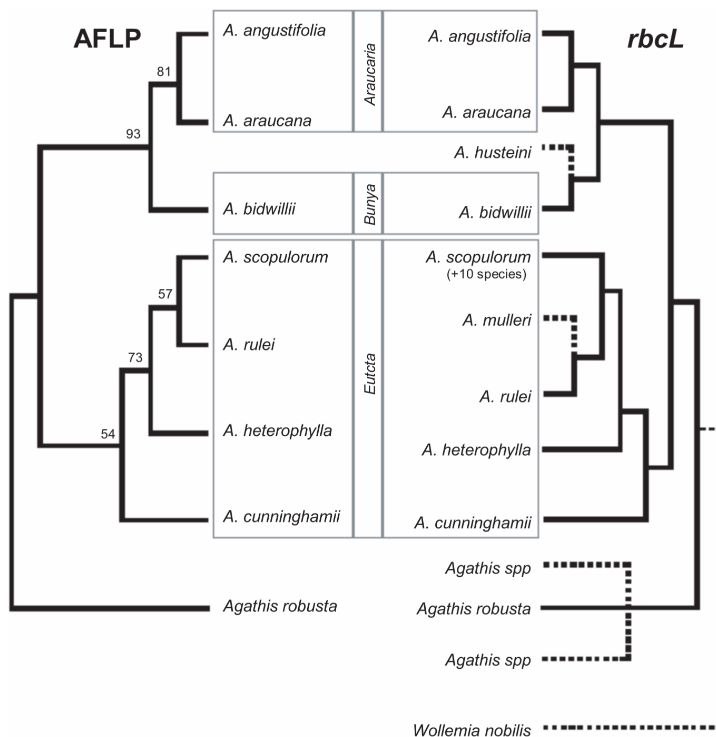

Figura 4. Sumário comparativo entre as árvores filogenéticas mais parcimoniosas baseado em marcadores AFLP e sequências do gene *rbcL*. Ramos tracejadas mostram espécies da análise por *rbcL* não incluídas no estudo por AFLP (Extraído de STEFENON; GAILING; FINKELDEY, 2006).

A *Araucaria angustifolia* (Bert.) O. Kuntze é uma conífera subtropical da família Araucariaceae, conhecida como araucária, pinheiro do Paraná ou pinheiro brasileiro. No Brasil, a *A. angustifolia* encontra-se distribuída por uma vasta área do planalto sul, em regiões acima de 500 m e entre 18º e 30° de latitude Sul. Sua ocorrência é predominante nos estados do sul do Brasil e alguns pontos descontínuos no sudeste, além da região de “Misiones” na Argentina, com registros de

uma ou duas populações cruzando a fronteira com o Paraguai (FARJON; FILER, 2013).

A. angustifolia é a gimnosperma nativa de maior importância econômica do Brasil. Devido às suas características de alto interesse, como qualidade de madeira, resina e sementes, esta espécie foi alvo de alta exploração ao longo do século passado (GUERRA et al., 2002). Essa exploração ocorreu de forma predatória do ponto de vista social, econômico ou ecológico e, em decorrência, seus remanescentes florestais estão estimados em 12,6% da sua extensão original (RIBEIRO et al., 2009). Esses fatores levaram esta espécie a ser classificada como espécie em perigo de extinção na lista brasileira de espécies ameaçadas (MINISTÉRIO DO MEIO AMBIENTE, 2008) e em perigo crítico de extinção segundo catálogo da IUCN (IUCN, 2014).

A manutenção das populações naturais é dependente da regeneração natural, e conhecer essa característica é fundamental para propor ações de recuperação das populações (PALUDO; MANTOVANI; REIS, 2011). No caso da *A. angustifolia*, foi demonstrado que grande parte da mortalidade na espécie está concentrada na classe de plantas medindo até 50 cm de altura, considerando esse um estágio crítico para a regeneração natural da espécie (PALUDO; MANTOVANI; REIS, 2011). Esse baixo sucesso no estabelecimento de novos indivíduos é um dos principais fatores que caracteriza as populações naturais de *A. angustifolia* como em leve declínio (PALUDO et al., 2016). Esse declínio só não é mais expressivo pelo fato dos indivíduos adultos da espécie apresentarem alta longevidade, corroborando a classificação da espécie como pioneira longeva (PALUDO et al., 2016).

Além dos estudos visando compreender o processo de regeneração natural para a *A. angustifolia*, diversos estudos de estrutura e diversidade genética em populações naturais dessa espécie têm sido realizados por meio de marcadores isoenzimáticos (DOS REIS et al., 2012; FERREIRA et al., 2012; MANTOVANI; MORELLATO; DOS REIS, 2006), microssatélites (BITTENCOURT; SEBBENN, 2007; SANT’ANNA et al., 2013; STEFENON; GAILING; FINKELDEY, 2006), AFLP (STEFENON; GAILING; FINKELDEY, 2006) e PCR-RFLP (SCHLOGL; DE SOUZA; NODARI, 2007) no intuito de auxiliar no desenvolvimento de estratégias de manejo e conservação dessa espécie. Estudos ecológicos e etnobotânicos apontam que a conservação da *A. angustifolia* muito mais eficiente se associada ao uso tradicional, visto que as paisagens encontradas hoje na Mata de Araucária foram

moldadas pelos seus habitantes, refletindo o padrão de uso do passado e do presente (REIS; LADIO; PERONI, 2014).

PODOCARPACEAE

A família Podocarpaceae apresenta distribuição principalmente no Hemisfério Sul, com extensões no Hemifério Norte até o México e Japão (FARJON; FILER, 2013). Atualmente, a família Podocarpaceae está dividida em 18 gêneros (LITTLE; KNOPF; SCHULZ, 2013), sendo que, o gênero *Podocarpus* responde pela maior parte da distribuição da família, englobando mais de 170 espécies, muitas das quais apresentam grande interesse florestal (FARJON; FILER, 2013). Além disso, o gênero apresenta ampla distribuição geográfica mundial, sendo registrada sua presença em Nova Caledônia, Sudeste da Ásia, China, Japão, Malásia, Austrália, Nova Zelândia, Bornéu, Nova Guiné, Ilhas do Pacífico, Ilhas Fiji, Antilhas, Américas Central e do Sul (KELCH, 1997, 1998).

No Brasil, apresentam ocorrência natural oito espécies do gênero *Podocarpus* (*P. acuminatus*, *P. aracensis, P. barretoi, P. brasiliensis, P. celatus, P. lambertii, P. roraimae*, *P. sellowii*) e o *R. piresii* (JARDIM BOTÂNICO DO RIO DE JANEIRO, 2016).

O gênero *Retrophyllum* foi reconhecido apenas na década de 80 do século passado, sendo que, até então, o gênero *Decussocarpus* compreendia três dos atuais gêneros da família Podocarpaceae: *Nageia*, *Retrophyllum* e *Afrocarpus* (PAGE, 1989). Dessa forma, a espécie *Retrophyllum piresii* (Silba) C.N. Page foi inicialmente classificada no gênero *Decussocarpus*, sendo em 1988, reclassificada para o gênero *Retrophyllum*. Atualmente, o gênero *Retrophyllum* compreende apenas cinco espécies: *R. comptonii*, *R. minor*, *R. piresii*, *R. rospigliosii* e *R. vitiense*. O gênero *Retrophyllum* difere de todos os outros membros existentes da família Podocarpaceae por ter folhas reduzidas e inclinadas nos ramos e folhas maiores e fotossinteticamente dominantes nos brotos laterais (Fig. 5) (HILL; POLE, 1992).

Até onde se sabe, *R. piresii* é uma espécie endêmica do Parque Nacional dos Pacaás Novos, Rondônia, Brasil. Essa espécie foi classificada em 1976 por João M. Pires, levando o nome de *Decussocarpus piresii* J. Silba em homenagem ao seu descritor. Nessa viagem foi realizada a coleta de algumas sementes dos espécimes encontrados no Parque Nacional dos Pacaás Novos, que foram levadas ao Museu Paraense Emílio Goeldi, onde atualmente se encontram exemplares dessa espécie (Willian Rodrigues, Comunicação pessoal).

Figura 5. Ramo da espécie *Retrophyllum piresii*. Foto: Angelo Heringer.

No momento da coleta, João M. Pires ainda não tinha certeza sobre a correta classificação da espécie, havendo dúvidas se o *R. piresii* não se tratava apenas de uma variedade botânica da espécie andina *R. rospigliosii* (LISBOA; ALMEIDA, 1995). No entanto, uma recente revisão taxonômica sobre o gênero *Retrophyllum* trouxe caracteres adicionais para separar a espécie *R. piresii* da espécie *R. rospigliosii* (MILL, 2016). Além disso, MILL, (2016) cita o trabalho apresentado no primeiro capítulo dessa tese (VIEIRA et al., 2016b) confirmando que o material coletado no Museu Paraense Emílio Goeldi foi verificado por ele e corretamente identificado como *R. piresii*.

O *R piresii* é uma árvore de aproximadamente 30 m de altura, com tronco em linha reta e ereto, ocorrendo em altitudes menores de 1000 m (MILL, 2016). Devido à dificuldade de acesso ao local de ocorrência natural da espécie, pouco se tem de estudo/informação sobre a mesma.

ESTRUTURA E EVOLUÇÃO DO GENOMA PLASTIDIAL

A mitocôndria e o cloroplasto foram originados a partir de organismos procariotos de vida livre em decorrência do processo de endossimbiose, podendo ser considerados a forma mais completa de transferência horizontal de genes, a transferência de um genoma completo (BOCK, 2010; TIMMIS et al., 2004). Após a endossimbiose, durante o processo de evolução, o genoma plastidial foi submetido a uma drástica redução de tamanho e, por conseguinte, os genomas plastidiais contemporâneos contêm apenas uma pequena proporção dos genes dos seus ancestrais. Calcula-se que essa redução foi de em torno de 3.000 genes para cerca de 120 genes, resultando em uma capacidade limitada de codificação do genoma plastidial e levando os cloroplastos a importar mais de 95% das suas proteínas do citosol (BOCK, 2007).

As proteínas codificadas pelo genoma nuclear que são destinadas a membrana do tilacóide, estroma e ao envelope interior são sintetizadas com uma extensão amino-terminal chamada de pré-sequência (SOLL; SCHLEIFF, 2004). Essa pré-sequência é reconhecida pelo translocon da membrana exterior (TOC) e da membrana interior (TIC), que permitem, respectivamente, sua passagem pelos sistemas de membrana exterior e interior, que circundam a organela (SOLL; SCHLEIFF, 2004). Finalmente, no estroma ocorre a clivagem da extensão amino-terminal, produzindo a forma madura da proteína (SOLL; SCHLEIFF, 2004).

Além disso, devido à sua origem procariótica, o genoma plastidial manteve diversas características de um procarioto, incluindo uma estrutura de genoma do tipo circular, organizada em nucleóides, organização de genes em operons, e uma maquinaria de expressão gênica procariótica (BOCK, 2007). Por exemplo, o DNA plastidial apresenta um padrão de metilação diferente do DNA nuclear. O DNA de cromoplastos, leucoplastos e amiloplastos contém vários sítios metilados, ao contrário do DNA do cloroplasto, que praticamente não é metilado (VANYUSHIN, 2006). Supõe-se que a metilação de DNA esteja associada com a diferenciação nos plastídios. Foi demonstrado através da introdução de dois genes de cianobactérias (adenina e citosina DNA metiltransferase) no genoma plastidial de *Nicotiana tabacum*, que as plantas transformadas com DNA plastidial metilado não apresentaram diferenças morfológicas ou no padrão de expressão gênica em relação ao controle não transformado (AHLERT et al., 2009). Estes fatos sugerem

que a metilação não está envolvida na regulação da transcrição em plastídios (AHLERT et al., 2009).

O DNA de plastídios e mitocôndrias é herdado de forma não-Mendeliana em todos os eucariotos. Na maioria dos organismos, essa herança é uniparental, sendo a herança materna muito mais difundida do que a herança paterna (REBOUD; ZEYL, 1994). Sabe-se que a maior parte das angiospermas possui herança materna, com raros casos de herança biparental ou paterna (ZHANG; SODMERGEN, 2010). Já as gimnospermas possuem majoritariamente herança paterna (JANSEN; RUHLMAN, 2012).

Nas plantas terrestres, a maior parte dos genomas plastidiais é divida em quatro partes: grande região de cópia simples (LSC), pequena região de cópia simples (SSC) e duas regiões invertidas repetidas (IR_A e IR_B) (Figura 6). As duas regiões IRs são idênticas em composição, sendo todos os genes contidos nessa região apresentam no mínimo duas cópias por genoma, porém em sentido de leitura contrário.

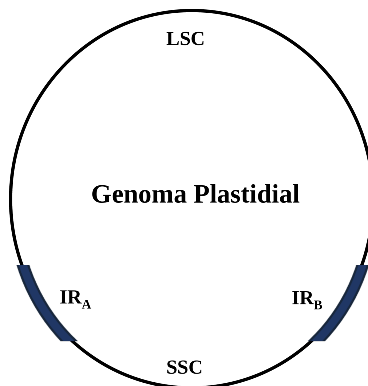

Figura 6. Organização estrutural típica do genoma plastidial, dividido em região grande de cópia simples (LSC), região pequena de cópia simples (SSC) e duas regiões invertidas repetidas (IR_A e IR_B).

Uma evidência circunstancial para a ação da conversão gênica nas IRs vem da observação de que a frequência de mutação de genes das regiões IR é significativamente menor do que para genes localizados em ambas as regiões de cópia única do genoma plastidial (MAIER et al., 1995). Em pteridófitas, quatro loci, correspondendo ao genes *psbA*, *ycf2*, *rps7* e exons 2-3 do *rps12*, foram avaliados antes e depois da sua translocação da SSC para a IR (LI et al., 2016a). Os quatro loci mostram uma desaceleração de duas a três vezes na sua taxa de substituição após

a translocação para o IR, um fenômeno não observado em quaisquer outros genes plastidiais não transferidos (LI et al., 2016a).

No entanto, foi relatado que algumas famílias de gimnospermas não apresentam uma das regiões IR, como é o caso da Araucariaceae (RUHSAM et al., 2015; WU; CHAW, 2014b), Podocarpaceae (VIEIRA et al., 2014a, 2016b; WU; CHAW, 2014b), Cupressaceae (HIRAO et al., 2008; WU et al., 2011a), Cephalotaxaceae (WU et al., 2011a) e algumas espécies da família Pinaceae (WU et al., 2011b). Adicionalmente, foi sugerido é possível determinar qual cópia da IR foi perdida nessas espécies (IR_A ou IR_B) e que as espécies da família Cupreassaceae perderam diferentes cópias das regiões IRs em relação as espécies da família Pinaceae (WU et al., 2011a; WU; CHAW, 2014b). Este relato é contestado pelo trabalho de YI et al., (2013), que diz que não é possível distinguir entre hipóteses que favoreçam a retenção da mesma região IR em genomas de Cupressophytas e Pinaceae a partir de uma hipótese propondo perda independente da IR_A ou IR_B.

Comparando o genoma plastidial de seis espécies de Cupressophytas (envolvendo quatro famílias), foi demonstrado que o genoma plastidial dessas espécies evoluiu de forma a haver uma redução de tamanho, em grande parte devido à diminuição de tamanho dos espaçadores intergênicos (WU; CHAW, 2014b).

No que diz respeito às análises evolutivas, o estudo do genoma plastidial apresenta algumas características interessantes: 1) apesar de pequeno, o DNA plastidial é um componente abundante do DNA celular total; 2) o genoma plastidial tem sido extensivamente caracterizado à nível molecular, fornecendo informações para apoiar a pesquisa evolutiva comparativa; 3) as taxas de substituição de nucleotídeos são relativamente baixas e, portanto, fornecem resolução adequada para estudos de filogenia de plantas (CLEGG; LEARN; MORTON, 1994). Além disso, devido a sua fácil amplificação, tamanho pequeno, e diferenças nas taxas evolutivas das regiões codificantes e não codificantes, é possível a comparação/medição de distância evolutiva em diferentes níveis taxonômicos (PALMER, 1985; SHAW et al., 2007, 2014).

Nos anos 80, foram iniciados os estudos filogenéticos baseados no genoma plastidial (PALMER, 1985). O gene *rbcL*, que codifica a grande subunidade da ribulose-1,5-bifosfato-carboxilase/oxigenase (RUBISCO), passou a ser amplamente sequenciado a partir de diferentes e numerosos taxas de plantas, gerando uma boa base de dados para estudos de filogenia de plantas (PALMER et al., 1988). No entanto,

em algumas situações as relações continuavam a gerar dúvidas pelo fato do gene *rbcL* ser muito conservado para esclarecer dúvidas entre gêneros próximos (GIELLY; TABERLET, 1994). A análise de regiões não-codificantes do genoma plastidial, como introns e espaçadores intergênicos, foi a estratégia encontrada para esclarecer as relações em níveis taxonômicos mais baixos. Estas regiões tendem a evoluir mais rapidamente do que as regiões codificantes, por acumulação de inserções/eliminações a uma taxa pelo menos igual à das substituições de nucleotídeos (GIELLY; TABERLET, 1994).

Ainda que o uso de regiões não-codificantes tenha solucionado algumas das dúvidas no âmbito dos estudos filogenéticos, percebeu-se mais tarde que muitas regiões não exploradas do genoma plastidial poderiam trazer informações adicionais a essa linha de estudo (SHAW et al., 2007, 2014). De fato, o sequenciamento completo do genoma plastidial mostrou regiões ainda inexploradas com maior número de caracteres para estudos filogenéticos em níveis taxonômicos mais baixos (SHAW et al., 2007, 2014). No entanto, não há uma região única ou grupo de regiões mais indicado para todas as linhagens taxonômicas, sendo necessário um rastreamento em cada grupo taxonômico de interesse para determinar quais regiões são as mais adequadas (SHAW et al., 2007).

Um estudo filogenético utilizando 78 sequências que codificam para proteínas, de 360 genomas plastidiais de plantas disponíveis no banco de dados do GenBank, mostrou sustentação para relações fundamentais entre clados já bem descritas, mas também forte sustentação para relações que não foram observadas em análises anteriores de clados maiores dentro de Viridiplantae (RUHFEL et al., 2014). Essas análises também mostraram em gimnospermas a inesperada relação irmã entre Gnetophyta e as coníferas do Clado II (Araucariales + Cupressales) (RUHFEL et al., 2014).

Em Pinaceae, análises filogenômicas plastidiais com representantes de todos os dez gêneros da família foram realizadas, resultando em um alta sustentação para o clado formado pelos cinco gêneros da subfamília Abietoideae, com uma relação irmã ao clado formado pelas outras três subfamílias: Pinoideae, Piceoideae e Laricoideae (SUDIANTO et al., 2016).

Em Araucariales, os genomas plastidiais de 14 espécies do gênero *Araucaria* foram sequenciados para realização de análises filogenéticas a partir do alinhamento da sequência completa do genoma plastidial (RUHSAM et al., 2015). As informações deste trabalho

geraram uma árvore filogenética com resolução maior do que qualquer árvore publicada anteriormente com este gênero. Além disso, o monofiletismo foi altamente suportado em cerca de duas vezes mais clados do que conjunto de dados nucleares (11 genes) usados para comparação (RUHSAM et al., 2015).

OBJETIVOS

OBJETIVO GERAL

Realizar estudos estruturais, filogenômicos e evolutivos em Araucariaceae e Podocarpaceae a partir do sequenciamento completo do genoma plastidial das espécies *Araucaria angustifolia*, *Araucaria araucana*, *Araucaria bidwillii* e *Retrophylum piresii*.

OBJETIVOS ESPECÍFICOS

a) Isolar e sequenciar os DNA plastidial dessas espécies;
b) Realizar a montagem e anotação dos genomas a partir dos dados de sequenciamento obtidos;
c) Comparar a estrutura e conteúdo gênico dos genomas plastidiais obtidos e outras espécies de Araucariaceae e Podocarpaceae;
d) Realizar análise de filogenia entre as três espécies de *Araucaria* sequenciadas nesse trabalho e as outras quatorze espécies com genoma plastidial disponível no banco de dados GenBank;

CAPÍTULO 1.

Este capítulo encontra-se publicado no periódico “Tree Genetics & Genomes”

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

VIEIRA, L. N. et al. The plastome sequence of the endemic Amazonian conifer, *Retrophyllum piresii* (Silba) C.N.Page, reveals different recombination events and plastome isoforms. Tree Genetics and Genomes, v. 12, n. 1, p. 1–11, 2016b.

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 gymosperms IR can vary from large in size to completely absent, being constituted principally by 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.

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 (BIFFIN; CONRAN; LOWE, 2011; FARJON, 1998). The *Retrophyllum* genus comprises five species: *Retrophyllum comptonii*, *R. minor*, *R. 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; TERBORGH; CUNNINGHAM, 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, 2010; PARKS; CRONN; LISTON, 2009; VIEIRA et al., 2014a; WU et al., 2011a; YI et al., 2013). Plastome sequences are available for all families of conifers: Cephalotaxaceae (YI et al., 2013), Cupressaceae (HIRAO et al., 2008), Pinaceae (CRONN et al., 2008; LIN et al., 2010; WAKASUGI et al., 1994), Taxaceae (ZHANG et al., 2014), Araucariaceae (RUHSAM et al., 2015; WU; CHAW, 2014b), and Podocarpaceae (VIEIRA et al., 2014a; WU; CHAW, 2014b). 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; CHAW, 2014b).

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 (KNOX, 2014; PALMER, 1983). The size of IRs in plastids of land plants is highly variable and it is dependent on plant group, genus, family or species (GUO et al., 2014; GURDON; MALIGA, 2014; JANSEN; RUHLMAN, 2012; VIEIRA et al., 2014a; WICKE et al., 2011). The IRs copies recombine themselves, and are intended to maintain or confer stability to the remaining plastome (KNOX, 2014; PALMER, 1983; STEIN; PALMER; THOMPSON, 1986).

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 (GUO et al., 2014; LIN et al., 2012; WU et al., 2007, 2009). In conifers, there are short IRs regions, containing different genes, but principally tRNA genes or a part of other genes sequence. Recently, in

species of the Juniperus genus, it was observed the presence of short IRs containing two copies of full *trnQ-UUG* (GUO et al., 2014). 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; 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 (ALKATIB et al., 2012; GRAY; AHNER; HANSON, 2009; ROGALSKI et al., 2008; ROGALSKI; KARCHER; BOCK, 2008; ROGALSKI; RUF; BOCK, 2006). The genome rearrangement is dependent on the direction of the two repeated sequences (i.e. direct repeats or inverted repeats). Whether the sequences are presented as direct repeats the sequence between them and one of them are deleted from the plastome (ALKATIB et al., 2012; ROGALSKI; KARCHER; BOCK, 2008), whereas if the sequences are found as inverted repeats they recombine and induce an inversion of the sequence between them (ROGALSKI et al., 2008; ROGALSKI; RUF; BOCK, 2006).

Here we demonstrated the presence of recombinationally active repeated sequences, consisting of different copies of tRNA genes, one as inverted and the other as direct repeat in the same plastome. These repeated sequences produce an IR-mediated inversion and a 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

(Conifers 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 (conifers 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).

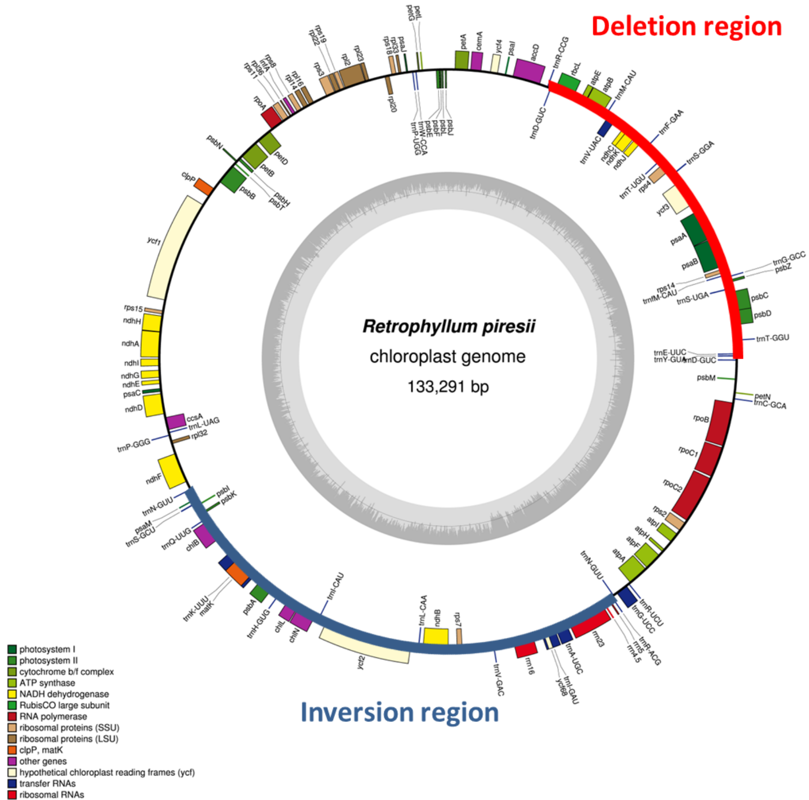

Figure 1. Gene map of *Retrophyllum piresii* plastome. Genes drawn inside the circle are transcribed clockwise, and genes drawn outside are counterclockwise. Genes belonging to different functional groups are color-coded. The darker gray in the inner circle corresponds to GC content, and the lighter 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.

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 direct repeat sequences, respectively. The following genes were identified and are listed in Figure 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 (VIEIRA et al., 2014a; WU; CHAW, 2014b). 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 (HIRAO et al., 2008; HUANG; MATZKE; MATZKE, 2013; VIEIRA et al., 2014a; WU et al., 2007, 2011a; WU; CHAW, 2014b). 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; TANGPHATSORNRUANG et al., 2010), Dioscoreaceae (HANSEN et al., 2007), and Melanthiaceae (DO; KIM; KIM, 2014) families.

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

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

* Genes containing introns. ** Duplicated gene.

Repeat Sequence Analysis

The plant population genetic studies may be greatly facilitated by the use of chloroplast DNA 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; POWELL; HOLLINGSWORTH, 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,

chloroplast DNA may be applied to studies involving genetic structure of natural populations due to its mode of inheritance in comparison to nuclear markers (PROVAN; POWELL; HOLLINGSWORTH, 2001).

Hence, chloroplast 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 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 (BASKAUF et al., 2014; KATO et al., 2011, 2013; ROULLIER et al., 2013).

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), 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), which is in accordance to the nature of chloroplast microsatellites of generally <15 mononucleotide repeats (PROVAN; POWELL; HOLLINGSWORTH, 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 intra-specific 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 SSR 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 endemics region, Pacaás Novos National Park – Brazil.

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

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) (KNOX, 2014; PALMER, 1983; SHINOZAKI et al., 1986). This plastome organization is highly conserved in angiosperms, with very few exceptions (DO; KIM; KIM, 2014; GUO et al., 2007; GURDON; MALIGA, 2014; HANSEN et al., 2007; TANGPHATSORNRUANG et al., 2010). In gymnosperms, the loss of the large IR has been reported in several species, mainly in conifers (HIRAO et al., 2008; WU; CHAW, 2014a; 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 (VIEIRA et al., 2014a; WU; CHAW, 2014a; YI et al., 2013). As in other species of Podocarpaceae family (VIEIRA et al., 2014a; WU; CHAW, 2014a) the plastome of *R. piresii* lacks one of the IRs (Figure 1).

Comparing the plastome of *R. piresii* with *P. lambertii* and *N. nagi* by dot-plot analyses (Figure 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 (Figure 3). PCR amplification with several primer combinations confirmed that indeed, this IR-mediated inversion produced two different isoforms of the *R. piresii* plastome (Figure 3). The plastid DNA used for the PCR amplification was isolated from the same plant and revealed that both isomers coexist in a single *R. piresii* plant (Figures 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).

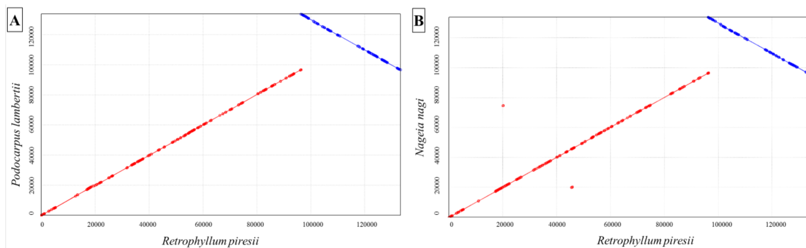

Figure 2. Dot-plot analyses of *Podocarpus lambertii* and *Nageia nagi* plastome sequence against *Retrophyllum piresii*. A positive slope denotes that the compared two sequences are in the same orientations, whereas a negative one indicates that the compared sequences can be aligned but their orientations are opposite. Graphs represents comparisons between *R. piresii* (axis X) and *P. lambertii* (axis Y) (A), and *R. piresii* (axis X) and *N. nagi* (axis Y) (B).

Although, the two isoforms differ in the orientation of a 56 kb segment of the plastome (Figure 1), they are functionally equivalent, considering that they both carry the same gene content and do not affect the integrity of other chloroplast genes (Figure 1). The different isoforms were readily detectable by PCR, and it is highly unlikely that PCR artifacts are 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; 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 plastomes 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 because homologous recombination is a randomly physical mechanism and is distributed due to random segregation of the isoforms during cell and organelle division (GUO et al., 2014; ROGALSKI et al., 2008; ROGALSKI; RUF; BOCK, 2006).

Analyzing different gymnosperms 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, conifers clade I present *trnI-CAU*, *trnS-GCU*, and *trnH-GUG* in inverted repeat, and *trnT-GGU* in direct repeat. The conifers 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., 2011a). However, in conifers clade II species, the recombinationally activity of the short IR containing *trnQ-UGG* (544 bp) was found to occur in *Cephalotaxus oliveri* (Cephalotaxaceae) but not in *C. japonica* (Cupressaceae) and *Taiwania 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; CHAW, 2014a), 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,

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*	*trnS-UGA*
Conifers Clade I															
Pinaceae	*Cathaya argyrophylla*	I											I		
	Cedrus deodara	I											I	I	
	Keteleeria davidiana	I											I		
	Larix decidua	I											I		
	Picea abies	I				I							I	D	
	Picea morrisonicola	I				D							D		
	Pinus koraiensis	I				I							I	D	
	Pinus krempfii	I				I							I		
	Pinus lambertiana	I				I							I		
	Pinus massoniana	I				I							I	D	
	Pinus monophylla	I				I							I	D	
	Pinus nelsonii	I											I	D	
	Pinus taeda	I				I							I	D	
	Pinus thunbergii	I				I							I	D	
	Pseudotsuga sinensis	I											I	D	D
Conifers Clade II															
Araucariaceae	*Agathis dammara*	I										D			
Cupressaceae	*Calocedrus formosana*	I													
	Cryptomeria japonica	I	I												
	Cunninghamia lanceolat	I	I												
	Juniperus bermudiana		I												
	Juniperus monosperma	I	I												
	Juniperus scopulorum	I	I												
	Juniperus virginiana	I	I												
	Taiwania cryptomerioide	I	I		I										
	Taiwania flousiana	I	I		I										
Cephalotaxaceae	*Cephalotaxus oliveri*		I												
	Cephalotaxus wilsoniana	D	I												
Podocarpaceae	*Nageia nagi*			I											
	Podocarpus lambertii			I											
	Podocarpus totara											D			
	Retrophyllum piresii			I								D			
Taxaceae	*Taxus mairei*	I	I												
Cycadidae															
Cycadaceae	*Cycas revoluta*			I	I	I	I	I	I	I					
	Cycas taitungensis			I	I	I	I	I	I	I					
Ginkgoidae															
Ginkgoaceae	*Ginkgo biloba*			I	I	I	I	I	I	D					
Gnetidae															
Ephedraceae	*Ephedra equisetina*	I		I	I	I	I	I	I	I					
Gnetaceae	*Gnetum montanum*	I		I	I	I	I	I	I	I					
	Gnetum parvifolium	I		I	I	I	I	I	I	I					
Welwistchiaceae	*Welwitschia mirabilis*	I		I	I	I	I	I	I		D				

I: Inverted repeat; D: Direct repeat.

Paris verticillata, although, at first analyses, no rearrangements were observed (DO; KIM; KIM, 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) (Figure 1). We investigated whether this DR could recombine and cause the deletion of its internal content. PCR data containing amplified products with suitable primers (Figure 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 (Figure 3). This hypothesis was confirmed by mapping the paired-end reads with both plastome isoforms (Electronic Supplementary Material 2).

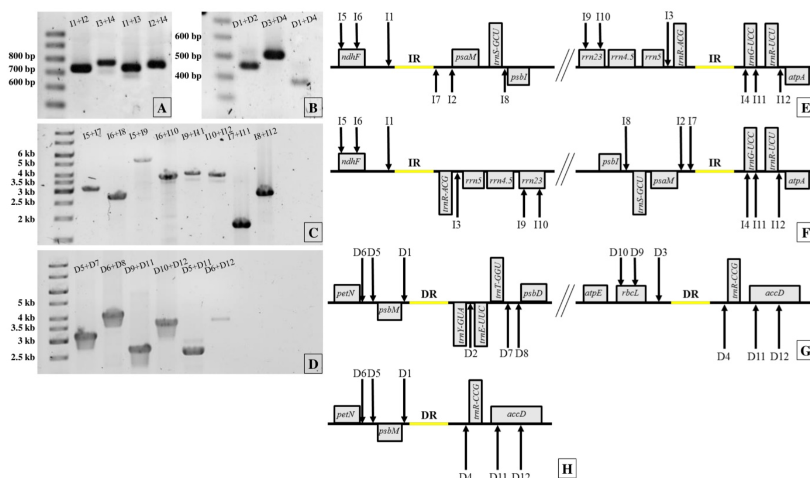

Figure 3. PCR analysis of recombinant genomes. A – PCR amplification products for IR-mediated inversion with 100 bp ladder; B - PCR amplification products for DR-mediated deletion with 100 bp ladder; C - PCR amplification products for IR-mediated inversion with 1 kb ladder; D - PCR amplification products for DR-mediated deletion with 1 kb ladder; E - PCR primer combination designed to amplify genome IR-mediated inversion, isoform 1; F and 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 formed in the position of *trnN-GUU*. DR indicates the short direct repeat formed in the position of *trnD-GUC*. In figure H one copy of the direct repeat is deleted and only one copy remains.

In transgenic plastids, the appearance of unexpected plastome conformations was observed when endogenous regulatory sequences were used (ALKATIB et al., 2012; FLEISCHMANN et al., 2011; ROGALSKI et al., 2008; ROGALSKI; KARCHER; BOCK, 2008; ROGALSKI; RUF; BOCK, 2006). 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., 2008) and, otherwise, if they were found as IR they can work as flip-flop recombination (ROGALSKI; KARCHER; BOCK, 2008; ROGALSKI; RUF; BOCK, 2006).

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 (DAY et al., 2004; IAMTHAM; DAY, 2000) 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 (ALKATIB et al., 2012; FLEISCHMANN et al., 2011; ROGALSKI et al., 2008; ROGALSKI; KARCHER; BOCK, 2008; ROGALSKI; RUF; BOCK, 2006).

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? To deepen these questions and others can help unraveling important aspects of the adaptive evolution of conifers.

MATERIAL AND METHODS

Plant Material and cp DNA Purification

Chloroplast isolation of *R. piresii* was performed from a single individual fresh leafs gently provided by Goeldi Museum (Museu Paraense Emílio 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 x 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; JANSEN; BOORE, 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; BROOKS; LOWE, 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 cpDNA was amplified in a reaction mixture containing 200 µM of each dNTP (Sigma-Aldrich), 2.0 mM $MgCl_2$ (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 94º C, 30 s at 63º C and 60 s at 72º C with a 3 min extension of the first cycle at 94º C and a 10 min final extension at 72º 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 of the software CLC Genomics Workbench 6.5v.

Table 4. Set of primers used in PCR amplification.

Primer	Sequence (5'-3')
I1	GGAATTCACCTGGGACAGAT
I2	GGCTACAGACTGTTCTCTCCTTG
I3	GAATAGCTGAGAGCGGATCAAG
I4	CCTTGAGAACATTCTGTTTCG
I5	GCCCCAATCTGGTACAAATC
I6	GGGACAAGAATCCAGCACAT
I7	AATTACAGCTGCTCTGGGCTA
I8	CCCGGACGTAAAGGATAGTG
I9	GCTAAGGCCCCTAAATGACC
I10	AAAGCTCGAACGACCGTTTA
I11	TTCATTTCATTCGGCTCCTC
I12	CCTCATTGAGCGCAAAAGAT
D1	ATTTCTCTGGGGAGGAGCAT
D2	TTATGGTTCCTGGGTCGATG
D3	ATGGCGGCATAATTAACCAA
D4	AATAGGGCTCGGGCTAGAAG
D5	CCGCCTCATGATCCATATTC
D6	TTGCGTTGATCGGTCTTAGTC
D7	TTGTCTCGGGATTCTGCTTT
D8	TGATGTGAGCGAATAGGAACA
D9	CCAAAACTTTCCAAGGTCCA
D10	CTGGGTTCAAAGCTGGTGTT
D11	CGGTTCATCTGCCAAGAAAT
D12	AGAGGATCCCTTGCAGTTCA

DATA ARCHIVING STATEMENT

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

ACKNOWLEDGEMENTS

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 to Prof. Dr. Willian Antônio Rodrigues for the information related to the species, to Angelo Schuabb Heringer and Ramon Felipe Scherer for collecting plant material, and Museu Paraense Emílio Goeldi for kindly providing plant material.

CAPÍTULO 2.

Este capítulo encontra-se formatado de acordo com a normativa do periódico “Tree Genetics & Genomes”

Plastid genome structure and phylogenomic analysis within *Araucaria* genus

VIEIRA, L. N. et al. Plastid genome structure and phylogenomic analysis within *Araucaria* genus.

ABSTRACT

Araucariaceae is considered one of the most ancient family among extant conifers, but the phylogenetic analyses within this family have shown contrasting results. Plastid genome sequencing is a low cost technique, and its sequencing and analyzes is the most direct way to elucidate relationships between species. Here, we sequenced and analyzed the complete plastome sequences of three species from two previously unavailable *Araucaria* sections, *Araucaria* (*Araucaria angustifolia*, *Araucaria araucana*), and *Bunya* (*Araucaria bidwillii*). A plastid phylogenomics analysis was performed based on 17 *Araucaria* species. *A. angustifolia*, *A. araucana,* and *A. bidwillii* plastomes were identical in gene content and order, including the *rps16* gene loss, and the presence of a *rrn5*-containing short IR (~580 bp). Both Maximum Likelihood estimation and Bayesian Inference analyses produced phylogenomic trees identical in topology. The two *Araucaria* sections (*Araucaria* and *Eutacta*) formed monophyletic and highly supported clades. Section *Bunya* is monotypic and formed a highly supported sister relationship with *Araucaria* section. These analyses also supported *Eutacta* as monophyletic and confirmed *Araucaria heterophylla* as sister to the two well supported lineages consisting of the coastal + small-leaved clade, and the large leaved clade. Our results showed an unprecedented phylogenomic inference within genus *Araucaria* including *Bunya* and *Araucaria* sections.

Keywords: Phylogeny, Illumina, next-generation sequencing, plastome, conifers

ACKNOWLEDGEMENTS

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).

INTRODUCTION

Conifers are worldwide distributed, but only four of the six conifer families naturally occurs in the Southern Hemisphere: Araucariaceae, Podocarpaceae, Cupressaceae, and Taxaceae (FARJON; FILER, 2013). Araucariaceae is restricted do Southern Hemisphere, but much of its fossil records are located in the Northern Hemisphere (HILL, 1995), from where this family disappeared in the late Cretaceous (KUNZMANN, 2007). Araucariaceae is considered one of the most ancient families among extant conifers, with some species called "living fossils" (KUNZMANN, 2007; WILLIAMS, 2009).

Nowadays, Araucariaceae is distributed throughout South America, Australia, New Guinea, New Caledonia, Norfolk Island, and other Pacific islands (SETOGUCHI et al., 1998). This family comprises three genera: *Araucaria*, *Agathis*, and *Wollemia*. The genus *Araucaria* has 19 species divided into four sections: *Araucaria* (2 ssp.), *Bunya* (1 sp.), *Eutacta* (15 ssp.), and *Intermedia* (1 sp.). The two species from section *Araucaria* are indigenous and endemic to South America: *Araucaria angustifolia* (Bert.) O. Kuntze and *Araucaria araucana* (Molina) K. Koch (BRODRIBB; HILL, 1999). The only extant species belonging to section *Bunya* is *Araucaria bidwillii* Hooker, whose has a very limited dispersion, endemic to the Bunya Mountains, north of Australia (FARJON; FILER, 2013).

Phylogenetic analysis in Araucariaceae showed conflicting results (DETTMANN; CLIFFORD, 2005). Cladistic analysis based on *matK* and *rbcL* sequence data, and one combined phylogenetic analyses (morphological and molecular data), placed *Wollemia* and *Agathis* in the same clade (= agathioid clade) (ESCAPA; CATALANO, 2013; WAGSTAFF, 2004). In contrast, phylogenies based on *rbcL* gene sequences placed *Araucaria* and *Agathis* as sister taxa (SETOGUCHI et al., 1998). Nevertheless, all these analysis support and agree with the monophyletic origin of the three distinct genera of Araucariaceae family.

Recently, plastid genomes (plastomes) from 13 species of New Caledonian *Araucaria* species and *Araucaria heterophylla* from Norfolk Island were sequenced to perform phylogenetic analysis (RUHSAM et al., 2015). Plastome phylogenetic results presented higher resolution than any previously published tree within the genus *Araucaria* (RUHSAM et al., 2015). Similarly, Pinaceae phylogeny, which has also been contentious, could be accurately revisited after the plastome sequencing of 11 Pinaceae species (SUDIANTO et al., 2016).

The plastome is present in high copy numbers per cell for most plants. In closely related taxa, plastomes show little variation in gene content and order (VIEIRA et al., 2016a), providing a wealth and relatively easy phylogenetically informative data (ROGALSKI et al., 2015; RUHFEL et al., 2014). The sequencing and analyses of plastomes are the most direct way to reconstruct deep evolutionary history and to elucidate relationships between taxa at multiple levels of biological organization (LEISTER, 2016). The plastome sequencing is becoming a low cost and rapid technique since the advancement of next-generation sequencing. Thus, plastome data available in the public databases comprise an efficient option for evolutionary studies and increased phylogenetic resolution (ROGALSKI et al., 2015). In fact, plastome sequence data have transformed plant systematics and contributed to the current view of plant relationships (RUHFEL et al., 2014).

Until now, only species of section *Eutacta* had their plastome sequenced and analyzed. Here, we sequenced and analyzed the complete plastome sequence of the two species of section *Araucaria* (*A. angustifolia*, *A. araucana*) and the single species that comprises section *Bunya* (*A. bidwillii*). We also performed a full plastome phylogeny using 17 species of genus *Araucaria*, including the three newly sequenced and 14 previous sequenced species.

MATERIAL AND METHODS

Plastid Genome Sequencing, Assembling And Annotation

Araucaria angustifolia and *A. bidwillii* leaves were collected at the Agricultural Science Department and the Department of Botany, respectively, both from the Federal University of Santa Catarina, Santa Catarina, Brazil. *A. araucana* leaves were collected in the Atacama region - Chile with previous authorization of the “Instituto Nacional de Investigación Agropecuaria” (INIA) – Chile. *A. angustifolia*, *A. araucana*, and *A. bidwillii* plastid DNA isolation were performed according to VIEIRA et al., (2014b) using fresh needles from a single individual. The isolated DNA was quantified using Qubit® fluorometer (Invitrogen, Carlsbad, CA) and 1 ng of plastid DNA was used to prepare sequencing libraries with Nextera XT DNA Sample Prep Kit (Illumina Inc., San Diego, CA). Libraries were sequenced on Illumina MiSeq (Illumina Inc., San Diego, CA) platform.

The obtained paired-end reads (2 x 300 bp) were de novo assembled using CLC Genomics Workbench 8.0v and FGAP (PIRO et al., 2014). Gene annotation was performed using Dual Organellar GenoMe Annotator (DOGMA) (WYMAN; JANSEN; BOORE, 2004) and tRNAscan-SE (SCHATTNER; BROOKS; LOWE, 2005). The plastome physical map was drawn using OrganellarGenomeDRAW (OGDRAW) (LOHSE et al., 2013). The complete nucleotide sequence of *A. angustifolia*, *A. araucana*, and *A. bidwillii* plastomes were deposited in the GenBank database under accession number KY404932, KY404933, and KY404931, respectively.

Comparative Analysis Of Plastome Structure

We used the Multiple Genome Alignment software MAUVE 2.4.0 to visualize rearrangments and inversions between Araucariaceae and Podocarpaceae plastomes. The complete plastome sequences available in GenBank usesd for this anaylsis were from Podocarpaceae species: *Retrophyllum piresii* (NC_024827), *Nageia nagi* (NC_023120), and *Podocarpus lambertii* (NC_023805); and Araucariaceae species: *Agathis dammara* (NC_023119), *Wollemia nobilis* (NC_027235), *Araucaria heterophylla* (NC_026450), and the newly sequenced *A. angustifolia*, *A. araucana*, and *A. bidwillii*.

Phylogenetic Inference

Phylogenetic inference was based on three newly sequenced species (*A. angustifolia*, *A. bidwillii*, and *A. araucana*) and 14 previously sequenced *Araucaria* spp. (Table 1). *Wollemia nobilis* and *Agathis dammara* were tested as possible outgroup species. The whole plastome sequence were alignment with MAFFT v7 (KATOH; STANDLEY, 2013), with posterior gaps exclusion. The substitution model (GTR + I + G) was selected using jModeltest 2.1.7 (number of substitution schemes = 5). Maximum likelihood (ML) analysis was performed using the RAxML v 7.2.8 (STAMATAKIS, 2006) with 1,000 non-parametric bootstrap replicates. Bayesian inference (BI) analysis was performed with MrBayes 3.2.2 (RONQUIST; HUELSENBECK, 2003). The Markov chain Monte Carlo (MCMC) analysis was run and the average standard deviation of split frequencies was below 0.01 after 2,000,000 generations. The first 25 % of samples were discarded as

burn-in. Resulting trees were represented and edited using FigTree v1.4.1.

Table 1. Plastid genome comparison of Araucariaceae species analyzed in this study.

Species	**Plastid genome size (bp)**	**GenBank accession**
Araucaria angustifolia	145,558	KY404932*
Araucaria araucana	146,005	KY404931*
Araucaria bernieri	146,696	KM678407
Araucaria bidwillii	146,042	KY404933*
Araucaria biramulata	146,728	KM678418
Araucaria columnaris	146,799	KM678417
Araucaria heterophylla	146,723	KM067155
Araucaria humboldtensis	146,892	KM678427
Araucaria laubenfelsii	146,614	KM678423
Araucaria luxurians	146,729	KM678408
Araucaria montana	146,764	KM678422
Araucaria muelleri	146,686	KM678404
Araucaria nemorosa	146,763	KM678429
Araucaria rulei	146,787	KM678425
Araucaria schmidii	146,697	KM678424
Araucaria scopulorum	146,794	KM678412
Araucaria subulata	146,730	KM678409
Wollemia nobilis	145,630	NC_027235
Agathis dammara	145,625	NC_023119

*Species with plastid genomes sequenced in this study

RESULTS AND DISCUSSION

Plastome Sequence And Structure

The Illumina MiSeq sequencing data used for genome assembly resulted in a high average plastome coverage for all three *Araucaria*

spp., ranging from 85x to 230x (Table 2). As expected, the determined plastome lenght is very similar between the three *Araucaria* species, ranging from 145,558 bp in *A. angustifolia* to 146,042 bp in *A. bidwillii* (Table 2).

Besides very similar to each other, *A. angustifolia* (145,558 bp), *A. bidwillii* (146,042 bp), and *A. araucana* (146,005 bp) plastomes were slightly smaller than the others *Araucaria* species plastomes, which ranged from 146,477 bp in *A. muelleri* to 146,892 bp in *A. humboldtensis* (RUHSAM et al., 2015). This slightly size difference may be related to the fact that the compared *Araucaria* species are from different sections. Compared to conifers, all genera from family Araucariaceae showed larger plastome size, being determined to be 145,625 bp in *Agathis dammara* (WU; CHAW, 2014a) and 145,630 bp in *Wollemia nobilis* (YAP et al., 2015). The plastome size in other conifers range from 107,122 bp in *Cathaya argyrophylla* (Pinaceae) (LIN et al., 2010) to 138,284 bp in *Sciadopitys verticillata* (Sciadopityaceae) (LI et al., 2016b). The closest conifer family to Araucariaceae, the Podocarpaceae, has average plastome size of ~133,5 kb (VIEIRA et al., 2014a, 2016b), about 20 kb smaller than Araucariaceae plastomes.

The *A. angustifolia*, *A. araucana*, and *A. bidwillii* plastomes determined GC content (36.5%) is the same from the other Araucariaceae species, *A. dammara* and *W. nobilis*, and slightly lower than Podocarpaceae species, *Podocarpus lambertii* (37.1%) and *Retrophyllum piresii* (37.2%) (VIEIRA et al., 2014a, 2016b).

Table 2. *Araucaria angustifolia*, *Araucaria bidwillii*, and *Araucaria araucana* plastid genome sequencing and assembly data. Illumina MiSeq and CLC Genomics Workbench 8.0v were used for sequencing and assembly, respectively.

Species	*A. angustifolia*	*A. bidwillii*	*A. araucana*
Plastome lenght (bp)	145,558	146,042	146,005
Nº of plastid reads	120,271	71,458	59,627
Mean read length	285.38	176.06	279.54
Average coverage	233.22	85.85	113.70

A total of 125 genes were annotated, of which 115 genes were single copy, five genes (*rrn5*, *trnI-CAU*, *trnR-UCU*, *trnY-GUA*) were duplicated, and *trnD-GUC* presented three copies (Table 3). Two genes, *rrn5* and *trnI-CAU* occurred as an inverted repeat sequence. These genes were grouped as follows (Figure 1, Table 3): 3 unique and 1 duplicated ribosomal RNA genes, 28 unique and 4 duplicated 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, 7 genes encoding other proteins, including the unknown function gene (*ycf2*), and 1 pseudogene (*ycf68*).

Table 3. List of genes identified in *Araucaria angustifolia*, *Araucaria araucana*, and *Araucaria bidwillii* plastomes.

Functional system	**Genes name**				
Ribosomal RNA genes	*rrn23*	*rrn4.5*	*rrn5*[2]	*rrn16*	
Transfer RNA genes	*trnA-UGC*[1]	*trnC-GCA*	*trnD-GUC*[3]	*trnE-UUC*	*trnF-GAA*
	trnG-GCC	*trnG-UCC*[1]	*trnH-GUG*	*trnI-GAU*[1]	*trnI-CAU*[2]
	trnL-UAA[1]	*trnL-UAG*	*trnL-CAA*	*trnK-UUU*[1]	*trnM-CAU*
	trnfM-CAU	*trnN-GUU*	*trnP-UGG*	*trnP-GGG*	*trnQ-UUG*
	trnR-ACG	*trnR-UCU*[2]	*trnR-CCG*	*trnS-GCU*	*trnS-UGA*
	trnS-GGA	*trnT-UGU*	*trnT-GGU*	*trnV-GAC*	*trnV-UAC*[1]
	trnW-CCA	*trnY-GUA*[2]			
Ribosomal small subunit	*rps2*	*rps3*	*rps4*	*rps7*	*rps8*
	rps11	*rps12*[1]	*rps14*	*rps15*	*rps18*
	rps19				
Ribosomal large subunit	*rpl2*[1]	*rpl14*	*rpl16*[1]	*rpl20*	*rpl22*
	rpl23	*rpl32*	*rpl33*	*rpl36*	
DNA-dependent RNA polymerase	*rpoA*	*rpoB*	*rpoC1*[1]	*rpoC2*	
Translational initiation factor	*infA*				
Photosystem I	*psaA*	*psaB*	*psaC*	*psaI*	*psaJ*
	psaM	*ycf3*[1]	*ycf4*		

Photosystem II	*psbA*	*psbB*	*psbC*	*psbD*	*psbE*
	psbF	*psbH*	*psbI*	*psbJ*	*psbK*
	psbL	*psbM*	*psbN*	*psbT*	*psbZ*
Subunits of cytochrome	*petA*	*petB*[1]	*petD*[1]	*petL*	*petN*
	petG				
ATP synthase	*atpA*	*atpB*	*atpE*	*atpF*[1]	*atpH*
	atpI				
Rubisco	*rbcl*				
Chlorophyll biosynthesis	*chlB*	*chlL*	*chlN*		
NADH oxidoreductase	*ndhA*[1]	*ndhB*[1]	*ndhC*	*ndhD*	*ndhE*
	ndhF	*ndhG*	*ndhH*	*ndhI*	*ndhJ*
	ndhK				
Maturase	*matK*				
Envelope membrane protein	*cemA*				
Subunit of acetyl-CoA	*accD*				
C-type cytochrome synthesis gene	*ccsA*				
Protease	*clpP*				
Component of TIC complex	*ycf1*				
Unknown function	*ycf2*				
Pseudogene	*ycf68*				

[1] Genes containing introns. [2] Double-copy gene. [3] Triple-copy gene.

A. angustifolia, *A. araucana*, and *A. bidwillii*, plastomes were identical in gene content and order. The *rps16* gene was absent in all three plastomes, but a remanescent of about ~20aa from *rps16* could be identified between *trnK* and *chlB* genes. The *Araucaria* plastomes lost one copy of the IR region, but present a remaining short IR of about 580 bp containing the *rrn5* gene.

The plastome gene content within Araucariaceae family is also strictly the same. The differences between Araucariaceae plastomes concern the presence of the duplicated *trnY-GUA* and the triple-copy

trnD-GUC in *Araucaria*. These tRNAs are present as single-copy in both *W. nobilis* and *A. dammara*. The *rps16* gene is lacking in all Araucariaceae plastomes, including species of genus *Araucaria*, corroborating the hypothesis that families Podocarpaceae and Araucariaceae lost this gene during evolution, while other non-Pinaceae families did not (VIEIRA et al., 2014a, 2016b; WU; CHAW, 2014a; YI et al., 2013). However, remnants of *rps16* were identified in the *trnK*/*chlB* region of *Araucaria* genomes. The presence of this remnants were previously reported in *W. nobilis* and *A. dammara* plastomes (YAP et al., 2015). In contrast, in Podocarpaceae plastomes, the *rps16* gene is completely absent (VIEIRA et al., 2014a). YAP et al., (2015) hypothesized that there is a possible slower mutation rate in this plastome region of Araucariaceae compared to Podocarpaceae, which could explain the complete absence of *rps16* in the latter group. This gene has been shown to be essential for cell survival (FLEISCHMANN et al., 2011), but it is also absent or nonfunctional in many species of some families, such as Pinaceae (TSUDZUKI et al., 1992), Fabaceae (GUO et al., 2007; TANGPHATSORNRUANG et al., 2010), Dioscoreaceae (HANSEN et al., 2007), and Melanthiaceae (DO; KIM; KIM, 2014).

Plastomes are usually structurally divided into one large single copy region (LSC), one small single copy region (SSC), and two inverted repeat regions (IR). However, the loss of one IR and the presence of a short IR (150 - 500 bp) has been reported in several species, mainly in conifers (HIRAO et al., 2008; VIEIRA et al., 2016b; WU; CHAW, 2014a; YI et al., 2013). The Araucariaceae genomes have a short IR of about 580 bp containing the *rrn5* gene. This region is probably residual of the typical rRNA-containing IRs. The presence of the duplicated *trnI-CAU* gene in *Araucaria* genus corroborates the currently available plastomes sequence of conifers clade I and some conifers clade II which also present *trnI-CAU* in palindromic repeat (VIEIRA et al., 2016b). However, in contrast to *R. piresii* duplicated *trnN-GUU* gene, *trnI-CAU* was not reported to show ability to recombine (LIN et al., 2010; VIEIRA et al., 2016b; WU et al., 2011b), and, no evidence of this recombination ability was observed in the *Araucaria* spp. plastome assembly.

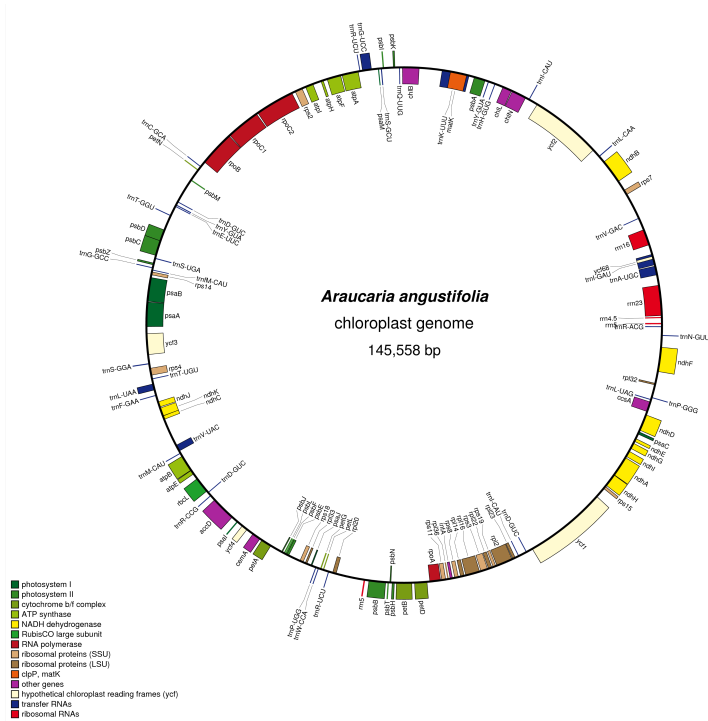

Figure 1. Gene map of *Araucaria angustifolia* plastid genome (plastome). Genes drawn inside the circle are transcribed clockwise, and genes drawn outside are counterclockwise. Genes belonging to different functional groups are color-coded. *Araucaria araucana* and *Araucaria bidwillii* are equal to *A. angustifolia* in gene order and content, so their gene maps are identical.

Although plastomes are considered more or less conserved in gene content and structure, the presence of rearrangements and inversions are common between different families. In addition, many conifers lost one copy of the IRs, including Araucariaceae and Podocarpaceae. The complete plastome alignment between Araucariaceae and Podocarpaceae species reveled rearrangements and inversions between these two families (Fig. 2). However, between the species of Araucariaceae, no rearrangements could be identified. The differences between Araucariaceae and Podocarpaceae plastomes are mainly due to the inversion of the *petL-rps18* and *rpl23-psbB* regions, and the translocation of *clpP* gene (Fig. 2). The region starting after *clpP* gene in Podocarpaceae and after *psbB* gene in Araucariaceae shows many Locally Collinear Blocks (LCBs). However, Araucariaceae and

Podocarpaceae gene content in this region were manually verified and we observed that they are strictly the same. Thus, we ascribed the presence of these LCBs to possible variations in the intergenic regions.

Phylogenetic Inference

A high number of complete plastome sequences are successfully being used for plastid phylogenomic analyses (BARRETT et al., 2016; VIEIRA et al., 2016a). This strategy proved to be effective to resolve evolutionary relationships, especially at lower taxonomic levels, which show limited sequence variation (ROGALSKI et al., 2015).

Araucariaceae is currently divided into three genera: *Araucaria*, *Wollemia*, and *Agathis*. Thus, we compared the phylogenetic trees using either *A. dammara* or *W. nobilis* as outgroups. The complete plastome aligned matrix with the 18 species (17 *Araucaria* species + one of the outgroup options) were 152,841 (140,932 without gaps) using *A. dammara* and 154,534 (140,711 without gaps) nucleotide positions in length using *W. nobilis*. Even though the alignment using *A. dammara* as outgroup generated a larger matrix, the log-likelihoods (lnL) in both BI and ML analyses generated lower lnL values, ML: lnL = -232,172 and BI lnL = -232,187. Using *W. nobilis* as outgroup, we obtained ML: lnL = -230,356 and BI: lnL = -230,370. Thus, we decided to present here the phylogenetic tree obtained with *W. nobilis* as outgroup.

Both ML and BI analyses produced phylogenomic trees identical in topology, but with minor differences regarding the levels of resolution (Fig. 3). Both trees presented high overall resolution, with some bootstrap values smaller than 70% only within section *Eutacta*. *Araucaria* + *Bunya*, and Eutacta formed monophyletic clades with maximum support (100% ML bootstrap and PP=1.0; Fig. 3). In addition, section *Eutacta* was monophyletic and *A. heterophylla* presented a sister relationship to the two well supported lineages consisting of the coastal (*A. columnaris*, *A. luxurians*, and *A. nemorosa*) + small-leaved clade (*A. schmidii*, *A. scopulorum*, *A. bernieri*, and *A. subulata*), and the large leaved clade (*A. humboldensis*, *A. biramulata*, *A. muelleri*, *A. rulei*, *A. laubenfelsii*, *A. montana*).

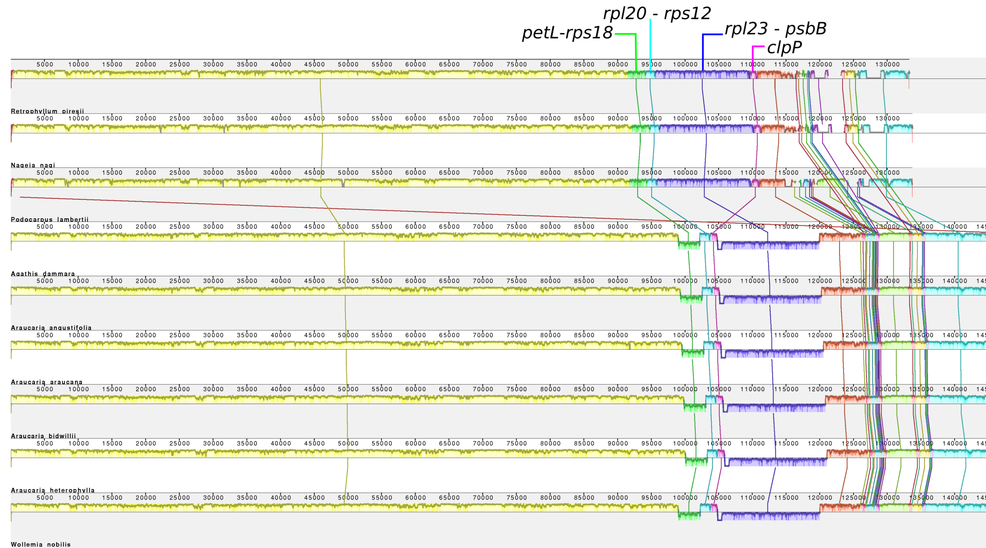

Figure 2. Multiple plastome alignment with species from families Araucariaceae and Podocarpaceae. Different colours represent different Local Colinear Blocks (LCBs), indicating sequence rearrangements and/or translocation. Gene region of LCBs with green, light blue, dark blue, and pink colour are identified in the figure.

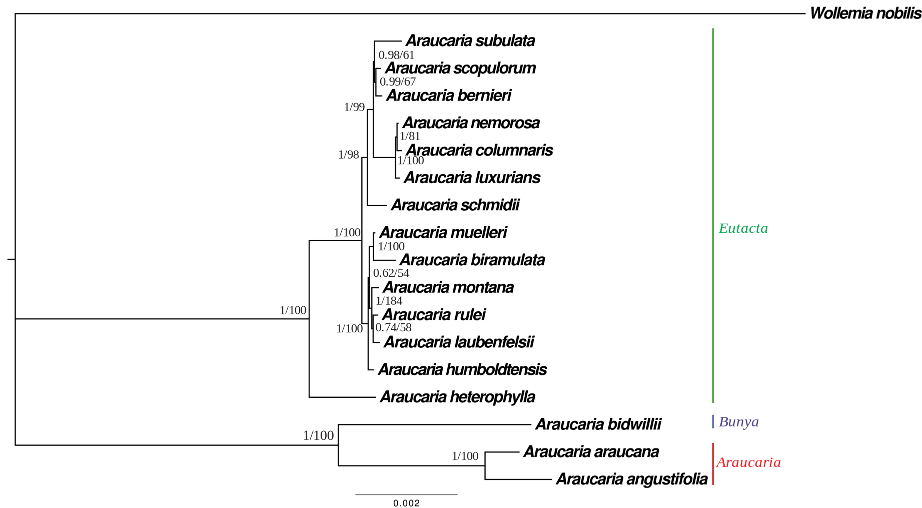

Figure 3. Phylogenetic relationship (Bayesian phylogeny) within *Araucaria* genus based on complete plastomes of 17 *Araucaria* species and the sister genus *Wollemia* as outgroup. Numbers at the branches indicate Bayesian posterior probabilities/maximum likelihood bootstrap (1000 replicates) values. Branch length is proportional to the inferred divergence level. Scale bar = substitutions per site.

In Araucariaceae, previous phylogenetic analysis based on *rbcL* gene sequences (SETOGUCHI et al., 1998) and AFLP markers (STEFENON; GAILING; FINKELDEY, 2006) also showed the monophyletism on clades *Araucaria* + Bunya, and *Eutacta*. However, our analysis presents higher support, probably due to the plastid phylogenomic strategy of analysis, and, consequently, with more informative sites.

The sister relationship observed between the small-leaved clades (coastal and interior), and the large-leaved clade was expected. These results corroborate the plastid phylogenomic inference of *Eutacta* species (RUHSAM et al., 2015), as well as the integrated phylogenetic analysis based on morphological and molecular data (ESCAPA; CATALANO, 2013). In addition, the node positioning *A. heterophylla* as sister to the other species of section *Eutacta* was expected due to the geographical isolation of this species. *A. heterophylla* is isolated in the Norfolk Island, while the others species of section *Eutacta* are from New Caledonian.

The sister relationship between sections *Bunya* and *Araucaria* were confirmed here with highly supported nodes. These sister relationship was previously demonstrated in many phylogenetic analysis (ESCAPA; CATALANO, 2013; SETOGUCHI et al., 1998; STEFENON; GAILING; FINKELDEY, 2006), but it is the first time that this inference is performed with plastid phylogenomic inference. Thus, bringing even more support for the phylogenetic relationships within genus *Araucaria*.

AVAILABILITY OF SUPPORTING DATA

Plastome nucleotide sequence were deposited in Genbank repository. The accession numbers are available in Table 1.

CONSIDERAÇÕES FINAIS E PERSPECTIVAS

Os resultados encontrados a partir do sequenciamento completo do genoma plastidial da espécie *R. piresii* possibilitou um maior conhecimento sobre uma conífera amazônica até então pouco estudada. Até onde se sabe, essa espécie é endêmica do Parque Nacional dos Pacaás Novos, Rondônia, Brasil, não existindo relato de coleta/identificação dessa espécie em outros locais. Neste estudo, foi possível identificar 168 regiões SSR, sendo que desses, 88 eram homopolímeros constituídos de A/T. Estes resultados revelam a presença de vários sítios SSR no genoma plastidial de *R. piresii* que podem ser futuramente avaliados quanto ao seu polimorfismo, possibilitando estudos de estrutura de populações nessa interessante conífera amazônica que tão pouco se conhece. Além disso, a realização de estudos com essa espécie pode ajudar a descrever seu estado de conservação em sua área de ocorrência.

O sequenciamento completo do genoma plastidial da espécie *R. piresi*i revelou a presença de recombinações no genoma plastidial genoma mediadas pela presença de repetições diretas e palíndromas. Esses sítios recombinantes ativos resultam na presença de uma inversão no genoma plastidial (mediada pela repetição palíndroma) e uma deleção (mediada pela repetição direta). Esse foi o primeiro relato de uma deleção mediada por uma repetição direta ocorrendo de forma natural em plantas. Até o momento, apenas relatos em plantas transplastômicas haviam sido realizados. Ressalta-se que o DNA de apenas uma planta foi utilizado, confirmando que estas isoformas coexistem dentro de uma única planta. Além disso, não foram observadas acículas anormais ou variegadas na planta utilizada para o sequenciamento. Esses resultados levantam várias questões interessantes, como: trata-se de uma peculiaridade do plastoma de *R. piresii* ou este é um fenômeno mais comum presente em outros plastomas que tem sido ignorado antes? Existe alguma vantagem evolutiva desta recombinação? Considerando que os plastomas têm um nível de ploidia elevado, existe uma mistura de isoformas viáveis e inviáveis que são suficientes para garantir a expressão gênica? Aprofundar estas questões e outras podem ajudar a desvendar aspectos importantes da evolução adaptativa das coníferas.

O sequenciamento completo do genoma plastidial das espécies *A. angustifolia*, *A. araucana* e *A. bidwillii* revelou que esses genomas plastidiais são idênticos em conteúdo e ordem dos genes, incluindo a

perda do gene *rps16* e a presença de um remanescente da IR contendo o gene *rrn5*. As análises filogenômicas realizadas com as três espécies de *Araucaria* sequenciadas nesse trabalho mais as 14 espécies de *Araucaria* da secção *Eutacta* disponíveis no Genbank, tanto por Máxima Verossimilhança quanto por Inferência Bayesiana produziram árvores filogenômicas idênticas em topologia. As duas seções de *Araucaria* (*Araucaria* e *Eutacta*) se mostraram monofiléticas com alto suporte. A seção *Bunya* é monotípica e formou uma relação irmã altamente suportada com a secção *Araucaria*. Ressalta-se que a secção *Intermedia*, que normalmente apresenta uma relação irmã com a secção *Bunya* não foi amostrada nesse trabalho. Essas análises também apoiaram a secção *Eutacta* como monofilética.

O sequenciamento dessas espécies de *Araucaria* possibilitaram a realização de análises filogenômicas dentro do gênero *Araucaria* com o maior suporte até o momento relatada. Esses dados ainda possibilitarão um trabalho futuro com a identificação das regiões com maior potencial para uso como marcadores moleculares plastidiais para o gênero *Araucaria*. Historicamente, o Laboratório de Fisiologia do Desenvolvimento e Genética Vegetal (LFDGV) realiza análises de genética de populações em *A. angustifolia*. Dessa forma, os dados gerados possibilitarão desenvolver futuros trabalhos com uma maior resolução, ampliando o conhecimento a respeito dessa espécie.

O desenvolvimento das atividades durante o ano de 2015 na Universidade Federal do Paraná intensificou a relação técnico-científica entre o Núcleo de Fixação de Nitrogênio e o nosso laboratório (LFDGV). Os trabalhos desenvolvidos com genômica plastidial também estabeleceram uma parceria com o Laboratório de Fisiologia Molecular de Plantas da Universidade Federal de Viçosa, através da co-orientação do Prof. Dr. Marcelo Rogalski. Através dessa parceria, intercâmbios relevantes puderam ser estabelecidos entre o LFDGV e o referido laboratório, gerando trabalhos científicos que constituem e constituirão teses e dissertações no Programa de Pós-graduação em Fisiologia Vegetal da UFV.

Essas parcerias com a UFV e a UFPR também geraram outros trabalhos que foram realizados ao longo do período de desenvolvimento da presente tese. Dentre eles, uma revisão bibliográfica publicada no periódico “Frontiers in Plant Science” intitulada “Plastid genomics in horticultural species: importance and applications for plant population genetics, evolution, and biotechnology” e o artigo abordando genômica plastidial de bambus nativos do Brasil entitulado “Phylogenetic

inference and SSR characterization of tropical woody bamboos tribe Bambuseae (Poaceae: Bambusoideae) based on complete plastid genome sequences” publicado no periódico “Current Genetics”. Esse último encontra-se em anexo a essa tese (Anexo A).

Sendo assim, acredita-se que o desenvolvimento de todos estes estudos acima mencionados poderão contribuir substancialmente para o avanço científico na área, além de permitir a consolidação desta linha de pesquisa no âmbito do LFDGV e do PPG-RGV.

REFERÊNCIAS

AHLERT, D. et al. Insensitivity of chloroplast gene expression to DNA methylation. **Molecular Genetics and Genomics**, v. 282, n. 1, p. 17–24, 2009.

AHMED, I. et al. Identification of chloroplast genome loci suitable for high-resolution phylogeographic studies of *Colocasia esculenta* (L.) Schott (Araceae) and closely related taxa. **Molecular Ecology Resources**, v. 13, n. 5, p. 929–937, 2013.

ALKATIB, S. et al. The Contributions of Wobbling and Superwobbling to the Reading of the Genetic Code. **PLoS Genetics**, v. 8, n. 11, 2012.

BARRETT, C. F. et al. Plastid genomes reveal support for deep phylogenetic relationships and extensive rate variation among palms and other commelinid monocots. **New Phytologist**, v. 209, n. 2, p. 855–870, 2016.

BASKAUF, C. J. et al. Population genetics of braun's rockcress (*Boechera perstellata*, brassicaceae), an endangered plant with a disjunct distribution. **Journal of Heredity**, v. 105, n. 2, p. 265–275, 2014.

BESNARD, G. et al. Genomic profiling of plastid DNA variation in the Mediterranean olive tree. **Bmc Plant Biology**, v. 11, n. May, p. 80, 2011.

BIFFIN, E.; CONRAN, J. G.; LOWE, A. J. Podocarp Evolution: A Molecular Phylogenetic Perspective. **Ecology of the Podocarpaceae in tropical forests**, v. 95, p. 1–20, 2011.

BITTENCOURT, J. V. M.; SEBBENN, A M. Patterns of pollen and seed dispersal in a small, fragmented population of the wind-pollinated tree *Araucaria angustifolia* in southern Brazil. **Heredity**, v. 99, n. 6, p. 580–591, 2007.

BOCK, R. Structure, function, and inheritance of plastid genomes. **Topics in Current Genetics**, v. 19, p. 29–63, 2007.

BOCK, R. The give-and-take of DNA: horizontal gene transfer in plants. **Trends in Plant Science**, v. 15, n. 1, p. 11–22, 2010.

BRODRIBB, T. J.; HILL, R. S. Southern Conifers in Time and Space. **Australian Journal of Botany**, v. 47, n. 5, p. 639–696, 1999.

CLEGG, M. T.; LEARN, G. H.; MORTON, B. R. patterns of chloroplast DNA evolution. v. 91, n. July, p. 6795–6801, 1994.

CRONN, R. et al. Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. **Nucleic Acids Research**, v. 36, n. 19, 2008.

DAY, A. et al. Simple and Efficient Removal of Marker Genes From Plastids by Homologous Recombination. In: PEÑA, L. (Ed.). . **Transgenic Plants: Methods and Protocols**. 1. ed. New York: Humana Press, 2004. p. 255–269.

DELPLANCKE, M. et al. Gene flow among wild and domesticated almond species: Insights from chloroplast and nuclear markers. **Evolutionary Applications**, v. 5, n. 4, p. 317–329, 2012.

DETTMANN, M.; CLIFFORD, T. H. Biogeography of Araucariaceae. **Australia and New Zealand forest histories: araucarian forests. Australian Forest History Society Inc. Occasional Publication**, v. 2, p. 1–9, 2005.

DEXTER, K. G.; TERBORGH, J. W.; CUNNINGHAM, C. W. Historical effects on beta diversity and community assembly in Amazonian trees. **Proceedings of the National Academy of Sciences**, v. 109, n. 20, p. 7787–7792, 2012.

DO, H. D. K.; KIM, J. S.; KIM, J. H. A *trnI_CAU* triplication event in the complete chloroplast genome of *Paris verticillata* M.Bieb. (Melanthiaceae, Liliales). **Genome biology and evolution**, v. 6, n. 7, p. 1699–1706, 2014.

DOS REIS, M. S. et al. Distribuição da Diversidade Genética e Conservação de Espécies Arbóreas em Remanescentes Florestais de

Santa Catarina. In: VIBRANS, A. C. et al. (Eds.). **Inventário florístico Florestal de Santa Catarina - Diversidade e conservação dos remanescentes florestais**. 1. ed. Blumenau: Edifurb, 2012. p. 344.

ESCAPA, I. H.; CATALANO, S. A. Phylogenetic Analysis of Araucariaceae: Integrating Molecules, Morphology, and Fossils. **International Journal of Plant Sciences**, v. 174, n. 8, p. 1153–1170, 2013.

FARJON, A. **World Checklist and Bibliography of Conifers**. 2. ed. Royal Botanic Garden, 1998.

FARJON, A.; FILER, D. **An atlas of the world's conifers: an analysis of their distribution, biogeography, diversity and conservation status**. 1. ed. Leiden: Koninklijke Brill NV, 2013.

FERREIRA, D. K. et al. Genetic analysis of 50-year old Brazilian pine (*Araucaria angustifolia*) plantations: Implications for conservation planning. **Conservation Genetics**, v. 13, n. 2, p. 435–442, 2012.

FLEISCHMANN, T. T. et al. Nonessential Plastid-Encoded Ribosomal Proteins in Tobacco: A Developmental Role for Plastid Translation and Implications for Reductive Genome Evolution. **the Plant Cell Online**, v. 23, n. 9, p. 3137–3155, 2011.

GIELLY, L.; TABERLET, P. The use of chloroplast DNA to resolve plant phylogenies: noncoding versus *rbcL* sequences. **Molecular biology and evolution**, v. 11, n. 5, p. 769–777, 1994.

GRAY, B. N.; AHNER, B. A.; HANSON, M. R. Extensive homologous recombination between introduced and native regulatory plastid DNA elements in transplastomic plants. **Transgenic Research**, v. 18, n. 4, p. 559–572, 2009.

GUERRA, M. P. et al. Exploração, manejo e conservação da Araucária (*Araucaria angustifolia*). In: SIMÕES, L. L.; LINO, C. F. (Eds.). . **Sustentável Mata Atlantica: a exploração de seus recursos florestais**. São Paulo: SENAC São Paulo, 2002. p. 85–101.

GUO, W. et al. Predominant and substoichiometric isomers of the plastid genome coexist within Juniperus plants and have shifted multiple times during cupressophyte evolution. **Genome Biology and Evolution**, v. 6, n. 3, p. 580–590, 2014.

GUO, X. et al. Rapid evolutionary change of common bean (*Phaseolus vulgaris* L) plastome, and the genomic diversification of legume chloroplasts. **BMC genomics**, v. 8, p. 228, 2007.

GURDON, C.; MALIGA, P. Two distinct plastid genome configurations and unprecedented intraspecies length variation in the *accD* coding region in *Medicago truncatula*. **DNA Research**, v. 21, n. 4, p. 417–427, 2014.

HANSEN, D. R. et al. Phylogenetic and evolutionary implications of complete chloroplast genome sequences of four early-diverging angiosperms: *Buxus* (Buxaceae), *Chloranthus* (Chloranthaceae), *Dioscorea* (Dioscoreaceae), and *Illicium* (Schisandraceae). **Molecular Phylogenetics and Evolution**, v. 45, n. 2, p. 547–563, 2007.

HILL, R. S. Conifer origin, evolution and diversification in the Southern Hemisphere. **Ecology of the southern conifers**, p. 10–29, 1995.

HILL, R. S.; POLE, M.S. Leaf and shoot morphology of extant *Afrocarpus*, *Nageia* and *Retrophyllum* (Podocarpaceae) species, and species with similar leaf arrangement. **Australian Systematic Botany**, v. 5, n. June, p. 337–358, 1992.

HIRAO, T. et al. Complete nucleotide sequence of the *Cryptomeria japonica* D. Don. chloroplast genome and comparative chloroplast genomics: diversified genomic structure of coniferous species. **BMC plant biology**, v. 8, p. 70, 2008.

HUANG, Y. Y.; MATZKE, A. J. M.; MATZKE, M. Complete Sequence and Comparative Analysis of the Chloroplast Genome of Coconut Palm (*Cocos nucifera*). **PLoS ONE**, v. 8, n. 8, p. 1–12, 2013.

IAMTHAM, S.; DAY, A. Removal of antibiotic resistance genes from transgenic tobacco plastids. **Nature biotechnology**, v. 18, n. 11, p. 1172–1176, 2000.

IUCN. **IUCN Red List of Threatened Species**.

JANSEN, R. K. et al. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. **Proceedings of the National Academy of Sciences of the United States of America**, v. 104, n. 49, p. 19369–19374, 2007.

JANSEN, R. K.; RUHLMAN, T. A. Plastid Genomes of Seed Plants. In: BOCK, R.; KNOOP, V. (Eds.). . **Genomics of Chloroplasts and Mitochondria**. Dordrecht: Springer Netherlands, 2012. p. 103–126.

JARDIM BOTÂNICO DO RIO DE JANEIRO. **Gimnospermas in Flora do Brasil 2020 em construção**. Disponível em: <http://floradobrasil.jbrj.gov.br/reflora/floradobrasil/FB128477>. Acesso em: 15 dez. 2016.

KATO, S. et al. Genetic structure of island populations of *Prunus lannesiana* var. speciosa revealed by chloroplast DNA, AFLP and nuclear SSR loci analyses. **Journal of Plant Research**, v. 124, n. 1, p. 11–23, 2011.

KATO, S. et al. Population genetic structure in a threatened tree, *Pyrus calleryana* var. dimorphophylla revealed by chloroplast DNA and nuclear SSR locus polymorphisms. **Conservation Genetics**, v. 14, n. 5, p. 983–996, 2013.

KATOH, K.; STANDLEY, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. **Molecular Biology and Evolution**, v. 30, n. 4, p. 772–780, 2013.

KELCH, D. G. The Phylogeny of the Podocarpaceae Based on Morphological Evidence. **Systematic Botany**, v. 22, n. 1, p. 113–131, 1997.

KELCH, D. G. Phylogeny of Podocarpaceae: Comparison of evidence from morphology and *18S* rDNA. **American Journal of Botany**, v. 85, n. 7, p. 986–996, 1998.

KERSHAW, P.; WAGSTAFF, B. THE SOUTHERN CONIFER FAMILY ARAUCARIACEAE: History, Status, and Value for Paleoenvironmental Reconstruction. **Annual Review of Ecology and Systematics**, v. 32, p. 397–414, 2001.

KNOX, E. B. The dynamic history of plastid genomes in the Campanulaceae sensu lato is unique among angiosperms. **Proceedings of the National Academy of Sciences of the United States of America**, v. 111, n. 30, p. 11097–102, 2014.

KODE, V. et al. 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 Journal**, v. 46, n. 5, p. 901–909, 2006.

KUNZMANN, L. Araucariaceae (Pinopsida): Aspects in palaeobiogeography and palaeobiodiversity in the Mesozoic. **Zoologischer Anzeiger**, v. 246, n. 4, p. 257–277, 2007.

KURTZ, S. et al. REPuter: the manifold applications of repeat analysis on a genomic scale. **Nucleic Acids Research**, v. 29, n. 22, p. 4633–4642, 2001.

KURTZ, S. et al. Versatile and open software for comparing large genomes. **Genome biology**, v. 5, n. 2, p. R12, 2004.

LEISTER, D. Towards understanding the evolution and functional diversification of DNA-containing plant organelles. **F1000Research**, v. 5, n. 0, p. 330, 2016.

LI, F.-W. et al. Genes Translocated into the Plastid Inverted Repeat Show Decelerated Substitution Rates and Elevated GC Content. **Genome biology and evolution**, v. 8, n. 8, p. 2452–2458, 2016a.

LI, J. et al. Evolution of short inverted repeat in cupressophytes, transfer of *accD* to nucleus in *Sciadopitys verticillata* and phylogenetic position of Sciadopityaceae. **Scientific reports**, v. 6, n. October 2015, p. 20934, 2016b.

LIN, C. P. et al. Comparative chloroplast genomics reveals the evolution of Pinaceae genera and subfamilies. **Genome Biology and Evolution**, v. 2, n. 1, p. 504–517, 2010.

LIN, C. P. et al. The complete chloroplast genome of *Ginkgo biloba* reveals the mechanism of inverted repeat contraction. **Genome Biology and Evolution**, v. 4, n. 3, p. 374–381, 2012.

LISBOA, P. L. B.; ALMEIDA, S. S. Vida e obra de João Murça Pires (1917-1994). A**cta Botânica Brasílica**, v. 9, n. 2, p. 303–314, 1995.

LITTLE, D. P.; KNOPF, P.; SCHULZ, C. DNA Barcode Identification of Podocarpaceae—The Second Largest Conifer Family. v. 8, n. 11, 2013.

LOHSE, M. et al. OrganellarGenomeDRAW--a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. **Nucleic acids research**, v. 41, n. Web Server issue, p. 1–7, 2013.

LÓPEZ, P. et al. Progenitorderivative speciation in Pozoa (Apiaceae, Azorelloideae) of the Southern Andes. **Annals of Botany**, v. 109, n. 2, p. 351–363, 2012.

MAIER, R. M. et al. Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. **Journal of Molecular Biology**, v. 251, n. 5, p. 614–628, 1995.

MANTOVANI, A.; MORELLATO, L. P. C.; DOS REIS, M. S. Internal genetic structure and outcrossing rate in a natural population of *Araucaria angustifolia* (Bert.) O. Kuntze. **Journal of Heredity**, v. 97, n. 5, p. 466–472, 2006.

MILL, R. R. A monographic revision of *Retrophyllum* (Podocarpaceae). **Edinburgh Journal of Botany**, v. 73, n. 2, p. 171–261, 2016.

MINISTÉRIO DO MEIO AMBIENTE. **Brasil. Instruçao normativa nº6, de setembro de 2008**, 2008.

MOORE, M. J. et al. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. **Proceedings of the National Academy of Sciences of the United States of America**, v. 104, n. 49, p. 19363–8, 2007.

MOORE, M. J. et al. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. **Proceedings of the National Academy of Sciences of the United States of America**, v. 107, n. 10, p. 4623–8, 2010.

MURRAY, B. G. Karyotype Variation and Evolution in Gymnosperms. In: GREILHUBER, J.; DOLEZEL, J.; WENDEL, J. F. (Eds.). . **Plant Genome Diversity Volume 2**. Vienna: Springer Vienna, 2013. p. 231–243.

PAGE, C. N. New and maintained genera in the conifer families Podocarpaceae and Pinaceae. **Notes of the Royal Botanical Garden Edinburgh**, v. 45, n. 2, p. 377–395, 1989.

PALMER, J. D. Chloroplast DNA exists in two orientations. **Nature**, v. 301, n. 5895, p. 92–93, 1983.

PALMER, J. D. Chloroplast DNA and Molecular Phylogeny. **BioEssays**, v. 2, n. 6, p. 263–267, 1985.

PALMER, J. D. et al. Chloroplast DNA Variation and Plant Phylogeny. **Annals of the Missouri Botanical Garden**, v. 75, n. 4, p. 1180–1206, 1988.

PALUDO, G. F. et al. Inferring population trends of *Araucaria angustifolia* (Araucariaceae) using a transition matrix model in an old-growth forest. **Southern Forests: a Journal of Forest Science**, v. 2620, n. March, 2016.

PALUDO, G. F.; MANTOVANI, A.; REIS, M. S. DOS. Regeneração de uma população natural de *Araucaria angustifolia* (Araucariaceae). **Revista Árvore**, v. 35, n. 5, p. 1107–1119, 2011.

PARKS, M.; CRONN, R.; LISTON, A. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. **BMC biology**, v. 7, n. 1, p. 84, 2009.

PIRO, V. C. et al. FGAP: an automated gap closing tool. **BMC research notes**, v. 7, n. 1, p. 371, 2014.

POWELL, W. et al. Hypervariable microsatellites provide a general source of polymorphic DNA markers for the chloroplast genome. **Current Biology**, v. 5, n. 9, p. 1023–1029, 1995.

PROVAN, J.; POWELL, W.; HOLLINGSWORTH, P. M. Chloroplast microsatellites: new tools for studies in plant ecology and evolution. **Trends in ecology & evolution**, v. 16, n. 3, p. 142–147, 2001.

QUINN, C. J.; PRICE, R. A. Phylogeny of the southern hemisphere conifers. **Acta Horticulturae**, v. 615, p. 129–136, 2003.

REBOUD, X.; ZEYL, C. Organelle inheritance in plants. **Heredity**, v. 72, n. 2, p. 132–140, 1994.

REIS, M. S.; LADIO, A.; PERONI, N. Landscapes with *Araucaria* in South America: Evidence for a cultural dimension. **Ecology and Society**, v. 19, n. 2, p. 43, 2014.

RIBEIRO, M. C. et al. The Brazilian Atlantic Forest: How much is left, and how is the remaining forest distributed? Implications for conservation. **Biological Conservation**, v. 142, n. 6, p. 1141–1153, 2009.

ROGALSKI, M. et al. Rpl33, a Nonessential Plastid-Encoded Ribosomal Protein in Tobacco, Is Required under Cold Stress Conditions. **the Plant Cell Online**, v. 20, n. 8, p. 2221–2237, 2008.

ROGALSKI, M. et al. Plastid genomics in horticultural species: importance and applications for plant population genetics, evolution, and biotechnology. **Frontiers in plant science**, v. 6, n. July, p. 586, 2015.

ROGALSKI, M.; KARCHER, D.; BOCK, R. Superwobbling facilitates translation with reduced tRNA sets. **Nature Structural & Molecular Biology**, v. 15, n. 2, p. 192–198, 2008.

ROGALSKI, M.; RUF, S.; BOCK, R. Tobacco plastid ribosomal protein S18 is essential for cell survival. **Nucleic Acids Research**, v. 34, n. 16, p. 4537–4545, 2006.

RONQUIST, F.; HUELSENBECK, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. **Bioinformatics**, v. 19, n. 12, p. 1572–1574, 2003.

ROULLIER, C. et al. Combining chloroplast and nuclear microsatellites to investigate origin and dispersal of New World sweet potato landraces. **Molecular Ecology**, v. 20, n. 19, p. 3963–3977, 2011.

ROULLIER, C. et al. Disentangling the Origins of Cultivated Sweet Potato (*Ipomoea batatas* (L.) Lam.). **PLoS ONE**, v. 8, n. 5, 2013.

RUHFEL, B. R. et al. From algae to angiosperms - inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. **BMC Evolutionary Biology**, v. 14, p. 23, 2014.

RUHSAM, M. et al. Does complete plastid genome sequencing improve species discrimination and phylogenetic resolution in *Araucaria*? **Molecular Ecology Resources**, v. 15, n. 5, p. 1067–1078, 2015.

SANT’ANNA, C. S. et al. Realized pollen and seed dispersal within a continuous population of the dioecious coniferous Brazilian pine [*Araucaria angustifolia* (Bertol.) Kuntze]. **Conservation Genetics**, v. 14, n. 3, p. 601–613, 2013.

SCHATTNER, P.; BROOKS, A. N.; LOWE, T. M. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and

snoRNAs. **Nucleic Acids Research**, v. 33, n. SUPPL. 2, p. 686–689, 2005.

SCHLOGL, P. S.; DE SOUZA, A. P.; NODARI, R. O. PCR-RFLP analysis of non-coding regions of cpDNA in *Araucaria angustifolia* (Bert.) O. Kuntze. **Genetics and Molecular Biology**, v. 30, n. 2, p. 423–427, 2007.

SETOGUCHI, H. et al. Phylogenetic relationships within Araucariaceae based on *rbcl* gene sequences. **American Journal of Botany**, v. 85, n. 11, p. 1507–1516, 1998.

SHAW, J. et al. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: The Tortoise and the hare III. **American Journal of Botany**, v. 94, n. 3, p. 275–288, 2007.

SHAW, J. et al. Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: The tortoise and the hare IV. **American Journal of Botany**, v. 101, n. 11, p. 1987–2004, 2014.

SHINOZAKI, K. et al. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. **The EMBO journal**, v. 5, n. 9, p. 2043–2049, 1986.

SOLL, J.; SCHLEIFF, E. Plant cell biology: Protein import into chloroplasts. **Nature Reviews Molecular Cell Biology**, v. 5, n. March, p. 198–208, 2004.

STAMATAKIS, A. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. **Bioinformatics**, v. 22, n. 21, p. 2688–2690, 2006.

STEFENON, V. M.; GAILING, O.; FINKELDEY, R. Phylogenetic relationship within genus *Araucaria* (Araucariaceae) assessed by means of AFLP fingerprints. **Silvae Genetica**, v. 55, n. 2, p. 45–52, 2006.

STEIN, D. B.; PALMER, J. D.; THOMPSON, W. F. Structural evolution and flip-flop recombination of chloroplast DNA in the fern genus Osmunda. **Current Genetics**, v. 10, n. 11, p. 835–841, 1986.

SUDIANTO, E. et al. Revisiting the Plastid Phylogenomics of Pinaceae with Two Complete Plastomes of *Pseudolarix* and *Tsuga*. **Genome biology and evolution**, v. 8, n. 6, p. 1804–11, 2016.

TANGPHATSORNRUANG, S. et al. The chloroplast genome sequence of mungbean (*Vigna radiata*) determined by high-throughput pyrosequencing: structural organization and phylogenetic relationships. **DNA Research**, v. 17, n. 1, p. 11–22, 2010.

TIMMIS, J. N. et al. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. **Nature reviews. Genetics**, v. 5, n. 2, p. 123–35, 2004.

TOMAR, R. S. S. et al. Development of chloroplast-specific microsatellite markers for molecular characterization of alloplasmic lines and phylogenetic analysis in wheat. **Plant Breeding**, v. 133, n. 1, p. 12–18, 2014.

TSUDZUKI, J. et al. 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*. **Molecular & general genetics : MGG**, v. 232, n. 2, p. 206–214, 1992.

VANYUSHIN, B. F. **DNA Methylation in plants**. [s.l: s.n.].

VENDRAMIN, G. G. et al. Distribution of genetic diversity in *Pinus pinaster* Ait. as revealed by chloroplast microsatellites. **Theoretical and Applied Genetics**, v. 97, n. 3, p. 456–463, 1998.

VIEIRA, L. D. N. et al. The complete chloroplast genome sequence of *Podocarpus lambertii*: Genome structure, evolutionary aspects, gene content and SSR detection. **PLoS ONE**, v. 9, n. 3, p. 1–10, 2014a.

VIEIRA, L. D. N. et al. An improved protocol for intact chloroplasts and cpDNA isolation in conifers. **PLoS ONE**, v. 9, n. 1, p. 1–8, 2014b.

VIEIRA, L. DO N. et al. Phylogenetic inference and SSR characterization of tropical woody bamboos tribe Bambuseae (Poaceae: Bambusoideae) based on complete plastid genome sequences. **Current Genetics**, v. 62, n. 2, p. 443–453, 2016a.

VIEIRA, L. N. et al. The plastome sequence of the endemic Amazonian conifer, *Retrophyllum piresii* (Silba) C.N.Page, reveals different recombination events and plastome isoforms. **Tree Genetics and Genomes**, v. 12, n. 1, p. 1–11, 2016b.

WAGSTAFF, S. J. Evolution and biogeography of the austral genus *Phyllocladus* (Podocarpaceae). **Journal of Biogeography**, v. 31, n. 10, p. 1569–1577, 2004.

WAKASUGI, T. et al. Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine Pinus thunbergii. **Proceedings of the National Academy of Sciences of the United States of America**, v. 91, n. 21, p. 9794–9798, 1994.

WICKE, S. et al. The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. **Plant Molecular Biology**, v. 76, n. 3–5, p. 273–297, 2011.

WILLIAMS, C. G. Conifer reproductive biology. **Conifer Reproductive Biology**, n. May 2010, p. 1–169, 2009.

WU, C.; CHAW, S. Highly rearranged and size-variable chloroplast genomes in conifers II clade (cupressophytes): evolution towards shorter intergenic spacers. p. 344–353, 2014a.

WU, C. S. et al. 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. **Molecular Biology and Evolution**, v. 24, n. 6, p. 1366–1379, 2007.

WU, C. S. et al. Evolution of reduced and compact chloroplast genomes (cpDNAs) in gnetophytes: Selection toward a lower-cost strategy. **Molecular Phylogenetics and Evolution**, v. 52, n. 1, p. 115–124, 2009.

WU, C. S. et al. 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 Biology and Evolution**, v. 3, n. 1, p. 1284–1295, 2011a.

WU, C. S. et al. Comparative chloroplast genomes of Pinaceae: Insights into the mechanism of diversified genomic organizations. **Genome Biology and Evolution**, v. 3, n. 1, p. 309–319, 2011b.

WU, C. S.; CHAW, S. M. Highly rearranged and size-variable chloroplast genomes in conifers II clade (cupressophytes): Evolution towards shorter intergenic spacers. **Plant Biotechnology Journal**, v. 12, n. 3, p. 344–353, 2014b.

WYMAN, S. K.; JANSEN, R. K.; BOORE, J. L. Automatic annotation of organellar genomes with DOGMA. **Bioinformatics**, v. 20, n. 17, p. 3252–3255, 2004.

YAP, J. Y. S. et al. Complete chloroplast genome of the wollemi pine (*Wollemia nobilis*): Structure and evolution. **PLoS ONE**, v. 10, n. 6, p. 1–15, 2015.

YI, X. et al. 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 Biology and Evolution**, v. 5, n. 4, p. 688–698, 2013.

ZHANG, Q.; SODMERGEN. Why does biparental plastid inheritance revive in angiosperms? **Journal of Plant Research**, v. 123, n. 2, p. 201–206, 2010.

ZHANG, Y. et al. The complete chloroplast genome sequence of *Taxus chinensis* var. mairei (Taxaceae): Loss of an inverted repeat region and comparative analysis with related species. **Gene**, v. 540, n. 2, p. 201–209, 2014.

ANEXO A

Phylogenetic inference and SSR characterization of tropical woody bamboos tribe Bambuseae (Poaceae: Bambusoideae) based on complete plastid genome sequences

Leila do Nascimento Vieira[1], Karina Goulart dos Anjos[1], Helisson Faoro[2,3], Hugo Pacheco de Freitas Fraga[1], Thiago Machado Greco[1], Fábio de Oliveira Pedrosa[3], Emanuel Maltempi de Souza[3], Marcelo Rogalski[4], Robson Francisco de Souza[5], Miguel Pedro Guerra[1*]

[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. [*]Corresponding author, e-mail: miguel.guerra@ufsc.br
[2] Instituto Carlos Chagas, Fundação Oswaldo Cruz, Rua Professor Algacyr Munhoz Mader, 3775, bloco C, Cidade Industrial, Curitiba, Brazil
[3] 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.
[4] Departamento de Biologia Vegetal, Universidade Federal de Viçosa, Avenida Peter Henry Rolfs s/n, ZC: 36570-000 Viçosa, Minas Gerais, Brazil.
[5] Departamento de Microbiologia, Universidade de São Paulo, Avenida Professor Lineu Prestes, 1374 - Ed. Biomédicas II - Sala 250 - 2o. Andar, ZC: 05508-900 São Paulo, São Paulo, Brazil

Acknowledgments
This work was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq Proc. 457726/2013-0). The authors thank to Instituto Çarakura for kindly providing plant material.

Abstract

Bamboos belong to the subfamily Bambusoideae, which are divided into Arundinarieae, Bambuseae, and Olyreae tribes. Brazil holds a very high bamboo diversity, highlighting the Bambuseae *Merostachys* and *Chusquea* as the most common genera, and *Guadua* with highest potential utilization. Plastome sequences is an efficient option for increasing phylogenetic resolution and evolutionary studies, as well as may greatly facilitate the use of plastid DNA markers in plant population genetic studies. Here, we sequenced and analyzed the complete plastome sequences of the Brazilian native species *Guadua chacoensis* and *Merostachys sp*. We also performed full plastome phylogeny and characterized the occurrence, type, and distribution of SRRs using 20 Bambuseae species. The determined plastome sequence of *Merostachy*s sp. and *G. chacoensis* is 136,334 bp and 135,403 bp in size, respectively, with an identical gene content and typical quadripartite structure consisting of a pair of IRs separated by the LSC and SSC regions. We identified 141.8 cpSSR in Bambuseae plastomes and an inferior value (38.15) for plastome coding sequences. Among them, mono- and dinucleotide repeats were the most common in both complete plastome and coding sequences. Penta- and hexanucleotide repeats occur only in complete plastome sequence. Maximum Likelihood and Bayesian Inference analyses produced phylogenomic trees identical in topology. These trees supported monophyly of Paleotropical and Neotropical Bamboos clades. The Neotropical bamboos segregated into three well-supported lineages, Chusqueinae, Guaduinae, and Arthrostylidiinae, with the last two forming a well supported sister relationship. Paleotropical bamboos segregated into two well-supported lineages, Hickeliinae and Bambusinae + Melocanninae.

Keywords: Woody bamboos, Plastome, Next-generation, Molecular phylogeny, Poaceae

Introduction

Bamboos belong to the subfamily Bambusoideae Luerss. (Poaceae), which are divided in three tribes: Arundinarieae Asch. & Graebn., that include the temperate woody bamboos, with 621 species; Bambuseae Kunth ex Dumort., the tropical woody bamboos, with 893 species; and Olyreae Kunth ex Spenn., the herbaceous bamboos, with 127 species (Soreng et al. 2015). They are indigenous of all continents, except Antarctica and Europe, and occupy a broad range of habitat types, especially forests, from temperate to tropical climate zones (Clark et al. 2015). Brazil holds a high diversity of bamboo species, of which we can highlight *Merostachys* Spreng. (43 native species; 41 endemic species), and *Chusquea* Kunth (45 native species; 41 endemic species) as the most common genera, and *Guadua* Kunth (18 native species; 5 endemic species) as the highest important for use in construction, furniture, and handcraft (Greco et al. 2015; List of Species of the Brazilian Flora 2015).

Merostachys s.l. is the largest genus of the subtribe Arthrostylidiinae Soderstr. and R.P. Ellis. (Lizarazu et al. 2011). This genus spreads from Mexico to Southern Brazil in forest and forest margin habitats, but it is particularly diverse in Brazil, where remains many undescribed species (Judziewicz et al. 1999; Lizarazu et al. 2011; Santos-Gonçalves et al. 2012; Viana et al. 2013; Viana and Filgueiras 2014; Greco et al. 2015). Its taxonomic identification has been demonstrated to be very complicated due to their long periods of vegetative development, ca. 25–30 yr (Lizarazu et al. 2011).

Guadua s.l. is one of the five genera placed in the subtribe Guaduinae Soderstr. and R.P. Ellis. This genus occurs throughout tropical America, from Mexico to Brazil and northern Argentina (Londoño and Peterson 1992). *Guadua chacoensis* (Rojas) Londoño & P.M. Peterson is a native species from Brazil, but also occurs in northern Argentina, southeastern Bolivia, and southern Paraguay, and is one of the three southeasternmost species of the genus (Londoño and Peterson 1992).

The Bamboo Phylogeny Group (BPG 2006, 2012) suggested that a taxonomic reorganization of bamboos by means of a reclassification strongly supported by estimate of phylogeny was necessary, which can be addressed with plastid genome (plastome) markers (Kelchner et al. 2013). The advancement of next-generation sequencing technologies has enabled the rapid acquisition of whole

plastome sequences at low cost when compared with traditional sequencing approaches. Thus, the increasing number of plastome sequences in the public databases is now an efficient option for increasing phylogenetic resolution and evolutionary studies between and within different plant groups, families, genus and species (Zhang et al. 2011; Wysocki et al. 2015; Rogalski et al. 2015).

Besides phylogenetic studies, the low evolving rate, and the nonrecombinant, uniparentally inherited nature of plastome may greatly facilitate the use of plastid DNA markers in plant population genetic studies (Powell et al. 1995; Provan et al. 2001; Rogalski et al. 2015). By comparing nuclear and plastid markers, inferences about the relative contributions of seed and pollen flow to the genetic structure of natural populations are possible (Provan et al. 2001; Delplancke et al. 2012; Roullier et al. 2011; Khadivi-Khub et al. 2013). Powell et al. (1995) reported the presence of simple sequence repeats in plastids (cpSSR), consisting of DNA sequences in tandemly repeated motifs of 1-6 base pairs (bp). This cpSSR became widely studied due to their ability to generate highly informative DNA markers. The development and application of these plastid molecular markers was demonstrated by several studies (see review by Rogalski et al. 2015).

Plastid genome sequences are also useful tools utilized to study plastid gene transfer to nucleus (Huang et al. 2003, 2005; Stegemann et al. 2003; Bock 2006; Stegemann and Bock 2006), and recently, evidence for horizontal transfer of mitochondrial DNA to the plastid genome in *Pariana* Aubl., a bamboo genus was found (Ma et al. 2015).

Here, we sequenced and analyzed the complete plastome sequences of two Brazilian native species of tribe Bambuseae: *Guadua chacoensis* and *Merostachys sp*. We also performed a full plastome phylogeny using 20 Bambuseae species with 2 newly sequenced and 18 existing plastomes, and characterized the occurrence, type, and distribution of SRRs in the Bambuseae. *Merostachys sp*. is the first species of subtribe Arthrostylidiinae to have its plastome sequenced.

Material and Methods

Sequencing and assembly of the bamboo plastomes

Fresh leaf material of *Merostachys sp*. (voucher Th. Greco 18 - FLOR_16/09/2011) were collected from a natural population, and *Guadua chacoensis* (voucher Th. Greco 159 – FLOR_15/02/2013) were

collected from cultivated plants in Florianópolis, Santa Catarina – Brazil (27° 35' 49'' S 48° 32' 56'' W). Both species were located on a private property and were collected with the owner permission. The chloroplast isolation and plastid DNA extraction were carried out according to Vieira et al. (2014). A total of 1 ng of plastid DNA was used to prepare sequencing libraries with Nextera XT DNA Sample Prep Kit (Illumina Inc., San Diego, California, USA) according to the manufacturer's instructions. Libraries were sequenced using MiSeq Reagent Kit v3 (600 cycles) on Illumina MiSeq Sequencer (Illumina Inc., San Diego, California, USA).

The obtained paired-end reads (2 x 300 bp) were used for *de novo* assembly performed by CLC Genomics Workbench 8.0v. Initial annotation of the obtained plastome sequences 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 plastomes were drawn using OrganellarGenomeDRAW (OGDRAW) (Lohse et al. 2013). REPuter (Kurtz et al. 2001) was used to identify the IRs in both plastomes sequenced in this work and those used for phylogeny estimation by forward vs. reverse complement (palindromic) alignment. The minimal repeat size was set to 30 bp and the identity of repeats ≥ 90%. The complete nucleotide sequence of *Merostachys sp.* and *G. chacoensis* plastome were deposited in the GenBank database under accession number KT373815 and KT373814, respectively.

Phylogeny estimation

The phylogeny estimation was done according to Wysocki et al. (2015). The IR_A was omitted to prevent over representation of the IR sequences. *G. chacoensis* and *Merostachys sp.* plastomes were aligned using MAFFT program (Katoh and Standley 2013) along with 18 previously published Bambuseae plastomes (Table 1) and *Lolium perenne* (Poaceae:Pooideae; NC_009950) plastome was used as outgroup. Nucleotide positions that contained one or more gaps introduced by the alignments were omitted from the matrix. The General Time Reversible model of substitution, incorporating invariant sites and a gamma distribution (GTR + I + G), was used in subsequent plastome analyses. Maximum likelihood (ML) analysis was performed using the

RAxML v 7.2.8 (Stamatakis 2006) with 1,000 non-parametric bootstrap replicates. MrBayes 3.2.2 (Ronquist and Huelsenbeck 2003) was used to perform a Bayesian inference (BI) analysis. The Markov chain Monte Carlo (MCMC) analysis was run for 2,000,000 generations. Average standard deviation of split frequencies remained below 0.001 after the 25% percent burn-in. Resulting trees were represented and edited using FigTree v1.4.1.

Simple sequence repeats analysis

Simple sequence repeats (SSRs) were detected using MISA perl script, available at (http://pgrc.ipk-gatersleben.de/misa/), with threshold 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. For the SSRs identification, we used the plastid genome sequence of all species described in Table 1 with one IR region removed, and in their coding sequences (CDS).

Table 1. Comparison of plastid genome of tribe Bambuseae (Poaceae) analyzed in this study.

Subtribe	Species	Plastome (bp)	LSC[a] (bp)	SSC[b] (bp)	IR[c] (bp)	Accession
Arthrostylidiinae	*Merostachys sp.**	136,334	82,859	12,960	20,258	KT373815
Bambusinae	*Bambusa arnhemica*	139,289	82,802	12,891	21,798	NC_026958
	B. bambos	139,606	83,145	12,851	21,805	NC_026957
	B. emeiensis	139,493	82,988	12,901	21,802	NC_015830
	B. multiplex	139,394	82,923	12,875	21,798	NC_024668
	B. oldhamii	139,350	82,889	12,881	21,790	NC_012927
	Dendrocalamus latiflorus	139,394	83,010	12,874	21,755	NC_013088
	Greslania sp.	139,351	82,667	12,980	21,852	NC_026961
	Neololeba atra	139,395	82,905	12,926	21,782	NC_026964
Chusqueinae	*Chusquea sp.*	138,257	81,791	12,894	21,786	KP_319242
	C. spectabilis	136,848	80,743	12,671	21,717	NC_026959
	C. liebmannii	138,001	81,501	12,892	21,804	NC_026969
	C. circinata	137,951	81,431	12,912	21,804	NC_027490
Guaduinae	*Guadua weberbaueri*	135,324	82,806	12,930	19,794	NC_026991

	*G. chacoensis**	135,403	82,877	12,980	19,773	KT373814
	Olmeca reflexa	136,213	82,726	12,945	20,271	NC_026965
	Otatea acuminata	136,351	82,859	12,948	20,272	NC_026971
	O. glauca	136,377	82,841	12,769	20,333	KP_319243
Hickeliinae	*Hickelia madagascariensis*	138,276	81,925	12,743	21,804	NC_026962
Melocanninae	*Neohouzeaua sp.*	139,458	82,991	12,879	21,794	NC_026963

*Species with plastid genomes sequenced in this study.
[a]Large Single Copy Region.
[b]Short Single Copy Region.
[c]Inverted Repeat Region.

Results

Plastome assembly and content

The Illumina MiSeq reads obtained and submitted to *de novo* assembly resulted in a high genome coverage for both plastomes, ~140x for *G. chacoensis* and ~460x for *Merostachys sp.* (Table 2). The determined complete plastome sequence of *Merostachy*s sp. and *G. chacoensis* is 136,334 bp and 135,403 bp in size, respectively, with a GC content determined of 38.81% for *Merostachys sp.* and 38.76% for *G. chacoensis*. The *Guadua* genus presents the smaller plastome size in comparison with other species belonging to tribe Bambuseae (135,324-135,403 bp), ~1000 bp smaller than *Merostachys sp.*, but with an identical gene content (Table 1 and 3). Both plastomes exhibit the general quadripartite structure typical of angiosperms, consisting of a pair of IRs (20,258–19,773 bp) separated by the LSC (82,859–82,877 bp) and SSC (12,960–12,980 bp) regions (Table 1). The IR/SSC boundary is just within the coding sequence of *ndhH*, creating a short *ndhH* fragment in IR_A. They encode an identical set of 132 genes and 4 pseudogenes with the same gene order and gene clusters, of which 90 genes were single copy and 21 genes were duplicated (Figure 1; Table 3). The following genes were identified and are listed in Figure 1 and Table 3: 4 duplicated ribosomal RNA genes, 21 unique and 9 duplicated transfer RNA genes, 15 unique and 6 duplicated genes encoding large and small ribosomal subunits, 1 translational initiation factor, 4 genes encoding DNA-dependent RNA polymerases, 45 unique and 1 duplicated genes encoding photosynthesis-related proteins, 4 unique and 1 duplicated genes encoding other proteins, including the unknown function gene ycf68.

Table 2. *Guadua chacoensis* and *Merostachys sp.* plastid genome sequencing and assembly data.

Species	Plastome size (bp)	Total plastid read count	Mean read lenght (bp) [a]	Reads mapped in aligned pairs	Average coverage
Guadua chacoensis	135,403	70,760	272.72	64,990	141.77
Merostachys sp.	136,334	225,233	282.55	221,164	462.13

[a] Mean read lenght (bp) after trimming usign CLC Genomics Workbench using quality score limit 0.05

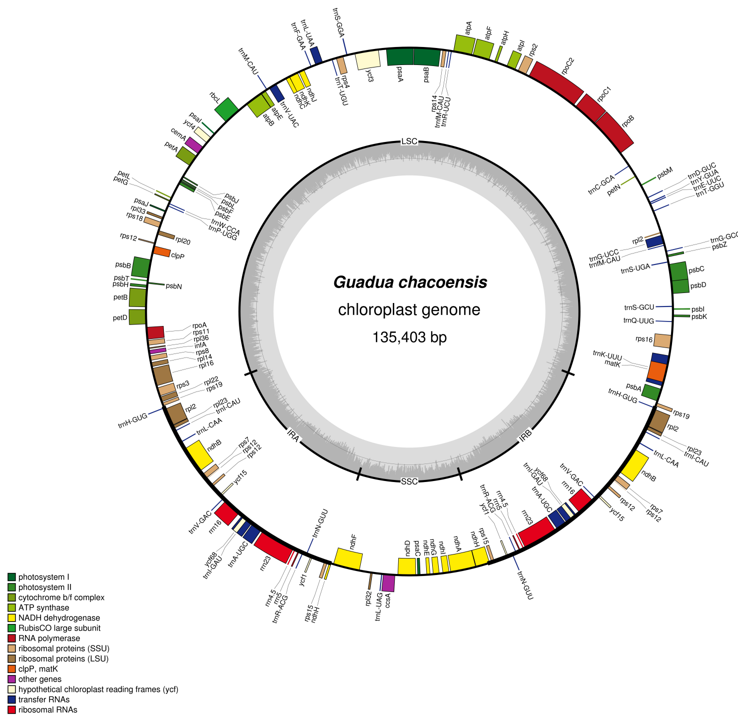

Figure 1. Gene map of *Guadua chacoensis* plastid genome. Genes drawn inside the circle are transcribed clockwise, and genes drawn outside are counterclockwise. Genes belonging to different functional groups are color-coded. The darker gray in the inner circle corresponds to GC content, and the lighter gray to AT content. Gene map of *Merostachys sp.* is equal to *G. chacoensis* in gene order and content.

Table 3. List of genes identified in *Guadua chacoensis* and *Merostachys sp.* plastid genome. The gene content of these species is strictly the same.

Category of genes	Group of Gene	Name of gene			
	Ribosomal RNA genes	*rrn23***	*rrn4.5***	*rrn5***	*rrn16***
	Transfer RNA genes	*trnA-UGC*,***	*trnC-GCA*	*trnD-GUC*	*trnE-UUC*
		trnF-GAA	*trnG-UCC**	*trnG-GCC*	*trnH-GUG***
		trnI-GAU,***	*trnI-CAU***	*trnL-UAG*	*trnL-CAA***
		*trnL-UAA**	*trnK-UUU**	*trnM-CAU*	*trnfM-CAU***
		*trnN-GUU***	*trnP-UGG*	*trnQ-UUG*	*trnR-ACG***
		trnR-UCU	*trnS-GCU*	*trnS-UGA*	*trnS-GGA*
		trnT-UGU	*trnT-GGU*	*trnV-GAC***	*trnV-UAC**
		trnW-CCA	*trnY-GUA*		
	Small subunit of ribosome	*rps2*	*rps3*	*rps4*	*rps7***
		rps8	*rps11*	*rps12*,***	*rps14*
		*rps15***	*rps18*	*rps19***	*rps16**
	Large subunit of ribosome	*rpl2*,***	*rpl14*	*rpl16**	*rpl20*
		rpl22	*rpl23***	*rpl32*	*rpl33*
		rpl36			
	DNA-dependent RNA polymerase	*rpoA*	*rpoB*	*rpoC1*	*rpoC2*
	Translational initiation factor	*infA*			
Photosynthesis	Subunits of photosystem I	*psaA*	*psaB*	*psaC*	*psaI*
		psaJ	*ycf3**	*ycf4*	
	Subunits of photosystem II	*psbA*	*psbB*	*psbC*	*psbD*
		psbE	*psbF*	*psbH*	*psbI*
		psbJ	*psbK*	*psbL*	*psbM*
		psbN	*psbT*	*psbZ*	
	Subunits of cytochrome	*petA*	*petB**	*petD**	*petL*
		petN	*petG*		

	Subunits of ATP synthase	*atpA*	*atpB*	*atpE*	*atpF**
		atpH	*atpI*		
	Large subunit of Rubisco	*rbcL*			
	Subunits of NADH dehydrogenase	*ndhA**	*ndhB*,***	*ndhC*	*ndhD*
		ndhE	*ndhF*	*ndhG*	*ndhH*
		ndhI	*ndhJ*	*ndhK*	
Other genes	Maturase	*matK*			
	Envelope membrane protein	*cemA*			
	C-type cytochrome synthesis gene	*ccsA*			
	Protease	*clpP*			
Unknown function		*ycf68***			
Pseudogene		*rpl2*	*ndhH*	*ycf1***	*ycf15***

* Intron-containing genes
** Duplicated genes

Phylogeny estimation

The matrix of the 21 completely aligned Bambusoideae plastomes was 122,950 nucleotide positions in length, and the exclusion of gaps reduced this matrix to 105,270 sites. ML (-lnL = -206179.38) and BI (-lnL = -206212.43) analyses produced phylogenomic trees that were identical in topology, which is represented by Figure 2. The ML bootstrap values between 68% and 98% for 4 nodes and 100% for the rest, and the BI posterior probability values between 0.95 and 0.99 for 3 nodes and 1.0 for the rest (Figure 2). These trees supported monophyly of Paleotropical and Neotropical Bamboos clades. The Neotropical bamboos segregated into three well-supported lineages, Chusqueinae, Guaduinae, and Arthrostylidiinae, with Guaduinae forming a well supported sister relationship with Arthrostylidiinae. All Neotropical bamboos genera (*Chusquea*, *Guadua*, *Olmeca*, *Otatea*, *Merostachys*) was resolved as monophyletic with 100% ML bootstrap support (Figure 2). The Paleotropical bamboos segregated into two well-supported lineages, Hickeliinae and Bambusinae + Melocanninae. The genus

Bambusa was resolved as monophyletic with 100% ML bootstrap support with very short branches and two internal nodes with 81% and 98% ML bootstrap support (Figure 2). *Dendrocalamus latiflorus* (Bambusinae) and *Neohauzeana sp.* (Melocanninae) formed a well supported node, with 100% ML bootstrap support. These two species are nowadays classified as belonging to different subtribes (Clark et al. 2015). *Greslania sp.* and *Neololeba atra* form a well supported sister relationship.

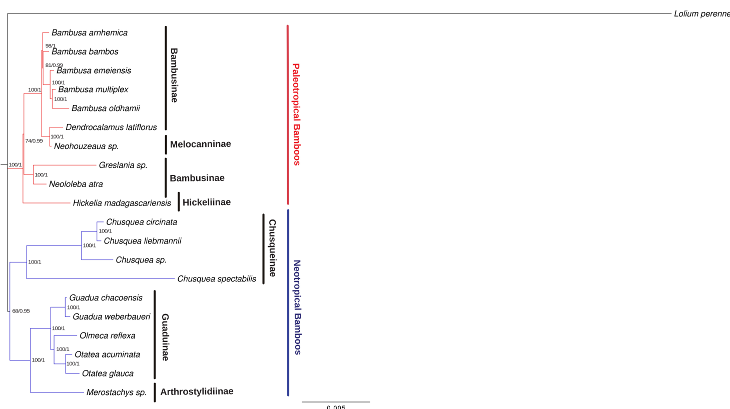

Figure 2. Bayesian phylogeny of 20 complete plastomes sequences of Bambuseae tribe species and the outgroup *Lolium perenne* (Poaceae: Pooideae). The numbers above the branches are maximum likelihood bootstrap values/Bayesian posterior probabilities. The branch length is proportional to the inferred divergence level, the scale bar indicating the number of inferred nucleic acids substitutions per site.

Simple sequence repeats analysis

We analyzed the occurrence, type, and distribution of SRRs in the Bambuseae species listed in Table 1. In the complete plastome

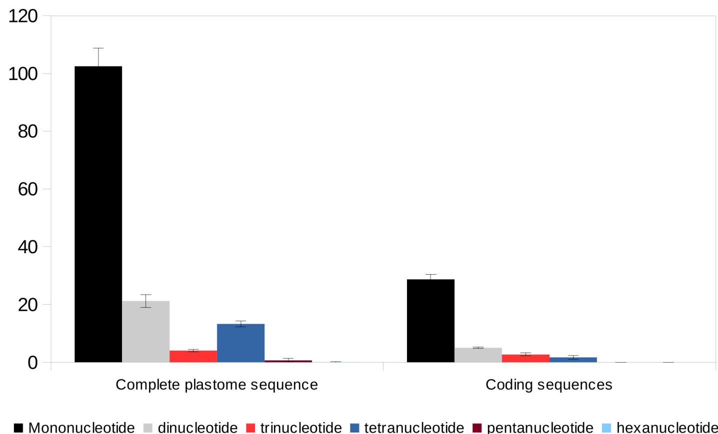

Figure 3. SSR unit size distribution in Bambuseae tribe complete plastome sequences and coding sequences. For this analysis, we considered as complete plastome sequence the SSC, LSC and only one copy of the IR sequence.

sequence, the average of the total identified SSR in all Bambuseae species were 141.8 (1 SSR unit every 823.14 bp), while for coding sequences a lower value was found (38.15; 1 SSR unit every 1466.60 bp). Among them, mono- and dinucleotide repeats were the most common in both complete plastome and coding sequences, whereas tri-, and tetranucleotide repeats occurred with lower frequency (Figure 3). Penta- and hexanucleotide repeats occurred only in complete plastome sequence, no penta or hexanucleotide repeats were identified in coding sequences for any analyzed species (Figure 3). In addition, in the complete plastome sequence, the tetranucleotide occurred in higher frequency than trinucleotide repeats, differently from coding sequences where tetranuclotide repeats occurred in lower frequency (Figure 3).

For the complete plastome sequence, the most mononucleotide repeats identified were constituted by A/T sequences (96.61%), and for the dipolymers, 56.85% were also constituted by multiple A and T bases, no dinucleotide repeats constituted by multiple C and G bases were found (Figure 4). Similarly, for trinucleotide repeats, 31.92% were constituted by multiple A and T bases (AAT/ATT), 48.17% by AAG/CTT and 19.92% by AGC/CTG. We also identified 11 tetranucleotide repeats, with AAAT/ATTT being the most common

(33.86%). Interestingly, we also did not identified any tri- of tetranucleotide unit size constituted by only C or G bases (Figure 4).

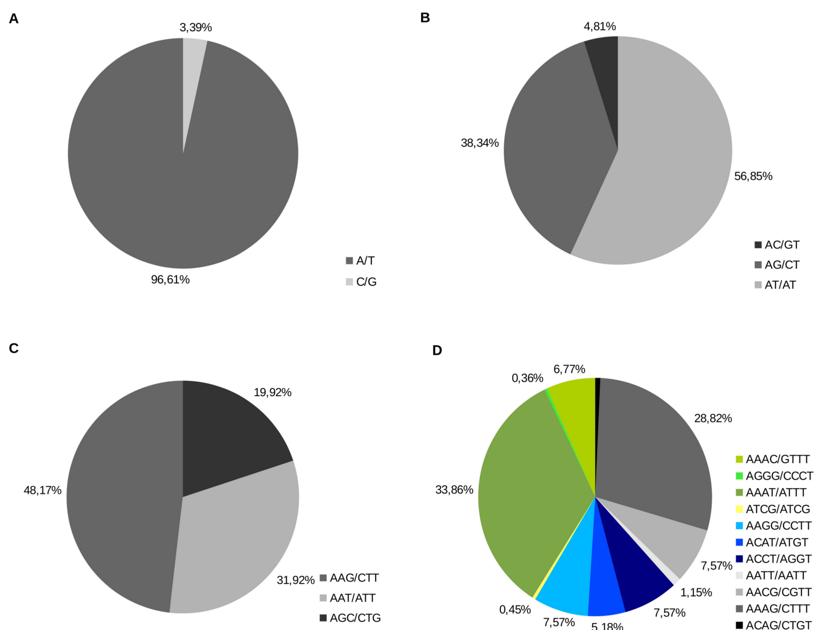

Figure 4. SSR unit type distribution for mono- (A), di- (B), tri- (C), and tetranucleotide repeats (D) in Bambuseae tribe complete plastome sequences. For this analysis, we considered as complete plastome sequence the SSC, LSC and only one copy of the IR sequence.

Discussion

Plastome assembly and content

The plastome size in Neotropical bamboos was typically shorter than in Paleotropical bamboos, with determined plastomes measuring between 135,324-138,257 bp and 138,276-139,606 bp, respectively (Table 1). This increase in Paleotropical plastome sizes seems to be distributed by LSC and IR regions, which ranges between 80,743-82,859 bp and 81,925-83,145 bp for LSC regions in Netropical and Paleotropical bamboos, respectively, and between 19,773-21,804 bp and

21,755-21,852 bp for IR regions, also respectively. Differently, for the SSC regions, little length variation was found (12,671-12,980 bp for Neotropical and 12,743-12,980 bp for Paleotropical bamboos), suggesting that this region is more conserved in size than LSC and IR regions for Bambuseae. The IR/SSC boundary was conserved in Bambuseae tribe, in *G. chacoensis*, *Merostachys sp.* and all Bambuseae plastomes analyzed in this study (Table 1) the IR/SSC boundary was just within the coding sequence of *ndhH*, with a short *ndhH* fragment in IR_A (Wu et al. 2009; Zhang et al. 2011; Burke et al. 2012). This feature is similar to the results found in other Poaceae species (Ogihara et al. 2002; Saski et al. 2007), but differ from the Olyreae species, which were shown to present the IR/SSC boundary within *ndhF* coding sequence, an suggested synapomorphy for the tribe (Burke et al. 2012, 2014). The gene content and gene order of the *G. chacoensis* and *Merostachys sp.* plastomes were identical to other sequenced Bambuseae plastomes, including the loss of introns in *clpP* and *rpoC1,* the loss of *accD*, *ycf1* and *ycf2* genes, and the IR expansion to include *rps19* (Wu et al. 2009; Zhang et al. 2011; Burke et al. 2012).

Phylogeny estimation

Plastid genomes are very conserved evolutionarily, presenting rates of substitutions extremely low as compared to nuclear genomes (Palmer et al. 1985). In addition, different plastome regions evolve at different rates, allowing measuring evolutionary distance at many taxonomy levels (Palmer, 1985; Shaw et al. 2005, 2007). Phylogenetic inferences based on whole plastome sequences have been used to address evolutionary features in Bambusoideae (Wu et al. 2009; Zhang et al. 2011; Burke et al. 2012,2014; Wysocki et al. 2015). These full plastome analyzes provide enough information to resolve difficult interspecific relationships, an issue to woody bamboos that generally hybridize readily and exhibit very long generation times (Wysocki et al. 2015).

In Bambuseae, the well supported monophyletic lineages that represent Neotropical and Paleotropical woody bamboos retrieved here were previously reported in phylogenetic studies using combined analysis of plastid DNA regions (Sungkaew et al. 2009) and complete plastome sequences (Wysocki et al. 2015). In Wysocki et al. (2015), the segregation of Neotropical bamboos into two well-supported lineages, Chusqueinae and Guaduinae, was reported.

In this work, we present for the first time a phylogenetic tree based on complete plastome sequence including an Arthrostylidiinae species, providing the phylogenetic tree with well supported monophyly of subtribes and with a sister relationship between Guaduinae and Arthrostylidiinae. Previous analyses based on *matK* and/or *ndhF* sequence data also reported this sister relationship (Guala et al. 2000; Zhang and Clark 2000; Soreng et al. 2015), and a study based on morphology and *rpl16* intron sequence data indicate that Guaduinae may be derived from within Arthrostylidiinae (Clark et al. 2007). In Chusqueinae, especially *Chusquea spectabilis*, there was a high substitution rate in both ML and BI analyses. This feature has been reported by Wysocki et al. (2005) and associated with the extremely different flower intervals in Chusqueinae species, and the shorter flowering intervals observed in *C. spectabilis*. In addition, *C. spectabilis* long branches may be associated with the fact that this species belongs to subgenus *Magnifoliae*, one of the two earliest-diverging clades in genus *Chusquea*, and *C. circinata* (subgenus *Rettbergia*) is known to be a sister of the large Euchusquea clade, represented in this analysis by *C. liebmannii* (Fisher et al. 2014).

Low support (82% ML bootstrap) between the Paleotropical bamboos subtribes, Hickeliinae and Bambusinae were reported by Wysocky et al. (2015), corroborating our results with low support between Hickeliinae and Bambusinae + Melocanninae (74% ML bootstrap support). Wysocki et al. (2015) also reported the well supported clade *Neoleba atra* + *Greslania sp.*, as well as *Bambusa spp.* monophyly. However, the inclusion of *Neohouzeaua sp.* (Melocanninae), *resulted* in its unexpected sister relashionship to *Dendrocalamus latiflorus* (Bambusinae), and consequently, the non-monophyly of Bambusinae.

Simple sequence repeats analysis

Powell et al. (1995) introduced the cpSSR as a readily usable chloroplast marker, exhibiting length variation and polymorphism. Even though in plastomes the occurrence of di-, tri-, tetra-, penta-, and hexanucleotide repeats is less common (George et al., 2015), it became a widely used molecular marker, which have aroused considerable interest due to their ability to generate highly informative DNA markers (Provan et al., 2001). These regions may be used for both intraspecific and interspecific variability analyses, with practical value for monitoring

gene flow, population differentiation and cytoplasmic diversity (Powell et al., 1995), in both basic plant sciences and applied agricultural. In native species, cpSSRs have been used most frequently in population genetics studies, but also for understanding uniparental genetic structure (e.g., seed or pollen dispersal) and in studies of hybridization (see review by Wheeler et al. 2014). Given that not all chloroplast loci are likely to be equally diverse across plants, it is important to consider a more targeted approach to developing cpSSR loci specific to a study system rather than relying on universal primers (Wheeler et al. 2014).

The cpSSRs may be identified in completely sequenced plant chloroplast genomes by simple database searches, followed by primers designed to screen for polymorphism even in nonmodel species, and is becoming a common pathway for developing variable markers (Wheeler et al. 2012; Nazareno et al. 2015). George et al. (2015) analyzed the abundance and distribution of simple and compound SSR in 164 sequenced plastomes from wide range of plants. Corroborating our results in Bambuseae, George et al. (2015) described that mononucleotide repeats occurs in higher number as compared to di-, tri-, tetra-, penta-, and hexanucleotide repeats, longer SSRs are excluded from coding regions, AT/TA followed by CT/TC was the most common dinucleotide repeat motif observed, and GC/CG was rarely found and even absent from few plastomes. The several cpSSR sites in Bambuseae plastome can be assessed for the intraspecific level of polymorphism, leading to innovative highly sensitive phylogeographic and population genetics studies for this tribe. *Merostachys sp.* and *G. chacoensis* are both native species from the the Atlantic Forest Biome, which is considered one of the 25 biodiversity hotspots of the world (Myers et al. 2000), increasing the interest in population genetics studies for these species.

Conclusions

The determined complete plastome sequence of *Merostachy*s sp. and *Guadua chacoensis* were identical in gene content and order, also with identical gene content as compared to other Bambuseae plastomes. They have several SSR regions, whose distribution and types are highly similar between sequenced plastomes, including the presence of more cpSSR in complete plastome sequence than in coding sequences. The mono- and dinucleotide repeats were the most common, while penta- and hexanucleotide repeats were identified only in

complete plastome sequence. These cpSSR may be assessed for the intraspecific level of polymorphism, leading to innovative highly sensitive phylogeographic and population genetics, specially for *Merostachys sp.* and *G. chacoensis,* which are both native species from the the Atlantic Forest Biome, the fourth in importance of the 25 biodiversity hotspots of the world. ML and BI trees supported monophyly of Paleotropical and Neotropical Bamboos clades. All Neotropical bamboos genera were resolved as monophyletic. The Paleotropical bamboos segregated into two well-supported lineages, Hickeliinae and Bambusinae + Melocanninae. This is the first report of a phylogenetic tree based on complete plastome sequences including an Arthrostylidiinae species, the *Merostachys sp*.

Availability of supporting data

All nucleotide sequences were deposited in the NCBI Genbank repository. Accessions can be found in Table 1.

LITERATURE CITED

Bamboo Phylogeny Group (2006) The bamboo phylogeny project. Bamboo 27:11–14.

Bamboo Phylogeny Group (2012) An updated tribal and subtribal classification for the Bambusoideae (Poaceae). In: Gielis J, Potters G (Eds) Proceedings of the 9th World Bamboo Congress. Antwerp, Belgium: World Bamboo Organization, pp. 3–27.

Bock R (2006) Extranuclear inheritance: Gene transfer out of plastids. Prog Bot 67:75-100. doi:10.1007/3-540-27998-9_4.Bock 2006

Burke SV, Colin PG, Melvin RD (2012) Plastome sequences of two New World bamboos—*Arundinaria gigantea* and *Cryptochloa strictiflora* (Poaceae)—extend phylogenomic understanding of Bambusoideae. Am J Bot 99:1951-1961. doi:10.3732/ajb.1200365

Burke SV, Clark LG, Triplett JK, Grennan CP, Duvall MR (2014). Biogeography and phylogenomics of new world Bambusoideae (Poaceae), revisited. Am J Bot 101:886-891. doi:10.3732/ajb.1400063

Clark LG, Dransfield S, Triplett JK, Sánchez-Ken JG (2007) Phylogenetic relationships among the one-flowered, determinate genera of Bambuseae (Poaceae: Bambusoideae). Aliso 23:315–332

Clark LG, Londoño X, Ruiz-Sanchez E (2015) Bamboo Taxonomy and Habitat. In: Liese W, Kohl M (eds) Bamboo, Tropical Forestry. Springer International Publishing, Switzerland, pp 1-30.

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. doi:10.1111/j.1752-4571.2011.00223.x

Fisher AE, Clark LG, Kelchner SA (2014) Molecular phylogeny estimation of the bamboo genus *Chusquea* (Poaceae: Bambusoideae: Bambuseae) and description of two new subgenera. Sist Bot *39:*829-844. doi:10.1600/036364414X681554

George B, Bhatt BS, Awasthi M, George B, Singh AK (2015) Comparative analysis of microsatellites in chloroplast genomes of lower and higher plants. Curr Genet 1-13. doi:10.1007/s00294-015-0495-9

Greco TM, Pinto MM, Tombolato FC, Xia N (2015) Diversity of Bamboo in Brazil. J Trop Subtrop Botany 23:1-16. doi: 10.11926/j.issn.1005–3395.2015.01.001

Guala G, Bogler D, Sadle J, Francisco-Ortega J (2000) Molecular evidence for polyphyly in the genus *Apoclada* (Poaceae: Bambusoideae). Bamboo Sci Cult 14:15–20

Huang CY, Ayliffe MA, Timmis JN (2003) Direct measurement of the transfer rate of chloroplast DNA into the nucleus. Nature 422:72-76. doi: 10.1038/nature01435

Huang CY, Grunheit N, Ahmadinejad N, Timmis JN, Martin W (2005) Mutational decay and age of chloroplast and mitochondrial genomes transferred recently to angiosperm nuclear chromosomes. Plant Physiol 138:1723-1733. doi: 10.1104/pp.105.060327

Judziewicz EJ, Clark LG, Londono X, Stern M (1999) American bamboos. Smithsonian Institution Press, Washington DC

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772-780. doi:10.1093/molbev/mst010

Kelchner SA, Group BP (2013) Higher level phylogenetic relationships within the bamboos (Poaceae: Bambusoideae) based on five plastid markers. Mol Phylogenet Evol 67*:*404-413. doi:10.1016/j.ympev.2013.02.005

Khadivi-Khub A, Zamani Z, Fattahi R, Wünsch A (2013) Genetic variation in wild *Prunus* L. subgen. *Cerasus* germplasm from Iran

characterized by nuclear and chloroplast SSR markers. Trees 28:471-485. doi:10.1007/s00468-013-0964-z

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

List of Species of the Brazilian Flora. Rio de Janeiro Botanical Garden. http://floradobrasil.jbrj.gov.br. Accessed 04 July 2015

Lizarazu MA, Agrasar ZER, Vega AS (2011) A New Species of *Merostachys* (Poaceae, Bambusoideae, Bambuseae) and Synopsis of the Genus in Argentina and Neighboring Regions. Sist Bot 36:896–906. doi:10.1600/036364411X604912

Londoño X, Peterson PM (1992) *Guadua chacoensis* (Poaceae: Bambuseae), Its Taxonomic Identity, Morphology, and Affinities. Novon 2:41-47

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 1-7. doi:10.1093/nar/gkt289

Ma PF, Zhang YX, Guo ZH, Li DZ (2015) Evidence for horizontal transfer of mitochondrial DNA to the plastid genome in a bamboo genus. Sci Rep 5:11608. doi:10.1038/srep11608

Myers N, Mittermeier RA, Mittermeier CG, Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. Nature 403: 853–858. doi: 10.1038/35002501

Nazareno AG, Carlsen M, Lohmann LG (2015) Complete Chloroplast Genome of *Tanaecium tetragonolobum*: The First Bignoniaceae Plastome. PLoS ONE 10(6):e0129930. doi:10.1371/journal.pone.0129930

Ogihara Y, Isono K, Kojima T, Endo A, Hanaoka M, Shiina T, Terachi T, Utsugi S, Murata M, Mori N, Takumi S, Ikeo K, Gojobori T, Murai R, Murai K, Matsuoka Y, Ohnishi Y, Tajiri H, Tsunewaki K (2002) Structural features of a wheat plastome as revealed by complete sequencing of chloroplast DNA. Mol Genet Genomics 266:740–746. doi:10.1007/s00438-001-0606-9

Palmer JD (1985) Chloroplast DNA and molecular phylogeny. BioEssays 2:263-267.

Powell W, Morgantet 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. Current Biology 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. doi: 10.1016/s0169-5347(00)02097-8

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.doi:10.3389/fpls.2015.00586

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinform 19:*1572-1574.doi:10.1093/bioinformatics/btg180

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. doi:10.1111/j.1365-294X.2011.05229.x.

Santos-Gonçalves APS, Carvalho-Okano RMD, Filgueiras TS (2012) A new species of *Merostachys* (Poaceae:Bambusoideae) from southeastern Brazil. Syst Bot 37:938-940. doi:10.1600/036364412X656400

Saski C, Lee SB, Fjellheim S, Guda C, Jansen RK, Luo H, Tomkins J, Rognli OA, Daniell H, Clarke JL (2007) Complete chloroplast genome sequences of *Hordeum vulgare*, *Sorghum bicolor* and *Agrostis stolonifera*, and comparative analyses with other grass genomes. Theor Appl Genet 115:571–590. doi:10.1007/s00122-007-0567-4

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: W686–689. doi: 10.1093/nar/gki366

Shaw J, Lickey EB, Beck JT, Farmer SB, Liu W, Miller J, Siripun KC, Winder CT, Schilling EE, Small RL (2005) The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. Am J Bot 92:142-166. doi: 10.3732/ajb.94.3.275

Shaw J, Lickey EB, Schilling EE, Small RL (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. Am J Bot 94:275-288. doi: 10.3732/ajb.94.3.275

Soreng RJ, Peterson PM, Romaschenko K, Davidse G, Zuloaga FO, Judziewicz EJ, Filgueiras TS, Davis JI, Morrone O (2015) A worldwide phylogenetic classification of the Poaceae (Gramineae). J Syst Evol *53:*117-137. doi: 10.1111/jse.12150

Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinform, *22:*2688-2690. doi:10.1093/bioinformatics/btl446

Stegemann S, Bock R (2006) Experimental reconstruction of functional gene transfer from the tobacco plastid genome to the nucleus. Plant Cell 18:2869-2878. doi: 10.1105/tpc.106.046466

Stegemann S, Hartmann S, Ruf S, Bock R (2003) High-frequency gene transfer from the chloroplast genome to the nucleus. Proc. Natl. Acad. Sci. U S A 100:8828-8833. doi: 10.1073/pnas.1430924100

Sungkaew S, Stapleton CM, Salamin N, Hodkinson TR (2009) Non-monophyly of the woody bamboos (Bambuseae; Poaceae): a multi-gene region phylogenetic analysis of Bambusoideae ss. J Plant Res *122:*95-108. doi:10.1007/s10265-008-0192-6

Viana PL, Filgueiras TS (2014) Three new species of Aulonemia (Poaceae: Bambusoideae) from the Brazilian Atlantic rainforest. *Phytotaxa 156:*235-249. doi:10.11646/phytotaxa.156.4.6

Viana PL, Filgueiras TS, Clark LG (2013) Cambajuva (Poaceae: Bambusoideae: Bambuseae: Arthrostylidiinae), a new woody bamboo genus from southern Brazil. Syst Bot *38:*97-103. doi:10.1600/036364413X662097

Vieira LN, Faoro H, Fraga HPF, Rogalski M, Souza EM, et al. (2014) An improved protocol for intact chloroplasts and cpDNA isolation in conifers. PLoS ONE 9(1): e84792 doi:10.1371/journal.pone.0084792.

Wheeler GL, McGlaughlin ME, Wallace LE (2012) Variable length chloroplast markers for population genetic studies in *Acmispon* (Fabaceae). Am J Bot 99:e408 – e410. doi:10.3732/ajb.1200129

Wheeler GL, Dorman HE, Buchanan A, Challagundla L, Wallace LE (2014) A review of the prevalence, utility, and caveats of using chloroplast simple sequence repeats for studies of plant biology Appl Plant Sci *2*:1400059. doi:10.3732/apps.1400059

Wu FH, Kan DP, Lee SB, Daniell H, Lee YW, Lin CC, Lin NS, Lin CS (2009) Complete nucleotide sequence of *Dendrocalamus latiflorus* and *Bambusa oldhamii* chloroplast genomes. Tree Physiol tpp015:1-10. doi:10.1093/treephys/tpp015

Wyman SK, Jansen RK, Boore JL (2004) Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20: 3252–3255. doi: 10.1093/bioinformatics/bth352

Wysocki WP, Clark LG, Attigala L, Ruiz-Sanchez E, Duvall MR (2015) Evolution of the bamboos (Bambusoideae; Poaceae): a full plastome

phylogenomic analysis. BMC Evol Biol *15:*50. doi:10.1186/s12862-015-0321-5

Zhang W-P, Clark LG (2000) Phylogeny and classification of the bamboos (Poaceae: Bambusoideae). In: Jacobs SWL, Everett J (Eds) Grasses: Systematics and Evolution. CSIRO Publishing, Collingwood, Victoria, Australia, pp 35–42.

Zhang Y-J, Ma P-F, Li D-Z (2011) High-Throughput Sequencing of Six Bamboo Chloroplast Genomes: Phylogenetic Implications for Temperate Woody Bamboos (Poaceae: Bambusoideae). PLoS ONE 6(5): e20596. doi:10.1371/journal.pone.0020596