Leila do Nascimento Vieira

DESENVOLVIMENTO DE UM PROTOCOLO DE ISOLAMENTO DE CLOROPLASTOS E DNA PLASTIDIAL EM CONÍFERAS E SEQUENCIAMENTO DO GENOMA PLASTIDIAL DE *Podocarpus lambertii* Klotzch ex Endl.

Dissertação submetida ao Programa de Pós-Graduação em Recursos Genéticos Vegetais da Universidade Federal de Santa Catarina para a obtenção do Grau de mestre em Recursos Genéticos Vegetais
Orientador: Prof. Dr. Miguel P. Guerra

Florianópolis
2014

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

Vieira, Leila do Nascimento
Desenvolvimento de um protocolo de isolamento de cloroplastos e DNA plastidial em coníferas e sequenciamento do genoma plastidial de Podocarpus lambertii Klotzch ex Endl. / Leila do Nascimento Vieira ; orientador, Miguel Pedro Guerra - Florianópolis, SC, 2014.
79 p.

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

Inclui referências

1. Recursos Genéticos Vegetais. 2. Cloroplasto. 3. Genoma plastidial. 4. Coníferas. 5. Podocarpus lambertii. I. Guerra, Miguel Pedro. II. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Recursos Genéticos Vegetais. III. Título.

Este trabalho é dedicado aos meus queridos pais, Rogério e Eloisa e ao Julião (*in memoriam*).

AGRADECIMENTOS

Agradeço aos meus queridos pais, Rogério e Eloisa, e ao meu irmão Felipe, pelo apoio incondicional durante toda minha formação acadêmica.

Ao Hugo, que esteve do meu lado durante toda a realização desse trabalho, por seu carinho, compreensão e infinita paciência.

Ao meu orientador, Prof. Miguel Pedro Guerra, por todos os ensinamentos em nessa caminhada acadêmica, acima de tudo pela confiança depositada para realização desse projeto e diversas oportunidades de aprendizado junto ao LFDGV.

Ao Prof. Rubens Onofre Nodari e Prof. Marcelo Rogalski pela atenção, por todas as trocas de experiências que tornaram esse trabalho possível e pelo apoio de sempre.

À toda a equipe da UFPR, Prof. Emanuel M. de Souza, Prof. Fábio de O. Pedrosa, ao Rodrigo L.A. Cardoso e principalmente ao Helisson Faoro por terem tornado possível o sequenciamento e análises de dados presentes nesse trabalho.

Aos amigos Douglas Steinmacher e Angelo Heringer, que mesmo de longe sempre estiveram dando seu apoio, pela amizade construída e os ótimos momentos compartilhados.

Aos queridos Ramon, Clarissa, Dorival, Montagna, Liliana, Catarina, Antônio, Sabrina, Morgana, Carol Cristofolini, Sara Ferrigo, Larissa Zanette e aos demais colegas do LFDGV pelos momentos de descontração e apoio ao longo dessa caminhada.

Ao Eddie Vedder e ao Corey Taylor pela indispensável trilha sonora durante os longos momentos de análises de bioinformática e escrita do trabalho.

À Bernadete Ribas pelo suporte administrativo durante o mestrado.

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

RESUMO

O sequenciamento do genoma plastidial de coníferas vem sendo realizado por meio do isolamento do DNA total, seguido por amplificações do DNA plastidial através de PCR de longo alcance. Esse método de sequenciamento é mais trabalhoso do que o sequenciamento a partir DNA plastidial. O sequenciamento do genoma plastidial permite a realização de várias análises comparativas, como estudos filogenéticos, filogeografia, evolução e estrutura, localização de regiões repetidas, identificação de sítios de edição de RNA, entre outras. Dessa forma, o presente trabalho visou desenvolver um protocolo de isolamento de cloroplastos e DNA plastidial em coníferas e sequenciar o genoma plastidial do *Podocarpus lambertii*. O protocolo foi desenvolvido a partir da *Araucaria angustifolia* e *Araucaria bidwilli* (Araucariaceae), *P. lambertii* (Podocarpaceae) e *Pinus patula* (Pinaceae). O protocolo se baseia no isolamento plastidial a partir de um tampão salino, seguido por gradiente salino de Percoll. Essas duas estratégias combinadas, reduziram a contaminação e aumentaram o rendimento do DNA plastidial. O DNA plastidial foi sequenciado em sequenciador Illumina MiSeq, obtendo-se uma cobertura média do genoma de: 24,63 para *A. angustifolia*, 135,97 para *A. bidwilli*, 1196,10 para *P. lambertii* e 64,68 para *P. patula*. O genoma plastidial do *P. lambertii* é formado por 133.734 pb e apresentou a perda de uma das regiões invertidas repetidas. O genoma contém 118 genes únicos e 1 tRNA duplicado (*trnN*-GUU). Estruturalmente, o genoma plastidial do *P. lambertii* apresenta quatro grandes inversões de aproximadamente 20.000 pb quando comparado ao *Podocarpus totara*. O genoma plastidial do *P. lambertii* apresenta um total de 28 repetições em tandem e 156 *simple sequence repeat* (SSRs). Os resultados obtidos são inovadores uma vez que um protocolo viável para o isolamento de DNA plastidial de coníferas com alta qualidade e quantidade de DNA foi obtido. Adicionalmente, a sequência do genoma plastidial do *P. lambertii* revelou diferenças estruturais significativas, mesmo em relação à outra espécie do mesmo gênero. Os diversos SSRs encontrados no genoma plastidial do *P. lambertii* podem ser avaliados como regiões polimórficas intraespecíficas, que podem levar, entre outros, a estudos filogenéticos com maior sensibilidade.

Palavras-chave: cpDNA, genoma plastidial, coníferas, cloroplasto, sequenciamento, *Podocarpus lambertii.*

ABSTRACT

Plastid genome sequencing protocols for conifer species have been based mainly on long-range PCR, which is known to be time-consuming and difficult to implement than sequencing from isolated plastid DNA. Plastid genome sequencing are useful for several comparative analyzes, as phylogeny, phylogeography, evolution and structure, location of repeated regions, identification of RNA editing sites, among others. Thus, this study aimed to develop a protocol for chloroplast and plastid DNA isolation in conifers and sequence the chloroplast genome of *Podocarpus lambertii*. The protocol was developed using *Araucaria angustifolia* and *Araucaria bidwilli* (Araucariaceae), *P. lambertii* (Podocarpaceae) and *Pinus patula* (Pinaceae) species. The protocol is based on plastid isolation with saline buffer followed by saline Percoll gradient. These two combined strategies reduced contamination and increased the plastid DNA yield. The plastid DNA was sequenced in MiSeq Illumina sequencer, and the average genome coverage were 24.63 to *A. angustifolia*, 135.97 to *A. bidwilli*, 1196.10 to *P. lambertii*, and 64.68 to *P. patula*. The chloroplast genome of *P. lambertii* is 133.734 bp in length and lacks one of the inverted repeat regions. The genome contains 118 unique genes and one duplicated gene, the tRNA (*trnN*-GUU). Structurally, the plastid genome of *P. lambertii* shows four large inversions of approximately 20,000 bp compared to *Podocarpus totara*. The plastid genome of *P. lambertii* shows a total of 28 tandem repeats and 156 simple sequence repeat (SSR). Results show that this improved protocol is suitable for enhanced quality and yield of chloroplasts and cpDNA isolation from conifers. Additionally, the sequence of the plastid genome of *P. lambertii* revealed significant structural differences, even in relation to other species of the same genus. The various SSRs found in the plastid genome of *P. lambertii* can be evaluated for intraspecific polymorphic regions, which may allow highly sensitive phylogeographic and population structure studies, as well as phylogenetic studies of species of this genus.

Keywords: cpDNA, plastid genome, conifers, chloroplast, next generation sequencing, *Podocarpus lambertii*.

LISTA DE FIGURAS

ESTADO DA ARTE

CAPÍTULO 1

CAPÍTULO 2

LISTA DE QUADROS

CAPÍTULO 1

CAPÍTULO 2

LISTA DE ABREVIATURAS E SIGLAS

ATP - *Adenosine triphosphate*
BSA - *Bovine serum albumin*
CDS - *Coding sequences*
cp - *Chloroplast*
cpDNA – *chloroplast* DNA
DOGMA – *Dual organellar genome annotator*
DTT - *Dithiothreitol*
EDTA - *Ethylenediamine tetraacetic acid*
IGS - *Intergenic spacers*
IR – *Inverted repeat*
KAc - *Potassium acetate*
LSC – *Large sequence copy*
NADH - *Nicotinamide adenine dinucleotide*
NCBI – *National center for biotechnology information*
OGDRAW – *Organellar genome draw*
ORF – *Open reading frame*
PCR – *Polymerase chain reaction*
PVP-40 - *Polyvinylpyrrolidone*
RFLP - *Restriction fragment length polymorphism*
SSC – *Small sequence copy*
SSR – *Simple sequence repeat*
TIC - *Translocon of inner membrane*
TOC - *Translocon of outer membrane*
TRF – *Tandem repeats finder*

SUMÁRIO

INTRODUÇÃO

O presente trabalho faz parte do projeto intitulado "Análises genômicas e transcriptômicas nas coníferas brasileiras *Araucaria angustifolia*, *Podocarpus sellowii*, *Podocarpus lambertii* e *Retrophyllum piresii* visando uso, conservação, estudos evolutivos, moleculares e biotecnológicos" financiado pela Fundação de Amparo à Pesquisa e Inovação do Estado de Santa Catarina (FAPESC). As análises foram realizadas no Laboratório de Fisiologia do Desenvolvimento e Genética Vegetal da Universidade Federal de Santa Catarina e no Departamento de Bioquímica e Biologia Molecular da Universidade Federal do Paraná sob a supervisão do Prof. Dr. Emanuel Maltempi de Souza e Prof. Dr. Fábio de Oliveira Pedrosa. As coníferas nativas do Brasil e de Santa Catarina, *A. angustifolia* e *P. lambertii*, foram escolhidas para o desenvolvimento do trabalho.

A. angustifolia é uma espécie de grande relevância ecológica, criando um ambiente ideal para o crescimento de outras espécies tolerantes à sombra e tendo suas sementes como alimento para a fauna selvagem. Do ponto de vista econômico essa espécie também possui madeira de alta qualidade, sendo empregada para construção civil, móveis e para a produção de celulose. Atualmente calcula-se que os remanescentes florestais de *A. angustifolia* representem apenas 2% da ocorrência natural. Essa drástica redução vem causando erosão genética em populações naturais.

P. lambertii pertence ao gênero *Podocarpus,* considerado um dos mais diversos entre todas as coníferas. *P. lambertii* é característico da Floresta Ombrófila Mista, possuindo sementes do tipo ortodoxas. Estudos recentes mostraram que o *P. lambertii* no estado de Santa Catarina apresentara baixa diversidade genética e um alto índice de fixação para a média das populações.

Diante da complexidade do isolamento plastidial em coníferas o primeiro ponto abordado pelo trabalho foi o desenvolvimento de um protocolo para o isolamento do cloroplasto. Para tanto, foram utilizadas quatro espécies: *A. angustifolia* (Araucariaceae), *P. lambertii* (Podocarpaceae), *Araucaria bidwillii* (Araucariaceae) e *Pinus patula* (Pinaceae). As primeiras três espécies foram escolhidas devido ao interesse futuro do sequenciamento do seu genoma plastidial. A última, para validação do protocolo com outra família de conífera.

A partir dos resultados obtidos com o isolamento do DNA plastidial, foi realizado o sequenciamento completo do genoma plastidial do *P. lambertii.*

REVISÃO BIBLIOGRÁFICA

1. Família Podocarpaceae

Dentre as Gimnospermas, a família Podocarpaceae se destaca pela presença de um receptáculo dilatado aderido à semente, conhecido como epimácio (Figura 1). Esta família compreende 18 gêneros e aproximadamente 170 espécies distribuídas principalmente no hemisfério sul (Figura 2), estendendo-se também ao norte até a China subtropical, Japão, México e Caribe (FARJON, 1998; BIFFIN, CONRAN e LOWE, 2010).

Figura 1. **Semente madura do *Podocarpus lambertii*.** Seta: epimácio. Asterisco: semente.

Dentre as espécies da família Podocarpaceae encontradas no Brasil, destaca-se o *P. lambertii*, *P. sellowii* e o *R. piresii*.

O *P. sellowii* distribui-se ao desde o Rio Grande do Sul até as pequenas serras e brejos de altitude no Nordeste, junto as Floresta Ombrófilas Densa e Floresta Estacional Semidecidual, além das florestas de galerias serranas no domínio do cerrado (CARVALHO, 1994).

A semente do *P. sellowii,* sem o receptáculo carnoso, possui em média 9,01 mm (GARCIA, 2006). Suas sementes apresentam características recalcitrantes, com o grau crítico de umidade em torno de 26% (GARCIA, 2008).

Figura 2. Distribuição geográfica da família Podocarpaceae. Fonte: *Missouri Botanical Garden*.

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

R. piresii é uma espécie endêmica de Rondônia, no Parque Nacional dos Pacaás Novos. Essa espécie foi classificada em 1976 por João M. Pires, levando o nome de *Decussocarpus piresii* J. Silba em homenagem ao seu descobridor. Nessa viagem foi realizada a coleta de algumas sementes dos espécimes encontrados no Parque Nacional dos Pacaás Novos, que foram levadas ao Museu Emílio Goeldi, onde atualmente se encontram exemplares dessa espécie (WILLIAN RODRIGUES, Comunicação pessoal). Devido à dificuldade de acesso ao local de ocorrência natural da espécie, pouco se tem estudo sobre a mesma. No entanto, registrou-se que João Murça Pires tinha dúvidas sobre a conceituação da espécie, devido à semelhança com a espécie andina *R. rospigliosii,* pois as diferenças das formas andina e amazônica segundo ele não eram convincentes, talvez tratando-se apenas de variedades botânicas (LISBOA E ALMEIDA, 1995).

2. *Podocarpus lambertii* Klotzch ex Endl.

O gênero *Podocarpus* é um dos mais diversos entre todas as coníferas, englobando mais de 100 espécies, muitas das quais apresentam grande interesse florestal. Além disso, o gênero apresenta ampla distribuição geográfica mundial, sendo registrada sua presença na 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, 1998).

P. lambertii ocorre desde o estado do Rio Grande do Sul até a Bahia, sendo encontrado também no nordeste da Argentina. A espécie é característica da Floresta Ombrófila Mista (Figura 3), onde se apresenta associada à *A. angustifolia* (CARVALHO, 1994). Apresenta características de uma árvore perenifólia de altura variável, medindo de 1 a 10 m de altura, apresentando tronco geralmente tortuoso, inclinado e curto, podendo apresentar-se reto na floresta (CARVALHO, 1994). Prefere locais pouco pedregosos, pouco inclinados, relativamente úmidos, com alta frequência de indivíduos e alta densidade do sub-bosque, indicando ser esta uma espécie secundária tardia tolerante a sombra (LONGHI, 2010).

Dentre as Gimnospermas, a família Podocarpaceae se destaca pela presença de um receptáculo carnoso aderido à semente, conhecido como epimácio. A semente do *P. lambertii,* sem esse receptáculo carnoso, possui em média 4,7 mm (GARCIA, NOGUEIRA e ALQUINI, 2006). Suas sementes são classificadas como ortodoxas, visto que a partir de um grau de umidade inicial de 28,7% podem sofrer dessecação atingindo 5,7% de umidade com uma taxa de germinação de 72,06% (GARCIA e NOGUEIRA, 2008).

BITTENCOURT (2011) caracterizou a diversidade genética e estrutura de populações naturais *P. lambertii* no estado de Santa Catarina. As populações avaliadas apresentaram baixa diversidade genética e um alto índice de fixação para a média das populações, fortes indícios de que as

Figura 3. ***Podocarpus lambertii (em primeiro plano)*** **e** ***Araucaria angustifolia*** **em mata nativa no município de Lages-SC.**

populações avaliadas sofreram sérios desequilíbrios resultantes de ações antrópicas de fragmentação e exploração. Alelos raros foram também observados nas populações e uma divergência genética significativa entre as populações, evidenciando um reduzido fluxo gênico histórico e um grande risco de perda de diversidade.

Alguns estudos filogenéticos foram realizados na família Podocarpaceae com base tanto em caracteres morfológicos (KELCH, 1998) quanto em caracteres moleculares (KELCH, 1998, 2002; BIFFIN, CONRAN e LOWE, 2010). Em geral, a filogenia molecular está de acordo com as classificações baseadas nos caracteres morfológicos para relações intergenéricas. No entanto, muitas das relações intragenéricas permanecem com pouco apoio.

3. Estrutura e aplicação do sequenciamento do genoma plastidial

O cloroplasto é uma organela originada a partir da endossimbiose, sendo as cianobactérias o grupo mais próximo do genoma plastidial. Não

havendo definição clara de qual linhagem desse grupo de cianobactérias originou essa organela (TIMMIS *et al*., 2004). Durante mais de um bilhão de anos de evolução, os três genomas das células vegetais (nuclear, plastidial e mitocondrial) sofreram mudanças estruturais dramáticas para otimizar a expressão compartimentalizada do material genético e a comunicação entre esses compartimentos (GREINER e BOCK, 2013).

Dentre essas modificações, destaca-se a transferência em larga escala de informação genética do cloroplasto para o núcleo, resultando em uma expressiva redução no tamanho do genoma plastidial (GREINER e BOCK, 2013). Por conseguinte, os genomas plastidiais contemporâneos contêm apenas uma pequena proporção dos genes dos seus ancestrais. No entanto, os genes remanescentes permanecem com a característica procariótica de organização em *operons* (WICKE *et al.*, 2011). Calcula-se que essa redução foi de em torno de 3.000 genes para cerca de 120 genes, o que levou a uma capacidade limitada de codificação do plastoma. Estima-se que os cloroplastos importam mais de 95% das suas proteínas do citosol (BOCK, 2007).

Para tanto, é necessário que as pre-proteínas endereçadas ao cloroplasto atravessem três distintos sistemas de membranas – a membrana exterior e interior, que circundam a organela, e a membrana do tilacóide, que contém as proteínas fotossinteticamente ativas. As pre-proteínas destinadas ao tilacóide são sintetizadas com uma extensão amino-terminal chamada de pre-sequência, que é reconhecida pelo translocon da membrana exterior (TOC) e da membrana interior (TIC). No estroma ocorre a clivagem da extensão amino-terminal, produzindo a forma madura da proteína (SOLL e SCHLEIFF, 2004).

O DNA plastidial (cpDNA) possui características próprias distintas do DNA nuclear. Essas características incluem diferentes porcentagens de conteúdo GC e, na maior parte das vezes, ausência de metilação no DNA, ou seja, ausência da 5´-metilcitosina. Alguns autores apontam que há um baixo percentual de metilação do DNA no cloroplasto (VANYUSHIN, 2006; HUANG *et al*., 2012). No entanto, FOJTOVA, KOVARIK e MATYASEK (2001) consideram que na maioria dos casos a detecção de metilação no cpDNA ocorre devido à "artefatos" gerados pelas enzimas de restrição utilizadas para as análises. AHLERT *et al.* (2009) mostraram através da introdução dos genes da adenina e citosina DNA metiltransferase de cianobactérias que plantas com cpDNA metilado não apresentam diferenças morfológicas em relação ao controle não transformado.

Apesar de pequeno, o cpDNA compreende uma fração significante do DNA total, sendo estimado em 9% em *Nicotiana tabacum* (TEWARI e

WILDMAN, 1966). Dependendo da espécie, tecido, estádio de desenvolvimento e condições ambientais o nível de ploidia de um cloroplasto pode chegar a mais de 10.000 cópias idênticas do seu genoma por célula (BENDICH, 1987).

Os plastídios e mitocôndrias são herdados de forma não-Mendeliana em todos os eucariotos. Na maioria dos organismos, os genomas organelares são herdados de apenas um dos pais, com a herança materna sendo muito mais difundida do que a herança paterna (HAGEMANN, 2004). Sabe-se que a maior parte das angiospermas possui herança materna, com raros casos de herança biparental ou paterna (ZHANG e SODMERGEN, 2010). Já as gimnospermas possuem majoritariamente herança paterna (NEALE, MARSHALL e SEDEROFF, 1989). Era esperado que essa ausência de recombinação resultasse no acúmulo de mutações deletérias, em um processo conhecido como Catraca de Muller. No entanto, as taxas de mutação encontradas no genoma plastidial são menores do que no genoma nuclear (GREINER e BOCK, 2013).

Todo e qualquer organismo apresenta um sistema de reparo que mantém a estabilidade do genoma em um grau que previne excessivas mutações, mas sem impedir a evolução (MARÉCHAL e BRISSON, 2010). Por um lado, em nível de um único organismo, parece preferível manter o genoma absolutamente intacto. Por outro lado, certo grau de instabilidade é crucial para a evolução das espécies, eliminando completamente os mecanismos que produzem essa diversidade, também se destrói a resiliência e capacidade de adaptação das espécies às mudanças das condições ambientais (MARECHAL e BRISSON, 2010). No entanto, os cloroplastos mantém um mecanismo de reparo que ainda não foi completamente esclarecido (GREINER e BOCK, 2013).

Considerou-se por muito tempo que a estrutura do genoma plastidial seria exclusivamente circular, devido à sua origem procariótica (BOCK, 2007; WICKE *et al.*, 2011). No entanto, estudos recentes de hibridização *in situ* demonstraram que apenas uma pequena parte das moléculas ocorre de forma circular, ocorrendo em sua maioria de forma concatenada de duas ou mais moléculas, em forma circular ou linear (WICKE *et al.*, 2011).

No âmbito da transformação genética, o sequenciamento completo do genoma plastidial é essencial para o desenvolvimento de biotecnologias ligadas à transformação genética de cloroplastos (CLARKE, DANIELL e NUGENT 2011). As regiões intergênicas dos genomas plastidiais não são bem conservadas e o desenho eficaz do vetor para a transformação de

novas espécies exige dados da sequência do genoma plastidial específico de cada espécie (CLARKE, DANIELL e NUGENT 2011).

Há vantagens consideráveis associadas com a transformação gênica de cloroplastos em vez de nuclear. A primeira é o elevado número de plastídios e o elevado número de cópias no genoma do plastídio por célula, oferecendo níveis de expressão do transgene extraordinariamente altos. A segunda é que a integração do transgene no genoma do cloroplasto ocorre exclusivamente por recombinação homóloga, tornando genoma plastidial uma técnica de engenharia genética de alta precisão para as plantas. A terceira é que como um sistema procarioto, o sistema genético do cloroplasto é desprovido de silenciamento gênico e de outros mecanismos que interferem epigeneticamente com a expressão estável do transgene. A quarta é que da mesma forma que o genoma bacteriano, muitos genes plastidiais são dispostos em *operons*, oferecendo a possibilidade de arranjar genes em *operons* artificiais. Finalmente, a transformação plastidial tem recebido atenção significativa como uma ferramenta excelente para contenção do transgene devido ao modo de herança materna do plastídio característica da maioria das espécies de angiospermas, o que reduz drasticamente a segregação do transgene através de pólen (BOCK, 2014).

As tecnologias de transformação plastidial têm sido intensamente utilizadas em genômica funcional, através da realização de nocautes de genes dirigidos ao cloroplasto. Estes estudos têm contribuído enormemente para a compreensão dos processos fisiológicos e bioquímicos dentro do compartimento plastidial (BOCK, 2001). A tecnologia de transformação de cloroplastos também tem potencial para melhoramento de plantas, desenvolvimento de plantas como biorreatores para a produção sustentável e de baixo custo de produtos biofarmacêuticos, enzimas e matérias-primas para a indústria química (BOCK, 2014).

O genoma plastidial tem se mostrado, desde a década de 80, uma ferramenta viável para realização de estudos filogenéticos (PALMER, 1985). Em virtude da sua fácil amplificação, tamanho pequeno, baixa taxa evolutiva e da maior conservação observada entre os genomas em evolução conhecidos, o genoma plastidial foi considerado adequado para estudos filogenéticos em diferentes níveis taxonômicos (PALMER, 1985).

O gene *rbcL*, que codifica a grande subunidade da ribulose 1,5-bifosfato-carboxilase/oxigenase (RUBISCO), começou a ser amplamente sequenciado a partir de um grande número de espécies, gerando uma boa base de dados para estudos de filogenia de plantas (PALMER *et al.*, 1988). Estudos filogenéticos em nível de família e de taxas maiores foram

realizados com sucesso em gimnospermas (HASEBE *et al.*, 1992; BRUNSFELD *et al.*, 1994; SETOGUCHI *et al.*, 1998).

No entanto, em algumas situações as relações continuavam não claras pelo fato do gene *rbcL* ser muito conservado para esclarecer dúvidas entre gêneros próximos (GIELLY e TABERLET, 1994). A análise de regiões não-codificantes do cpDNA, como introns e espaçadores intergênicos, foi a estratégia encontrada para esclarecer as relações em níveis taxonômicos mais baixos. Estas zonas 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 e 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 cloroplasto poderiam trazer informações adicionais a essa linha de estudo (SHAW *et al.*, 2007). A partir do sequenciamento completo do genoma plastidial pode-se encontrar regiões com maior número de caracteres para estudos filogenéticos em menores taxa, assim como as regiões com menor número (SHAW *et al.*, 2007).

Quando apenas um gene é investigado, faz-se uma amostragem de um segmento do genoma; comparando mais genes reduz-se o erro de amostragem inerente à amostragem de apenas um ou poucos genes; comparando genomas completos se pode descobrir o padrão de sítios que estão disponíveis para comparação (MARTIN *et al.*, 2005). Além disso, SHAW *et al.* (2007) aponta que não há uma região única ou um conjunto de regiões mais indicados para todas as linhagens taxonômicas, e sim que deve ser feito um rastreamento em cada linhagem para determinar quais as regiões mais adequadas.

Dessa forma, há a possibilidade de se encontrar um grupo de regiões com características em que: (i) os resultados da análise usando este conjunto de regiões é semelhante ao dos resultados da análise multigênica usando o número máximo de regiões, (ii) a região deve ser compacta e ser significativamente menor do que o conjunto de multigenes (LOGACHEVA *et al.*, 2007).

Análises filogenéticas a partir do sequenciamento completo do genoma plastidial foram realizadas com sucesso em *Vitis* (JANSEN *et al.*, 2006), fornecendo forte apoio ao posicionamento de Vitaceae como a primeira linhagem divergente das rosídeas. Outros estudos, como na família Poaceae, entre as subfamílias Panicoideae, Bambusoideae e Pooideae, revelou uma maior proximidade entre Bambusoideae e Pooideae (MATSUOKA *et al.*, 2002).

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 regiões invertidas repetidas (IR_A e IR_B) (Figura 3). As duas IR são idênticas na composição, dessa forma, todos os genes contidos nessas regiões apresentam duas cópias no genoma, porém em sentido de leitura contrário (Figura 4).

A função dessas regiões IR permanece obscura. Teorias como o aumento na quantidade de genes altamente expressos, como genes que codificam para rRNA e a estabilização do genoma foram propostas (PALMER e THOMPSON, 1982).

Uma evidência circunstancial para a ação da conversão gênica nas IR 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 (WOLFE, LI e SHARP, 1987; MAIER *et al.*, 1995). Essas regiões IR estão frequentemente sujeitas a expansão, contração e até mesmo perda completa (WICKE *et al.*, 2011), eventos esses que influenciam no tamanho do genoma plastidial (WU e CHAW, 2013).

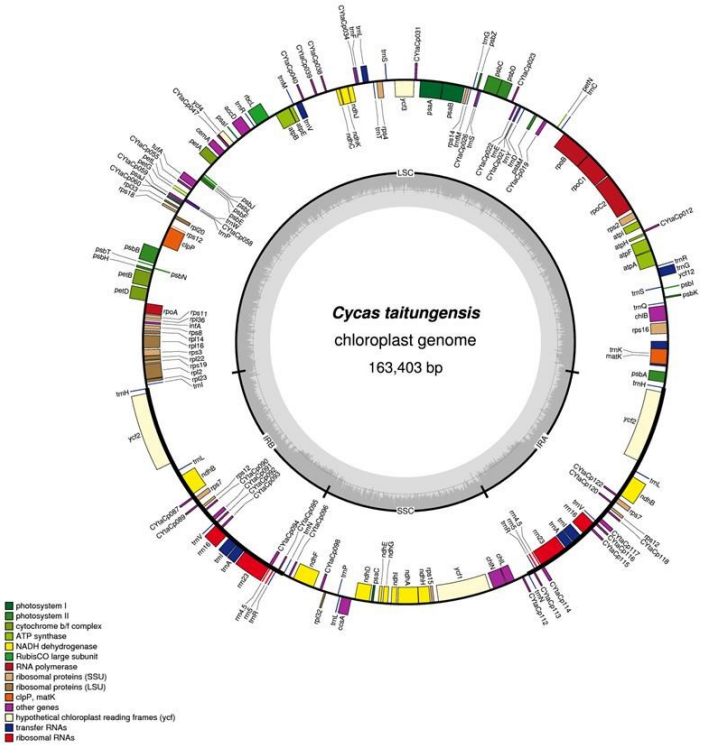

Figura 4. Mapa genético do genoma plastidial de *Cycas taitungensis*. Genes mostrados no lado de dentro e do lado de fora do círculo grande são transcritos no sentido horário e anti-horário, respectivamente (WU *et al.,* 2007).

Algumas espécies de gimnospermas com seu genoma plastidial sequenciado perderam umas dessas regiões IR, como é o caso da *Cryptomeria japonica*, *Cephalotaxus oliveri*, *Cephalotaxus wilsoniana*, *Taiwania cryptomerioides* e algumas Pinaceaes (CRONN *et al.*, 2008; HIRAO *et al.*, 2008; WU *et al.*, 2011; YI *et al.*, 2013). Até hoje, não está inteiramente claro se essas espécies perderam diferentes cópias das regiões IR (IR_a ou IR_b). Por um lado, WU *et al.,* (2011) e WU e CHAW, 2013 alegam que Pinaceas e Cupressophytas perderam diferentes cópias das regiões IR, enquanto YI *et al.,* (2013) demonstra que não há dados suficientes para afirmar nem que sim, nem que não.

CAPÍTULO 1
Manuscrito publicado no periódico *PLoS ONE*.

An improved protocol for intact chloroplasts and cpDNA isolation in conifers

Vieira LN, Faoro H, Fraga HPFF, Rogalski M, Souza EM, Pedrosa FO, Nodari RO, Guerra MP. 2014. An improved protocol for intact chloroplasts and cpDNA isolation in conifers. PLoS ONE. 9(1): e84792.

Abstract

Background: Performing chloroplast DNA (cpDNA) isolation is considered a major challenge among different plant groups, especially conifers. Isolating chloroplasts in conifers by such conventional methods as sucrose gradient and high salt has not been successful. So far, plastid genome sequencing protocols for conifer species have been based mainly on long-range PCR, which is known to be time-consuming and difficult to implement.
Methodology/Principal findings: We developed a protocol for cpDNA isolation using three different conifer families: *Araucaria angustifolia* and *Araucaria bidwilli* (Araucariaceae), *Podocarpus lambertii* (Podocarpaceae) and *Pinus patula* (Pinaceae). The present protocol is based on high salt isolation buffer followed by saline Percoll gradient. Combining these two strategies allowed enhanced chloroplast isolation, along with decreased contamination caused by polysaccharides, polyphenols, proteins, and nuclear DNA in cpDNA. Microscopy images confirmed the presence of intact chloroplasts in high abundance. This method was applied to cpDNA isolation and subsequent sequencing by Illumina MiSeq (2 x 250 bp), using only 50 ng of cpDNA. Reference-guided chloroplast genome mapping showed that high average coverage was achieved for all evaluated species: 24.63 for *A. angustifolia*, 135.97 for *A. bidwilli*, 1196.10 for *P. lambertii*, and 64.68 for *P. patula*.
Conclusion: Results show that this improved protocol is suitable for enhanced quality and yield of chloroplasts and cpDNA isolation from conifers, providing a useful tool for studies that require isolated chloroplasts and/or whole cpDNA sequences.

Keywords: chloroplast; cpDNA; organelle isolation; Percoll gradient; conifers; plants

1 Introduction

The chloroplast genome of land plants usually harbors a conserved set of approximately 120 genes in a 120-160 kb pair genome, out of a genome of some 3,200 genes present in their cyanobacterial ancestor (KANEKO *et al.,* 1996). Land plant plastomes are mostly conserved and present little variation in size and gene content, ranging from 70,028 nucleotides and 25 protein coding genes in the nonphotosynthetic parasitic plant, *Epifagus virginiana* (WOLFE *et al.,* 1992), to 217,942 nucleotides and 131 protein coding genes in *Pelargonium* x *Hortorum* (CHUMLEY *et al.,* 2006). Although chloroplast genomes contain highly conserved essential genes for plant growth and development, they also contain variable regions, i.e., intergenic regions and structural variations. In addition, they contain one of the few sets of characters that can transcend the life history of green plants and, hence, generate important evolutionary information. Therefore, chloroplast genome sequences can be used for comparative evolutionary studies within and between different groups of plants (TIMMIS *et al.,* 2004;WOLF *et al.,* 2010; GREINER *et al.,* 2013), as demonstrated by several works (JANSEN *et al.,* 2007; MOORE *et al.,* 2007; MOORE *et al.,* 2010; WU *et al.,* 2011; YI *et al.,* 2013). Furthermore, chloroplast genome sequences have been used to investigate gene function (ROGALSKI *et al.,* 2008; ALKATIB *et al.,* 2012), and they have been targeted for biotechnological applications (CLARKE, DANIELL E NUGENT*,* 2011; MALIGA E BOCK, 2011; ROGALSKI E CARRER, 2011; BOCK, 2013). Based on the importance of land plant chloroplast DNA (cpDNA) in plant genetics, evolution and biotechnology, it has been a target in many plant genome sequencing projects (WU *et al.*, 2007; HIRAO *et al.*, 2008; LIN *et al.*, 2010). To date, complete cpDNAs of more than 300 plants have been sequenced (ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=2759&opt=plastid).

With rapid progress in sequencing technologies, chloroplast genome sequencing can be realized quickly as a result of small size and structural simplicity when compared to nuclear genomes. However, chloroplast genome sequences have been determined for only a very few families belonging to gymnosperms (WU *et al.*, 2007; WERNER *et al.,* 2009; LIN *et al.,* 2010). Especially for conifers, chloroplast genome sequences are available for families that include Cephalotaxaceae (YI *et al.,* 2013), Cupressaceae (HIRAO *et al.,* 2008), Pinaceae (WAKASUGI *et al.,* 1994; CRONN *et al.,* 2008; LIN *et al.,* 2010), Podocarpaceae (database accession

no. NC_020361.1) and Taxaceae (database accession no. NC_020321.1), but not the Araucariaceae family.

Chloroplast DNA isolation has been a major challenge, hindering widespread applications in different plant groups. Chloroplast isolation in conifers by such conventional methods as sucrose gradient (PALMER E SEIN, 1986) and high salt (BOOKJANS, STUMMANN E HENNINGSEN, 1984) has, thus far, not been successful. This most likely results from the high volume of contaminants, including polyphenols, oleoresins, terpenoids and polysaccharides, present in conifer needles, making difficult the acquisition of intact isolated chloroplasts and high quality cpDNA (KEELING E BOHLMANN, 2006). For conifers, whole cpDNA sequencing protocols have been based on total DNA isolation, followed by cpDNA fragments amplification by use of polymerase chain reaction (PCR) with degenerate primers (CRONN *et al.,* 2008; LIN *et al.,* 2010; WU *et al.,* 2011; YI *et al.,* 2013). However, this strategy is known to be time-consuming and difficult to implement because of differences in gene organization among different plant species (ATHERTON *et al.,* 2010) and "promiscuous" cpDNA present in the nucleus and mitochondrial genome (AYLIFFE E TIMMIS, 1992; AYLIFFE, SCOTT E TIMMIS, 1998; GOREMYKIN *et al.,* 2009; ROUSSEAU-GUEUTIN, AYLIFFE E TIMMIS, 2011).

Therefore, the overall aim of the present work was to develop an efficient protocol for chloroplast isolation and subsequent high quality cpDNA extraction in conifers, using three different conifer families: *Araucaria angustifolia* and *Araucaria bidwilli* (Araucariaceae), *Podocarpus lambertii* (Podocarpaceae) and *Pinus patula* (Pinaceae).

2 Material and methods

2.1 Plant material

Local *A. angustifolia* and *P. patula* seeds were purchased and germinated in the greenhouse of Federal University of Santa Catarina, Brazil. Needles were collected from 6 months plants; this procedure does not require authorization. *P. lambertii* young plants (n=10) were collected at a private area, located at Lages, Santa Catarina, Brazil (27° 48' 57" S, 50° 19' 33" W), where the species is abundant, with the previous owner permission (José Antônio Ribas Ribeiro). This species are not considered under threat. After, the young plants were transplanted to greenhouse and maintained under this condition until the collection of needles. *A. bidwilli* young needles were collected at Botanical Garden, authorized by Federal

University of Santa Catarina, Brazil. For each plant species, 25 g of fresh young needles were collected and stored in 4ºC refrigerator for further chloroplast extractions.

2.2 Protocols

The three chloroplast DNA isolation methods used here are described as follows:

A. High salt plus saline Percoll gradient method (Figure 1)
All the following steps were carried out at 0ºC, if not otherwise stated.
1. Prior to extraction, 25 g (fresh weight) of young needles were collected and kept in dark for 10 days at 4ºC to decrease starch and resin level. Fresh needles were cleaned with 0.5% sarkosyl (Fluka, Ronkonkoma, NY) for 5 min to reduce microbial contamination and then washed 4 times with distilled water.
2. Needles were homogenized in 400 ml ice-cold isolation buffer (Table 1) for 30 s in a pre-chilled blender. Homogenate was filtered primarily into two layers of gauze bandage and then filtered again using two layers of Miracloth by softly squeezing the cloth.
3. Homogenate was centrifuged at 200 g for 15 min at 4ºC. The nucleus pellet and cell-wall debris were discarded. The supernatant included chloroplasts suspended in it.
4. The supernatant was centrifuged at the higher centrifugal force of 3000 g for 20 min at 4ºC, resulting in a chloroplast pellet with some contamination.
5. The pellet was gently resuspended in 12 ml of wash buffer (Table 1) using a paintbrush.
6. Homogenate was divided into 6 tubes (50 ml), each containing 20 ml Percoll (GE Healthcare, Uppsala, Sweden) gradient (70%-30%) and then centrifuged at 5000 g for 25 min at 4ºC. The interface 70%-30% containing chloroplasts was collected.
7. Collected interface containing chloroplasts was washed twice with 100 ml of wash buffer and centrifuged at 3000 g for 20 min at 4ºC to obtain the purified chloroplast pellet.

B) Modified high salt method (SHI et al., 2012)
All the following steps were carried out at 0ºC, if not otherwise stated.
1. Prior to extraction, 25 g (fresh weight) of young needles were collected and kept in dark for 72 h at 4ºC to decrease starch level stored in the needles. Fresh needles were cleaned with distilled water.

2. Needles were homogenized in 400 ml of isolation buffer (Table 1) for 30 s. Homogenate was filtered into centrifuge bottles, using two layers of Miracloth (Calbiochem, San Diego, CA) by softly squeezing the cloth.
3. The homogenate was centrifuged twice at 200 g for 20 min at 4ºC. The nucleus pellet and cell-wall debris were discarded. Supernatant included chloroplasts suspended in it.
4. The supernatant was submitted to a higher centrifugal force (3500 g) for 20 min at 4ºC, resulting in a chloroplast pellet contaminated with some nuclear DNA.
5. The pellet was gently resuspended in 250 ml of wash buffer (Table 1), using a paintbrush to wash the nuclear DNA attached to the chloroplast membrane, followed by centrifugation at 3500 g for 20 min at 4º C. The supernatant was discarded.
6. The pellet was resuspended again with 250 ml wash buffer and centrifuged at 3500 g for 20 min at 4ºC to obtain the final chloroplast pellet.

C) Sucrose gradient method (JANSEN *et al.*, 2005)
1. Prior to extraction, about 25 g (fresh weight) of young needles were collected and kept in dark for 72 h at 4ºC in order to decrease the starch level stored in the leaves. Fresh needles were cleaned with distilled water.
2. Needles were homogenized in 400 ml of ice-cold isolation buffer (Table 1) for 30 s. The homogenate was filtered into centrifuge bottles using two layers of Miracloth by softly squeezing the cloth.
3. The homogenate was centrifuged at 200 g for 15 min at 4ºC. The nucleus pellet and cell-wall debris were discarded. The supernatant included chloroplasts suspended in it.
4. The supernatant was centrifuged at a higher centrifugal force (2000 g) for 20 min at 4º C, and the resulting chloroplast pellet showed some contamination.
5. The pellet was resuspended in 7 ml of ice-cold wash buffer (Table 1), using a soft paintbrush.
6. The homogenate was gently loaded into 6 tubes (50 ml) containing sucrose step gradient consisting of 18 ml of 52% sucrose and overlaid with 7 ml of 30% sucrose.
7. Step gradients were centrifuged at 3500 g for 60 min at 4º C.
8. The band from the 30–52% interface containing chloroplasts was collected, diluted twice with 200 ml of wash buffer, and centrifuged at 1500 g for 15 min at 4ºC to gain the purified chloroplast pellet.

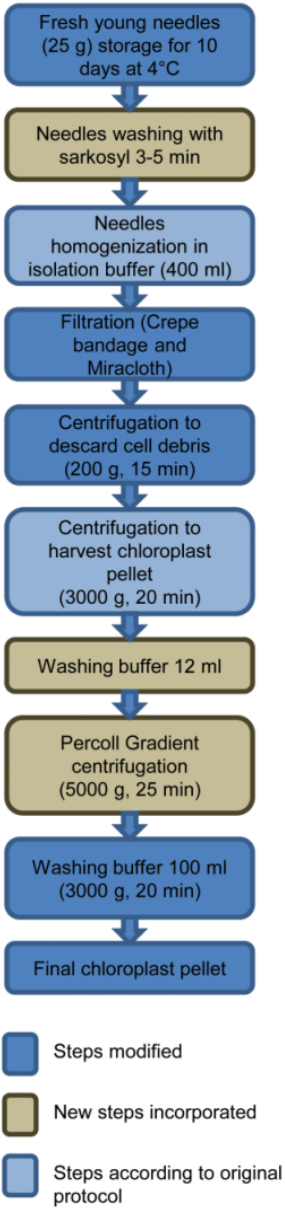

Figure 1. Flowchart showing the major steps for chloroplast isolation according to high salt plus saline Percoll.

Table 1. Composition of chloroplast isolation buffers and wash buffers for modified high salt method, high salt plus saline Percoll method and sucrose gradient method.

High salt plus saline Percoll method	Modified high salt method	Sucrose gradient
Isolation Buffer (pH 3.8)	**Isolation Buffer (pH 3.8)**	**Isolation Buffer**
1.25 M NaCl	1.25 M NaCl	50 mM Tris-HCl (pH 8.0)
0.25 M ascorbic acid	0.25 M ascorbic acid	0.35 M sorbitol
10 mM sodium metabisulfite	10 mM sodium metabisulfite	7 mM EDTA
0.0125 M Borax	0.0125 M Borax	0.1% 2-mercaptoethanol
50 mM Tris-HCl (pH 8.0)	50 mM Tris-HCl (pH 8.0)	0.1% BSA
7 mM EDTA	7 mM EDTA	
1% PVP-40 (w/v)	1% PVP-40 (w/v)	
0.1% BSA (w/v)	0.1% BSA (w/v)	
	1 mM DTT	
Wash Buffer (pH 8.0)	**Wash Buffer (pH 8.0)**	**Wash Buffer**
1.25 M NaCl	1.25 M NaCl	50 mM Tris-HCl (pH 8.0)
0.0125 M Borax	0.0125 M Borax	0.35 M sorbitol
50 mM Tris-HCl (pH 8.0)	50 mM Tris-HCl (pH 8.0)	25 mM EDTA
25 mM EDTA	25 mM EDTA	
1% PVP-40 (w/v)	1% PVP-40 (w/v)	
0.1% BSA (w/v)	0.1% BSA (w/v)	
	1 mM DTT	

Both BSA and DTT were added just before the start of the experiment.
Percoll gradient solutions consisted of wash buffer with Percoll at a final concentration of 70% (v/v) and 30% (v/v).
Sucrose gradient solutions consisted of 50 mM Tris-HCl (pH 8.0), 25 mM EDTA and sucrose addition for a final concentration of 52% sucrose (w/v) and 30% (w/v) sucrose.

2.3 Chloroplast DNA isolation

Chloroplast DNA isolation was the same for all chloroplast pellets obtained using the three different isolation methods. DNA isolation buffer consisted of 100 mM NaCl, 100 mM Tris-HCl (pH 8.0), 50 mM EDTA, and 1 mM DTT.

1. The chloroplast lyse was obtained by incubating the chloroplast pellet with 8 ml of DNA isolation buffer, 1.5 ml 20% SDS, 20 µl 2-Mercaptoethanol and 30 µl Proteinase K (10 mg/ml) into a centrifuge tube at 55ºC for 4 h.
2. The centrifuge tube was incubated on ice for 5 min, and then 1.5 ml 5 M KAc (pH 5.2) was added to the lyse mixture and chilled for more than 30

min. After that, the tube was centrifuged at 10000 g for 15 min at 4°C, and the pellet was discarded.
3. The supernatant was extracted with an equal volume of saturated phenol and chloroform:isoamyl-alcohol (24:1) and centrifuged twice at 10000 g for 20 min.
4. An equal volume of isopropyl alcohol (about 10 ml) was added to the upper aqueous phase and incubated at -20ºC overnight.
5. To obtain the DNA pellet, the tube was centrifuged at 10000 g for 20 min at 4 ºC. The cpDNA pellet was washed with 70% and 96% ethanol, air dried, and redissolved in 50 µl TE buffer.
6. The cpDNA samples were treated with RNAse, and the DNA band was visualized on a 0.7% agarose gel.
7. DNA purity and concentration were evaluated with Nanodrop®, based on 260/280 and 260/230 ratios.

2.4 Microscopy analysis

The integrity of isolated chloroplasts was assessed with a phase-contrast light microscopy using an inverted Olympus IX81 microscope (WALKER, CEROVIC E ROBINSON, 1987). Intact chloroplasts were considered those with pale yellow-green color and refractive, with a bright halo appearance around each plastid, whereas broken chloroplasts were those with a dark green, granular, and non-refractive appearance (WALKER, CEROVIC E ROBINSON, 1987).

2.5 Chloroplast genome sequencing

A. angustifolia, *A. bidwilli*, *P. patula* and *P. lambertii* cpDNAs were isolated using the high salt plus Percoll method. For each species, approximately 50 ng of DNA were prepared with the Nextera DNA Sample Prep Kit (Illumina, San Diego, USA) according to the manufacturer's instructions. Chloroplast DNAs were sequenced using Illumina MiSeq (2 x 250 read length) at the Federal University of Paraná - Brazil. The obtained paired-end reads were assembled to reference genome sequence and estimate genome coverage, using the CLC Genomics Workbench 5.5 software. The reference chloroplast genome sequences of *Podocarpus totara* (NC_020361.1) and *Pinus thunbergii* (NC_001631.1) were downloaded from GenBank.

3 Results and Discussion

3.1 Chloroplast isolation

Chloroplast isolation protocols are generally based on methods that employ high salt concentration buffers (SHI *et al.,* 2012), sucrose density gradient (JANSEN *et al.,* 2005), high salt buffers followed by sucrose gradient (DIEKMANN *et al.,* 2008) and high sorbitol concentration buffers followed by Percoll gradient (KUBIS, LILLEY E JARVIS, 2008). As the purity of intact chloroplasts is one of the critical steps of whole sequencing, a previous paper (HIRAO *et al.,* 2008) considered the use of sucrose density gradients as the best method for separating nuclear DNA contamination from cpDNA. The present protocol is based on a high salt isolation buffer followed by saline Percoll gradient. The combination of these two strategies provided two advantages: better isolation of chloroplasts by use of the Percoll gradient and decreased contamination by polysaccharides, polyphenols, and proteins.

The first significant change in isolation protocol was the increase in storage time to 10 days at 4°C prior to extraction. This change led to a significant reduction in the viscosity of extraction buffer, possibly caused by a decrease in the polysaccharides and oleoresin concentrations. A similar strategy has been used in other protocols, in which 48-72 h at 4°C was enough to reduce the stored polysaccharides (JANSEN *et al.,* 2005; SHI *et al.,* 2012). However, because of the high amount of oleoresins and thick outer periclinal walls in conifers (MASTROBERTI E MARIATH, 2003; YAMAMOTO, OTTO E SIMONEIT, 2004), a longer time is required to reduce the concentration of these compounds.

Subsequently, needles were washed with sarkosyl, reducing material contamination. At the time of homogenization, it was observed that the high salt buffer was also responsible for the decrease in viscosity of the solution. The viscosity normally found in the homogenate (conifer needles and isolation buffer) is related to the high amount of resins and polysaccharides present in conifer needles, and its reduction enables faster and more efficient filtration, with lower material loss. In a previous paper (ARIF *et al.,* 2010), it was observed that the use of high salt buffers for DNA extraction increased the quality and yield of DNA extracted from plant tissues rich in polysaccharides.

Similarly, the initial centrifugation step, performed to pellet cellular debris was reduced to 200 g. We also reduced the initial centrifugation step at 200 g to only one centrifugation, while in the modified high salt method (SHI *et al.,* 2012), it was performed twice. This reduction increased

chloroplast yield and did not entail any reduction in the quality of isolated chloroplast as a result of the Percoll gradient step. The major steps for chloroplast isolation using the high salt plus Percoll method, including original, new and modified steps, are summarized in Figure 1.

Microscopy images showed the presence of some intact chloroplasts and a large amount of broken chloroplasts and other cellular debris in extraction when using the modified high salt and sucrose methods (Fig. 2A, B). On the other hand, the high salt plus saline Percoll method resulted in the presence of abundant intact chloroplasts (Fig. 2C).

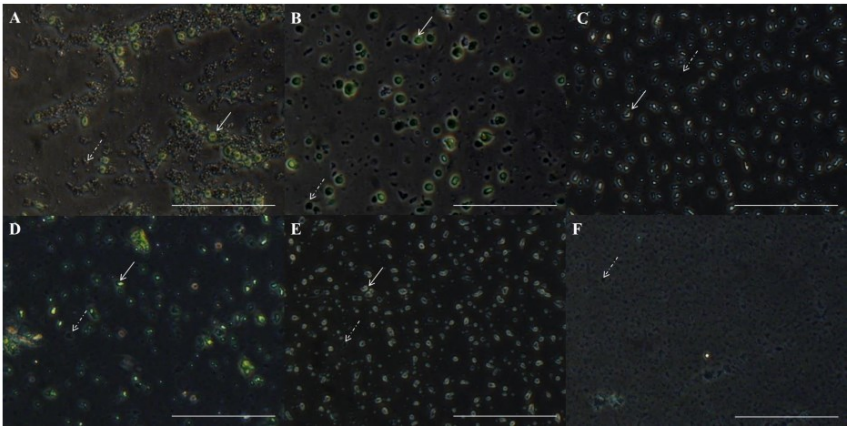

Figure 2. Chloroplast visualization of *Araucaria angustifolia* in phase contrast microscopy. (A) Chloroplasts isolated with improved high salt method; (B) Chloroplasts isolated with sucrose method; (C) Chloroplasts isolated with high salt plus Percoll method; (D-F) Micrographs during chloroplast isolation with high salt plus saline Percoll method; (D) Broken and intact chloroplasts before Percoll gradient centrifugation; (E) Intact isolated chloroplasts in interface 70 / 30% after Percoll gradient centrifugation; (F) Broken chloroplasts in upper 30% phase after Percoll gradient centrifugation. Dotted arrows indicate broken chloroplasts. Solid arrows indicate intact chloroplasts. Bar – 50 μM.

Aiming to better characterize the efficiency of Percoll gradient, microscopy analysis was performed prior to the Percoll gradient centrifugation step, and the presence of both intact and ruptured chloroplasts could be observed (Fig. 2D). However, after centrifugation in Percoll gradient, the 70%/30% interface (Fig. 2E) contains abundant intact chloroplasts and only a few broken chloroplasts, while the upper phase contains many broken chloroplasts (Fig. 2F). In addition, below the 70%

gradient, the formation of a white colored pellet composed of polysaccharides and other contaminants could be seen. Despite the high purity of chloroplasts observed immediately after Percoll gradient centrifugation, two subsequent centrifugations are essential to remove any residue of Percoll. The isolation of cpDNA without performing these two washes would be greatly affected by its presence. It is noteworthy that changes in the protocol enabled the isolation of the best quality chloroplasts, without the need of ultracentrifugation, which could be a limiting point in the procedure.

In addition to facilitating whole cpDNA sequencing, isolation of intact chloroplasts can also be applied in plastid proteome characterization studies. Comparative proteomics in *Triticum aestivum* and *Arabidopsis thaliana* chloroplasts have been recently developed using intact isolated chloroplasts. The results demonstrated that the quality of chloroplast isolation is a fundamental step of complete proteome characterization (GARGANO *et al.,* 2013; HE *et al.,* 2013).

3.2 Chloroplast DNA isolation

We also obtained isolated cpDNA with better quality and yield using this high salt plus saline Percoll method. Using the modified high salt and sucrose methods, bands in agarose gel revealed the presence of degraded DNA, indicating contamination with nuclear DNA and polysaccharides (Fig. 3A, B, respectively), while isolated cpDNA formed a well-defined band, which is indicative of high purity and polysaccharide-free cpDNA (Fig. 3C).

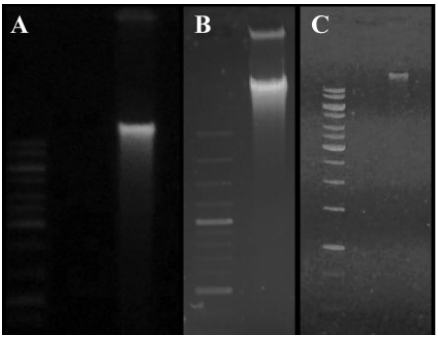

Figure 3. Chloroplast DNA visualization of *Araucaria angustifolia* in 0.7% agarose gel stained with ethidium bromide. (A) Ladder 1 kb and cpDNA isolated with modified high salt method; (B) Ladder 1 kb and cpDNA isolated with sucrose

method; (C) Ladder 1 kb and cpDNA isolated with high salt plus saline Percoll method.

In addition, Nanodrop evaluation indicated higher cpDNA yield with the high salt plus saline Percoll method, about 3 times higher when compared to high salt methods and almost 9 times higher when compared to the sucrose method. Enhanced 260/280 and 260/230 ratios were observed in the high salt plus saline Percoll method, 2.05 and 1.99, respectively (Table 2).

Table 2. cpDNA from different isolation methods in *Araucaria angustifolia* sample.

Isolation Method	DNA concentration (ng/µl)	260/280	260/230
Modified high salt method	975.6	1.66	0.92
High Salt plus saline Percoll method	3438.3	2.05	1.99
Sucrose Gradient Method	472.4	1.52	0.47

Ratios evaluated with Nanodrop®, in a final volume of 40 µl.

These ratios indicate a high purity of isolated cpDNA, which is a prerequisite for whole chloroplast sequencing. In the two other methods evaluated, contamination was observed with polyphenols and polysaccharides (Table 2). Moreover, when we used the sucrose method, a highly contaminated and oxidized DNA pellet was obtained. Taken together, we considered the high salt plus saline Percoll protocol as having the best yield and quality for cpDNA isolation from *A. angustifolia*. Thus, this method was applied to cpDNA isolation of *A. bidwilli*, *P. patula* and *P. lambertii*. As expected, a high quality in cpDNA isolated from all evaluated species was realized at 260/280> 1.95 and 260/230> 1.74 ratios (Table 3). All species showed cpDNA yield similar to *A. angustifolia,* with the exception of *P. lambertii.* However, even its cpDNA yield was sufficient for sequencing (Table 3).

Table 3. cpDNA ratios of selected conifers evaluated using Nanodrop®.

Plant species	DNA concentration (ng/µl)	260/280	260/230
Araucaria angustifolia	3438.3	2.05	1.99
Araucaria bidwilli	3038.0	1.95	1.74
Podocarpus lambertii	430.6	2.01	1.89
Pinus patula	1799.5	2.05	2.10

Samples were isolated with the high salt plus Percoll method in different conifer species. Final volume of 40 µl.

3.3 Chloroplast genome sequencing

Improving technologies have made DNA sequencing faster, more accurate and far cheaper, creating opportunities to sequence the whole chloroplast genome in order to perform evolutionary and phylogenomic studies. To test the quality of cpDNA isolated by our new method, we sequenced the plastid genome of four conifer species (*A. angustifolia*, *A. bidwilli*, *P. patula* and *P. lambertii*) using the Illumina sequencing technology.

To estimate the efficiency of chloroplast genome sequence assembly with our cpDNA isolation protocol, we sequenced these four chloroplast genomes by using MiSeq Illumina sequencing with only 50 ng of cpDNA. In other sequencing protocols, about 5-10 µg (JANSEN *et al.,* 2005; SHI *et al.,* 2012) were used for sequencing, thereby increasing the amount of plant material required for isolation and often limiting the use of the technique. A reference-guided chloroplast genome mapping was performed to estimate the genome average coverage (Figure 4). The cpDNA sequencing generated a high average coverage for all species evaluated: 24.63 for *A. angustifolia*, 135.97 for *A. bidwilli*, 1,196.10 for *P. lambertii*, and 64.68 for *P. patula* (Table 4). Thus, in this study, all of the reference genomes were sufficiently covered for assembly.

This protocol presents higher genome coverage when compared to protocols recently applied to conifers chloroplast genome sequencing, as those using total DNA followed by PCR amplification with degenerated primers that resulted in genome coverage only about 8-fold (LIN *et al.,* 2010; YI *et al.,* 2013). Furthermore, this strategy is time-consuming and difficult to implement because of differences in gene organization among different plant species. *Cryptomeria japonica* cp genome was sequenced using sucrose gradient method, followed by DNA isolation with phenol/chloroform, DNA purification with DNeasy Plant Mini Kit (QIAGEN) and ATP-dependent DNase (TOYOBO) (HIRAO *et al.,* 2008).

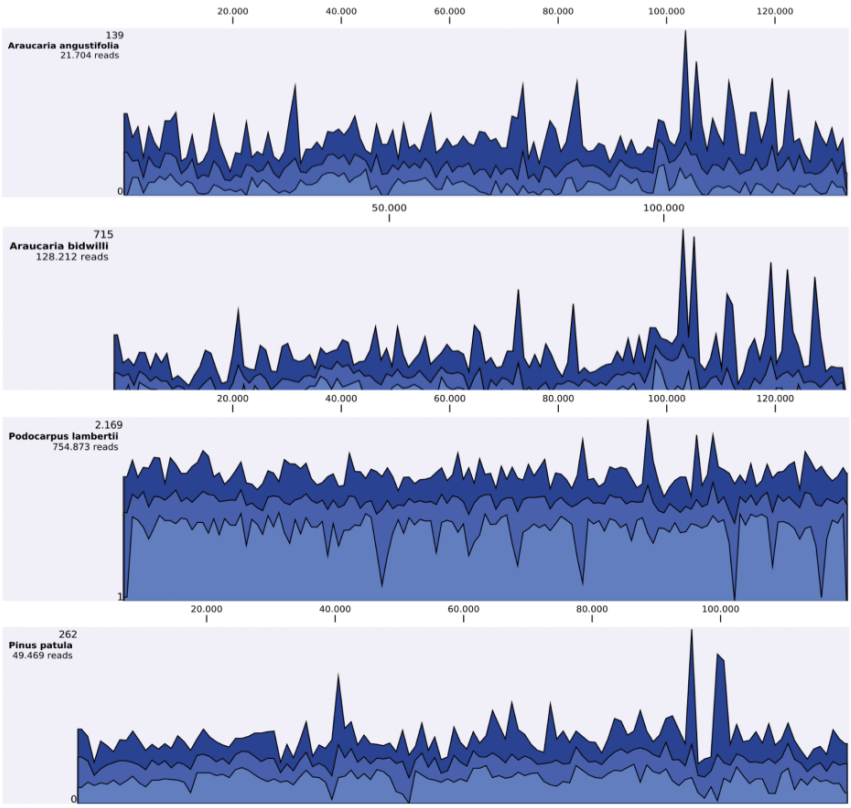

Figure 4. Reference graph track showing observed coverage values. Different colors show the minimum (light blue), mean (blue), and maximum (dark blue) observed coverage values for all genomic regions (data aggregation above 100 bp). *Araucaria angustifolia*, *Araucaria bidwilli*, and *Podocarpus lambertii* sequence reads were mapped on *Podocarpus totara*; *Pinus patula* sequence reads were mapped on *Pinus thunbergii*.

As shown in the present work, the protocol based on saline buffer followed by Percoll gradient results in higher quality DNA than sucrose gradient. Moreover, all these purification steps applied to the isolated DNA, such as the utilization of ATP-dependent DNase, led to a lower DNA yield (SHI *et al.,* 2012).

Table 4. Average coverage of cpDNA evaluated from selected conifers with CLC Genomics Workbench 5.5 software.

Plant Species	Average Coverage	Reference Genome
Araucaria angustifolia	24.63	*Podocarpus totara*
Araucaria bidwilli	135.97	*Podocarpus totara*
Podocarpus lambertii	1196.10	*Podocarpus totara*
Pinus patula	64.68	*Pinus thunbergii*

cpDNA reads were mapped to reference genomes.

In summary, the results obtained in the present work show that these improvements in the general protocol for chloroplasts and cpDNA isolation in conifers enhance the overall quality and yield of chloroplasts and cpDNA isolation, providing a useful tool for studies that require isolated chloroplasts and/or plastid genome sequence. Facilitating chloroplast sequencing of this species group and, hence, increasing the amount of information about the plastid genome of conifers may, in turn, lead to greater understanding about plant evolution, as well as the structural and functional genomics in plants other than conifers.

Financial Disclosure

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 funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

CAPÍTULO 2
Manuscrito formatado de acordo com as normas do periódico *PLoS ONE*.

The complete chloroplast genome sequence of *Podocarpus lambertii:* genome structure, evolutionary aspects, gene content and SSR detection

Vieira LN, Faoro H, Rogalski M, Fraga HPF, Cardoso RLA, Souza EM, Pedrosa FO, Nodari RO, Guerra MP. The complete chloroplast genome sequence of *Podocarpus lambertii:* genome structure, evolutionary aspects, gene content and SSR detection.

Abstract

Background: *Podocarpus lambertii* (Podocarpaceae) is a native conifer from the Brazilian Atlantic Forest Biome, which is considered one of the 25 biodiversity hotspots in the world. The advancement of next-generation sequencing technologies has enabled the rapid acquisition of whole chloroplast (cp) genome sequences at low cost. Several studies have proven the potential of cp genomes as tools to understand enigmatic and basal phylogenetic relationships at different taxonomic levels, as well as further probe the structural and functional evolution of plants. In this work, we present the complete cp genome sequence of *P. lambertii*.

Methodology/Principal Findings: The *P. lambertii* cp genome is 133,734 bp in length, and similar to other sequenced cupressophytes, it lacks one of the large inverted repeat regions (IR). It contains 118 unique genes and one duplicated tRNA (*trnN*-GUU), which occurs as an inverted repeat sequence. The *rps16* gene was not found, which was previously reported for the plastid genome of another Podocarpaceae (*Nageia nagi*) and Araucariaceae (*Agathis dammara*). Structurally, *P. lambertii* shows 4 inversions of a large DNA fragment ~20,000 bp compared to the *Podocarpus totara* cp genome. These unexpected characteristics may be attributed to geographical distance and different adaptive needs. The *P. lambertii* cp genome presents a total of 28 tandem repeats and 156 SSRs, with homo- and dipolymers being the most common and tri-, tetra-, penta-, and hexapolymers occurring with less frequency.

Conclusion: The complete cp genome sequence of *P. lambertii* revealed significant structural changes, even in species from the same genus. These results reinforce the apparently loss of *rps16* gene in Podocarpaceae cp genome. In addition, several SSRs in the *P. lambertii* cp genome are likely intraspecific polymorphism sites, which may allow highly sensitive phylogeographic and population structure studies, as well as phylogenetic studies of species of this genus.

1 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. Podocarpaceae are considered the most diverse family of Conifers, and much of this diversity has taken place within the *Podocarpus* and *Dacrydium* genera (KELCH, 1998). The Podocarpaceae family comprises 18 genera and 173 species distributed mainly in the Southern Hemisphere, but extending to the north in subtropical China, Japan, Mexico and the Caribbean (FARJON, 1998; BIFFIN *et al.*, 2011).

The *Podocarpus sensu lato* (*s.l.*) genus comprises nearly 100 species, widely spread throughout the Southern Hemisphere and northward to the West Indies, Mexico, southern China and southern Japan (PAGE, 1990). LEDRU *et al.*, (2007) described that *Podocarpus* populations in Brazil are widely dispersed in eastern Brazil, from north to south, and three endemic species have been reported: *Podocarpus sellowii* Klotzch ex Endl, *Podocarpus lambertii* Klotzch ex Endl, and *Podocarpus brasiliensis* de Laubenfels (de LAUBENFELS, 1985). *P. lambertii* is a native species from the Araucaria Forest, a subtropical moist forest ecoregion of the Atlantic Forest Biome, which is considered one of the 25 biodiversity hotspots of the world (MYERS *et al.*, 2000). It is a dioecious evergreen tree of variable height, measuring 1-10 m, shade-tolerant, adapted to high frequency and density of undergrowth (LONGHI, 2010).

Phylogeny analyses by maximum parsimony of Podocarpaceae family using 18S rDNA gene sequencing and morphological characteristics indicated Podocarpaceae as monophyletic and *Podocarpus s.l.* and *Dacrydium s.l.* genera as unnatural (KELCH, 1998). This author concluded that single-gene studies rarely result in perfect phylogenies, but they could provide a basis for choosing between competing hypotheses. PARKS *et al.*, (2009) suggested chloroplast (cp) genome sequencing as an efficient option for increasing phylogenetic resolution at lower taxonomic levels in plant phylogenetic and genetic population analyses.

The advancement of next-generation sequencing technologies has enabled the rapid acquisition of whole cp genome sequences at low cost when compared with traditional sequencing approaches. Chloroplast sequences are available for all families of Conifers: Cephalotaxaceae (YI *et al.*, 2013), Cupressaceae (HIRAO *et al.*, 2008), Pinaceae (WAKASUGI *et al.*, 1994; CRONN *et al.*, 2008; LIN *et al.*, 2010), Podocarpaceae (NC_020361.1; WU E CHAW, 2013), Taxaceae (NC_020321.1), and

Araucariaceae (WU E CHAW, 2013). For Podocarpus genus, the cp sequence of only one species has recently been obtained: the endemic New Zealand *Podocarpus totara* G. Benn. ex Don (NC_020361.1).

Several studies have proven the potential of cp genomes as tools to understand enigmatic and basal phylogenetic relationships at different taxonomic levels, as well as probe the structural and functional evolution of plants (MOORE *et al.*, 2007; JANSEN *et al.*, 2007; MOORE *et al.*, 2010; WU *et al.*, 2011a; YI *et al.*, 2013). HIRAO *et al.*, (2008) sequenced the cp genome of the first species in the Cupressaceae family, *Cryptomeria japonica*. They reported the deletion of one large inverted repeat (IR), numerous genomic rearrangements, and many differences in genomic structure between *C. japonica* and other land plants, thus supporting the theory that a pair of large IR can stabilize the cp genome against major structural rearrangements and, in turn, providing new insights into both the evolutionary lineage of coniferous species and the evolution of the cp genome (PALMER E THOMPSON, 1982; STRAUSS *et al.*, 1988; HIRAO *et al.*, 2008).

Chloroplast genome sequencing in gymnosperms also brought insights into evolutionary aspects in Gnetophytes. WU *et al.*, (2009) considered that the reduced cp genome size in Gnetophyte was based on a selection toward a lower-cost strategy by deletions of genes and noncoding sequences, leading to genomic compactness and accelerated substitution rates. More recently, comparative analysis of the cp genomes in cupressophytes and Pinaceae provided inferences about the loss of large IR (WU *et al.*, 2011a; YI *et al.*, 2013). On one hand, WU *et al.*, (2011a) and WU E CHAW (2013) argue that each Pinaceae and cupressophyte lost a different copy of IR. On the other hand, YI *et al.*, (2013) showed that distinct isomers are considered as alternative structures for the ancestral cp genome of cupressophyte and Pinaceae lineages. Therefore, it is not possible to distinguish between hypotheses favoring retention or independent loss of the same IR region in cupressophyte and Pinaceae cp genomes.

The present study focuses on establishing the complete cp genome sequence of a further member of the Podocarpaceae family, the Brazilian endemic species *P. lambertii*. Here, we characterize the cp genome organization of *P. lambertii* and compare its cp genome structure with other conifer species.

2 Material and Methods

2.1 Plant material and cp DNA purification

Chloroplast isolation of *P. lambertii* was performed from young plants collected at a private area located at Lages, Santa Catarina, Brazil (27° 48' 57" S, 50° 19' 33" W), where the species is abundant, with previous permission from the owner (José Antônio Ribas Ribeiro). This species is not considered threatened. Afterwards, the young plants were transplanted to the greenhouse until the collection of needles. The cpDNA isolation was performed according to VIEIRA *et al.*, (2014).

2.2 Chloroplast genome sequencing, assembling and annotation

Approximately 50 ng of cp DNA were used to prepare sequencing libraries with Nextera DNA Sample Prep Kit (Illumina Inc., San Diego, CA) according to the manufacturer's instructions. Chloroplast DNA was sequenced using Illumina MiSeq (Illumina Inc., San Diego, CA) at the Federal University of Paraná, Brazil. In total, 495,071 paired-end reads (2 x 250 bp) were obtained, and *de novo* assembly was performed using Newbler 2.6v. The obtained paired-end reads were mapped on *P. lambertii* cp genome and the genome coverage estimated using the CLC Genomics Workbench 5.5 software. By using this approach, a total of 377,437 paired-end reads (76.23%) was obtained from cpDNA, resulting in 1,200-fold genome coverage. Initial annotation of the *P. lambertii* cp genome 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 cp genomes. The tRNA genes were further verified by using tRNAscan-SE (SCHATTNER *et al.*, 2005). A physical map of the cp circular genome was drawn using OrganellarGenomeDRAW (OGDRAW) (LOHSE *et al.*, 2013). The resulting annotated sequence has been submitted to the National Center for Biotechnology Information (NCBI) under the accession number KJ010812.

2.3 Comparative analysis of genome 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 *P. lambertii* and the non-Pinaceae conifer

representatives *P. totara* (Podocarpaceae), *Cephalotaxus oliveri, Cephalotaxus wilsoniana* (Cephalotaxaceae), *Taxus mairei* (Taxaceae), *Taiwania cryptomerioides, T. flousiana* (Cupressaceae), *C. japonica* (Cupressaceae), as well as *Pinus thunbergii*, a Pinaceae representative.

2.4 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. Tandem repeats were analyzed using Tandem Repeats Finder (TRF) (BENSON, 1999) with parameter settings of 2, 7 and 7 for match, mismatch, and indel, respectively. The minimum alignment score and maximum period size were set as 50 and 500, respectively. All of the repeats found were manually verified, and the nested or redundant results were removed. REPuter (KURTZ *et al.*, 2001) was used to visualize the remaining IRs in *P. lambertii* by forward *vs*. reverse complement (palindromic) alignment. The minimal repeat size was set to 30 bp and the identity of repeats ≥ 90%.

3 Results and Discussion

3.1 Chloroplast genome sequencing, assembling and annotation

P. lambertii cp genome size was determined to be 133,734 bp, which is very similar to *P. totara* (133,259 bp) (MARSHALL, Unpublished, NC_020361.1) and larger than the sequenced cp genomes of Pinaceae species, which range from 116,479 bp in *Pinus monophylla* (CRONN *et al.*, 2008) to 124,168 bp in *Picea morrisonicola* (WU *et al*., 2011b). *P. lambertii* cp genome size is smaller than the cp sequences in the cycads *Cycas taitungensis* (163,403 bp) (WU *et al.*, 2007) and *Cycas Revoluta* (162,489 bp) (LI *et al.,* Unpublished, NC_020319.1). The genome size of *P. lambertii* cp is consistent with the size of non-Pinaceae conifer species, which ranges from 127,665 bp in *T. mairei* (LI *et al.*, unpublished, NC_020321.1) to 136,196 bp in *C. wilsoniana* (WU *et al.*, 2011). A total of 119 genes were identified in the *P. lambertii* cp genome, of which 118 genes were single copy and one gene, *trnN*-GUU, was duplicated and occurred as an inverted repeat sequence. 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

Figure 1. Gene map of *Podocarpus lambertii* chloroplast 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 corresponds to AT content.

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, 14 were genes containing introns (Table 1). The GC content determined for *P. lambertii* cp genome is 37.1%, which is higher than *C. oliveri* (35.2%), *C. wilsoniana* (35.1%), *T.*

cryptomerioides (34.6%), and *C. japonica* (35.4%), but lower than *C. taitungensis* (39.5%) and *P. thunbergii* (38.8%).

Table 1. List of genes identified in *Podocarpus lambertii* chloroplast genome.

Category of Genes	Group of gene	Name of gene					
Self-replication	Ribosomal RNA genes	*rrn16*	*rrn23*	*rrn5*	*rrn4.5*		
	Transfer RNA genes	*trnA*-UGC *	*trnC*-GCA	*trnD*-GUC	*trnE*-UUC	*trnF*-GAA	*trnfM*-CAU
		trnG-UCC*	*trnG*-GCC	*trnH*-GUG	*trnI*-CAU	*trnI*-GAU *	*trnK*-UUU *
		trnL-CAA	*trnL*-UAG	*trnM*-CAU	*trnN*-GUU**	*trnP*-GGG	*trnP*-UGG
		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 **	*petG*	*petL*	*petN*
	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*					
Pseudogenes		*ycf68*					

* Genes containing introns.
** Duplicated gene.

3.2 Gene content differences

The gene content of *P. lambertii* cp genome and that of other conifer cp genomes sequenced to date show high similarity. However, some differences are observed when we compare *P. lambertii* cpDNA with other non-Pinaceae and Pinaceae conifers. One exception is the *rps16* gene, which is absent from the *P. lambertii* cp genome. This result reinforce the apparently loss of *rps16* gene in Podocarpaceae and Araucariaceae families. WU E CHAW, 2013 reported the *rps16* gene loss in *Nageia nagi* (Podocarpaceae) and *Agathis dammara* (Araucariaceae). This gene is present in other non-Pinaceae conifer cp genomes published so far (WU *et al.*, 2007; HIRAO *et al.*, 2008; WU *et al.*, 2011a; YI *et al.*, 2013). The *rps16* gene loss has already been reported in other gymnosperms, such as

Pinaceae and Gnetophyte species (TSUDZUKI *et al.*, 1992; WU *et al.*, 2007; WU *et al.*, 2009). WU *et al.*, (2011a) considered *rps16* gene loss as a structural mutation unique to the cpDNAs of gnetophytes and Pinaceae, but since the loss of this gene has been identified in Podocarpaceae and Araucariaceae families, we can consider that some cupressophytes may also present this mutation. This gene is also absent, or nonfunctional, in some angiosperm species of the Fabaceae family, such as *Medicago truncatula*, in which it is completely absent, and in *Phaseolus vulgaris* and *Vigna radiata*, in which it is nonfunctional. In this angiosperm family, the coding sequence contains many internal stop codons and a modified initial stop codon (GUO *et al.*, 2007; TANGPHATSORNRUANG *et al.*, 2010). Since this gene was shown to be essential for cell survival in tobacco (FLEISCHMANN *et al.,* 2011), it was probably transferred to the nucleus, as observed for different species of the Fabaceae family (GUO *et al.*, 2007; TANGPHATSORNRUANG *et al.*, 2010), and has since become a functional nuclear gene required for normal plastid translation.

The *trnP*-GGG and *trnR*-CCG genes are considered to be relics of plastid genome evolution in gymnosperms, pteridophytes and bryophytes (SUGIURA E SUGITA, 2004). The *trnP*-GGG gene is present in the *P. lambertii* cp genome, as well as such conifer species as *C. japonica, P. thunbergii*, *C. oliveri* and *C. wilsoniana* and other gymnosperm species, such as *C. taitungensis, Gnetum* and *Ginkgo.* The *trnR*-CCG gene is present as complete and functional tRNA in *P. lambertii* (Podocarpaceae), as well as the cp genomes of *P. thunbergii (*Pinaceae), *C. taitungensis* (Cycadaceae) (WU *et al.*, 2007), whereas it is absent from *C. japonica* (Cupressaceae)*, C. oliveri* and *C. wilsoniana* (Cephalotaxaceae), and *T. mairei* (Taxaceae) (HIRAO *et al.*, 2008; YI *et al.*, 2013; LI *et al.*, unpublished). HIRAO *et al.*, (2008) suggested that *trnR*-CCG might have been completely lost in the Cupressaceae *s.l.*, which has only relatively recently diverged during the long evolutionary history of plants. These data corroborate the hypothesis based on phytochrome phylogenetic trees, in which the most ancient branch of the conifers seems to be the Pinaceae, and the next split appears to have separated Araucariaceae plus Podocarpaceae from the Taxaceae/Taxodiaceae/Cupressaceae group (SCHMIDT E SCHNEIDER-POETSCH, 2002). This trnR-CCG gene may have been lost during the second split separating Araucariaceae and Podocarpaceae taxa. In addition, *trnT*-GGU occurs as a pseudogene in the *C. japonica* cp genome, with only 43 bp, while it is present and completely functional in *P. lambertii* and *C. oliveri*, *C. wilsoniana,* duplicated in *P. thunbergii*, and totally absent from the *C. taitungensis* cp genome. Interestingly, the *trnT*-GGU gene is highly conserved in angiosperms, and

knockout of this gene in tobacco plants produced viable plants, whereas the growth of these plants was strongly affected, suggesting an important role during plastid translation (ALKATIB *et al.*, 2012). The loss of the *trnT*-GGU gene in several gymnosperm species suggests that a uridine modification in the anticodon position of the *trnT*-GGU gene occurred during evolution, which would facilitate the reading of threonine codons and make the *trnT*-GGU gene dispensable in these species (AMBROGELLY *et al.*, 2007; WEIXLBAUMER *et al.*, 2007; ALKATIB *et al.*, 2012). Evolutionarily, the loss of this tRNA gene could be used as a tool, or marker gene, to study the possible ways that the conifers diverged during evolution. However, it remains to be determined whether structural differences in the cp ribosome or modification in the structure of this tRNA, between angiosperms and gymnosperms, would facilitate the decoding.

3.3 Comparative analysis of genome structure

Chloroplast genome organization is much conserved in angiosperms, as well as the presence of IRs, with very few exceptions. As reported by TERAKAMI *et al.*, (2012) in *Pyrus*, *Malus* and *Nicotiana,* neither translocation nor inversion was detected in the three species. In addition, considering the many dicot and monocot species, only one large inversion was reported (TERAKAMI *et al.*, 2012).

In addition to the loss of the large IR in conifers, many genome rearrangements were observed in the cp genome, and such rearrangements appear to play an important role in their evolution. Dot-plot analyses indicate that the structure of the *P. lambertii* cp genome differs significantly from cp genomes of other conifer species, and, surprisingly, it has significant differences when compared to *P. totara* (Figure 2A-H).

For the genus *Cephalotaxus s.l.*, specifically *C. wilsoniana* and *C. Oliveri,* it was shown that the genome structures were almost the same (YI *et al.*, 2013). Similar results were observed in the present study, as revealed by the high similarity in the dot-plot analyses between *Podocarpus* and *Cephalotaxus* genera, as represented by *P. lambertii* x *C. wilsoniana* (Figure 2E) and *P. Lambertii* x *C. oliveri* (Figure 2F), and between the *Podocarpus* and *Taiwania* genera, as represented by *P. lambertii* x *T. flousiana* (Figure 2G) and *P. lambertii* x *T. cryptomerioides* (Figure 2H). This high similarity in dot-plot analysis indicates the occurrence of exactly the same structural modifications between *P. lambertii* and these two *Cephalotaxus* and *Taiwania species.*

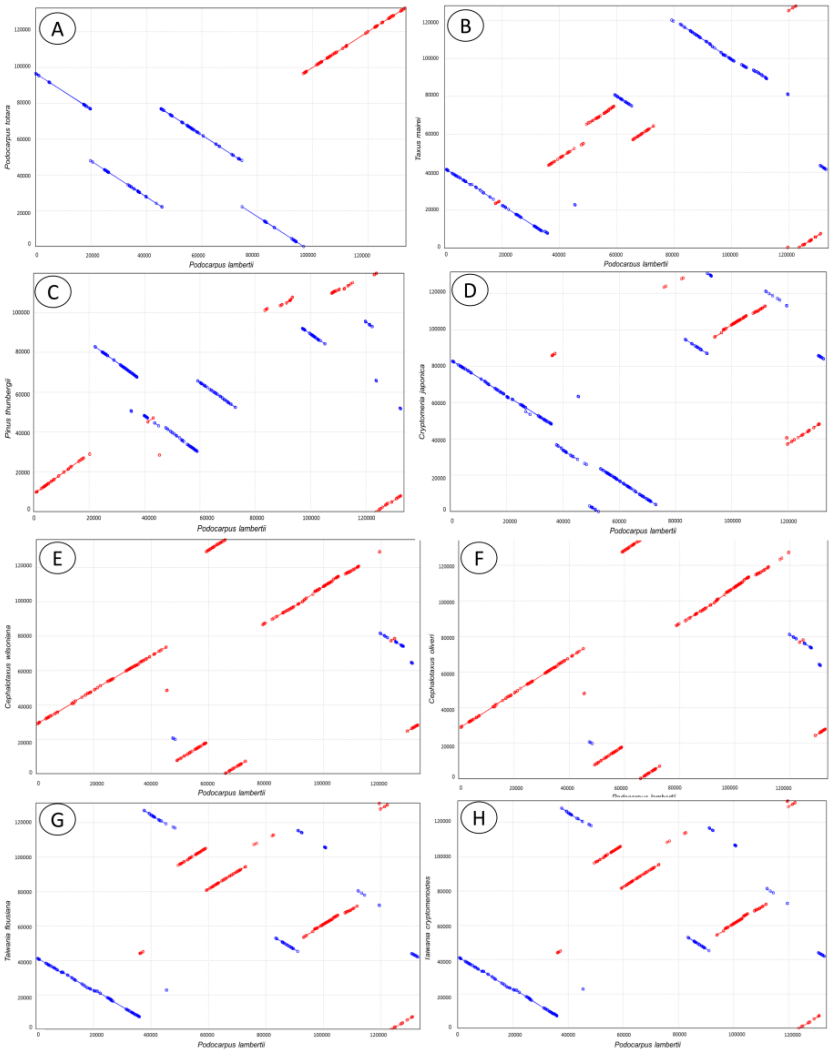

Figure 2. Dot-plot analyses of eight sampled conifer chloroplast DNAs against *Podocarpus lambertii*. A positive slope denotes that the two compared sequences are in the same orientation, whereas a negative slope indicates that the compared sequences can be aligned, but their orientations are opposite. Graphs represents comparisons between *Podocarpus lambertii* (axis X) and *Podocarpus totara* (A), *Taxus mairei* (B), *Pinus thunbergii* (C), *Cryptomeria japonica* (D), *Cephalotaxus wilsoniana* (E), *Cephalotaxus oliveri* (F), *Taiwania flousiana* (G), and *Taiwania cryptomerioides* (H) in axis Y.

Differently, for *P. lambertii* and *P. totara* (Figure 2A), we observed four large inversions of about 20,000 bp in length each. In both *Cephalotaxus* and *Taiwania* genera, the two sequenced species share the same region of natural occurrence, which is not true for either *Podocarpus* species sequenced. Thus, these large inversions can be explained by, and probably result from, the large distance between the natural occurrence of these two species in that *P. lambertii* occurs in Brazil, while *P. totara* occurs in New Zealand. Moreover, podocarps have a rich fossil record that suggests an origin in the Triassic period (about 220 million years) and a distribution in both the Northern and Southern Hemispheres through the Cretaceous and earliest Tertiary periods, about 100 million years ago (HILL E BRODRIBB, 1999; FARJON, 2008; MORLEY, 2011). Thus, geographic distance and different adaptive traits could explain the structural differences found between these two species of the same genera.

In addition, the loss of one large IR copy already reported in other conifer species were also observed in the *P. lambertii* cp genome (HIRAO *et al.* 2008; WU *et al.* 2011; YI *et al.*, 2013). However, short remaining IR sequences of 326 bp can be found in *P. lambertii*, 544 bp in *C. oliveri,* 530 bp in *C. wilsoniana,* 277 bp in *T. cryptomerioides* and 284 bp in *C. japonica* (YI *et al.*, 2013). These short remaining IR sequences also differ in the nucleic acid sequences and gene content between different conifer species. In *P. lambertii*, *trnN*-GUU remain from the lost IR copy region, while in *T. cryptomerioides* and *C. japonica*, *trnI*-CAU remained after the rearrangements that determined the loss of one IR copy (YI *et al.*, 2013). In *C. oliveri* and *C. wilsoniana*, the *trnQ*-UUG is duplicated; however, this gene is not normally present in the IR region, and its duplication was probably produced by other rearrangements not involved with the IR regions (WU *et al.*, 2011). After much evidence provided by different conifer plastid genomes, it can be concluded that the loss of one IR copy occurred after a reduction in sequence and gene content and that such loss was most likely caused by this reduction (TSUDZUKI *et al.*, 1992; WU *et al.*, 2007; CRONN *et al.*, 2008; HIRAO *et al.*, 2008; WU *et al.*, 2009; WU *et al.*, 2011; YI *et al.*, 2013). However, this speculation remains to be established. To date, it is not entirely clear whether cupressophytes and Pinaceae species have lost different IR regions (YI *et al.*, 2013). However, we can observe in *P. lambertii* an inversion in the direction of transcription of ribosomal RNA genes spanning *rrn5-rrn16* and protein-coding genes, *ndhB* and *ycf2,* when compared to *C. oliveri, C. wilsoniana, T. cryptomerioides* and *C. japonica* (Figure 3).

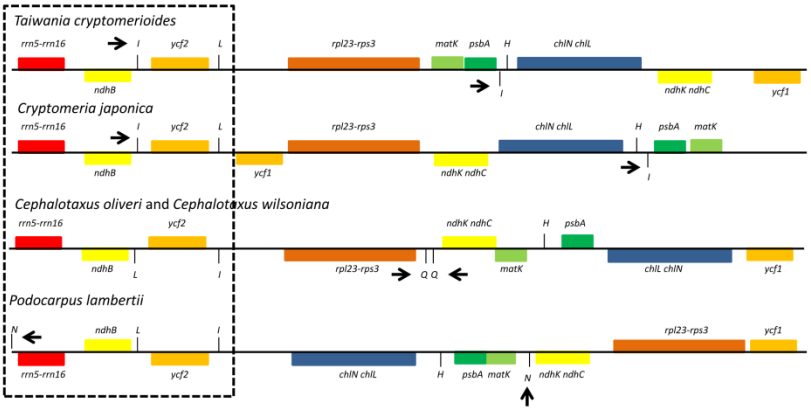

Figure 3. Comparison of IR and genome structure in 5 cupressophytes. Five cupressophyte species from top to bottom are *Taiwania cryptomerioides*, *Cryptomeria japonica*, *Cephalotaxus oliveri*, *Cephalotaxus wilsoniana* and *Podocarpus lambertii*. Genes are represented by boxes extending above or below the baseline, according to the direction of transcription; genes with the same function have the same color. Transfer RNA genes are abbreviated as the type of one letter. Dashed boxes represent the retained IR region, and arrows indicate the short IR on each species. Adapted from YI *et al.*, (2013).

3.4 Repeat sequence analysis

The cp genome mode of inheritance, paternal in most gymnosperms, allows us to elucidate the relative contributions of seed and pollen flow to the genetic structure of natural populations by comparison of nuclear and cp markers (PROVAN *et al.*, 2001). The cp microsatellites, or SSRs, may be identified in completely sequenced plant cp genomes by simple database searches, followed by primers designed to screen for polymorphism. To date, studies of cp microsatellites have revealed much higher levels of diversity than have those of cp restriction fragment length polymorphisms (RFLP) (PROVAN *et al.*, 1997, 1999, 2001).

We have analyzed the occurrence, type, and distribution of SRRs in the *P. lambertii* cp genome. In total, 156 SSRs were identified. Among them, homo- and dipolymers were the most common with, respectively, 80 and 63 occurrences, whereas tri- (4), tetra- (7), penta- (1), and hexapolymers (1) occur with lower frequency (Table 2). Most homopolymers are constituted by A/T sequences (87.5%), and of the dipolymers, 61.1% were also constituted by multiple A and T bases.

Table 2. List of simple sequence repeats identified in *Podocarpus lambertii* chloroplast genome.

SSR sequence	Number of repeats													TOTAL
	3	4	5	6	7	8	9	10	11	12	13	14	15	
A/T	-	-	-	-	-	39	14	6	4	6	-	-	1	70
C/G	-	-	-	-	-	-	3	3	-	1	1	-	2	10
AC/GT	-	1	-	1	-	-	-	-	-	-	-	-	-	2
AG/CT	-	21	1	-	-	-	-	-	-	-	-	-	-	22
AT/AT	-	24	7	2	2	-	3	1	-	-	-	-	-	39
AAG/CTT	-	1	-	-	-	-	-	-	-	-	-	-	-	1
AAT/ATT	-	3	-	-	-	-	-	-	-	-	-	-	-	3
AATC/ATTG	2	-	-	-	-	-	-	-	-	-	-	-	-	2
AATG/ATTC	1	-	-	-	-	-	-	-	-	-	-	-	-	1
AATT/AATT	1	-	-	-	-	-	-	-	-	-	-	-	-	1
ACAT/ATGT	1	-	-	-	-	-	-	-	-	-	-	-	-	1
ACCT/AGGT	1	-	-	-	-	-	-	-	-	-	-	-	-	1
AGAT/ATCT	3	-	-	-	-	-	-	-	-	-	-	-	-	3
AAATG/ATTTC	1	-	-	-	-	-	-	-	-	-	-	-	-	1
AGATAT/ATATCT	1	-	-	-	-	-	-	-	-	-	-	-	-	1
TOTAL														**158**

In this study, we identified 78 repeats with more than one nucleotide repeat, totaling almost 50% of all SSRs identified. The 13 tri-, tetra-, penta-, and hexapolymers are shown in Table 3, as well as their size and location. From these 13 polymers identified, 9 are localized in intergenic spacers, 3 in coding sequences, and only 1 inside an intron. These results reveal the presence of several SSR sites in *P. lambertii*. Hereafter, these sites can be assessed for the intraspecific level of polymorphism, leading to highly sensitive phylogeographic and population structure studies for this species.

Tandem repeats with more than 30 bp and with a sequence identity of more than 90% have also been examined. Twenty-eight tandem repeats were identified in the *P. lambertii* cp genome (Table 4), of which 15 are located in coding regions of *accD* (2), *rps18* (1), *rps19* (1), *rps11* (1), *ycf1* (8), *rpl32* (1), *ycf2* (1); 11 are distributed in the intergenic spacers of *atpA*/*atpF* (1), *trnR*-CCG/*accD* (1), *rpl2*/*rps19* (1), *clpP*/*ycf1* (2), *ndhE*/*psaC* (1), *trnR*-ACG/*rrn5* (1), *rps12*/*rps7* (1), *ycf2*/*trnI*-CAU (1), *trnQ*-UUG/*psbK* (1), *psbK*/*psbI* (1); and 2 are located in the intron sequence of *rpoC1*. The cp genome of *P. lambertii* has 11 tandem repeats, more than the cp genome of *C. oliveri*, as well as a higher number of repeats in the *ycf1* (6) gene coding sequence (YI *et al.*, 2013). The *ycf1* gene, previously considered as an enigmatic function in the cp genome, has

recently been identified as encoding an essential protein component of the cp translocon at the inner envelope membrane (TIC) (KIKUCHI *et al.*, 2013). In *Salvia miltiorrhiza* and *Cocos nucifera*, two angiosperms, only 7 and 8 tandem repeats, respectively, of about 20 bp were identified, none of them located at the *ycf1* coding sequence (HUANG *et al.*, 2013; QIAN *et al.*, 2013), corroborating the theory that the IR influences the stability of the plastid genome.

Table 3. Distribution of tri-, tetra-, penta-, and hexapolymer simple sequence repeats (SSRs) loci in *Podocarpus lambertii* chloroplast genome.

SSR type	SSR sequence	Size	Start	End	Location
penta	(AATGA)3	15	21884	21898	*trnE*-*UUC*/*trnT*-GGU (IGS)
hexa	(AGATAT)3	18	37894	37911	*trnF*-GAA/*ndhJ* (IGS)
tetra	(ATCA)3	12	44346	44357	*atpE*/*rbcL* (IGS)
tri	(AAG)4	12	75761	75772	*Ycf1* (CDS)
tetra	(AATG)3	12	86350	86361	*ndhA* (intron)
tetra	(TGAT)3	12	97140	97151	*ndhF*/*trnN*-GUU (IGS)
tetra	(CTAC)3	12	99809	99820	*rrn23* (CDS)
tri	(ATT)4	12	103664	103675	*trnI*-GAU/*rrn16* (IGS)
tri	(ATA)4	12	120539	120550	*rps7*/*ndhB* (IGS)
tri	(TTA)4	12	122046	122057	*chlL* (CDS)
tetra	(AATT)3	12	122977	122988	*chlL*/*trnH*-GUG (IGS)
tetra	(CATA)3	12	125437	125448	*psbA*/*trnK*-UUU (IGS)
tetra	(ATAG)3	12	125570	125581	*psbA*/*trnK*-UUU (IGS)

CDS, coding sequences; IGS, intergenic spacers.

YI *et al.*, 2013 attributed the expansion of the *accD* ORF to the presence of tandemly repeated sequences. In the *P. lambertii* cp genome, we identified 2 tandem repeats in *accD* CDS, totaling 132 bp, or 44 codons. The *accD* reading frame length of the *P. lambertii* cp genome is 864 codons, similar to other cupressophyte species, such as *C. oliveri* (936 codons), *C. wilsoniana* (1,056 codons), *C. japonica* (700 codons) and *T. cryptomerioides* (800 codons). In contrast, the reading frame lengths of cycads, Ginkgo and Pinaceae, range from 320 to 359 codons, less than half the size found in cupressophytes. These results support the hypothesis of HIRAO *et al.*, (2008) and YI *et al.*, (2013) which holds that the *accD* reading frame has displayed a tendency toward enlarging sizes in cupressophytes.

Table 4. Distribution of tandem repeats in *Podocarpus lambertii* chloroplast genome.

Serial Number	Repeat Length (bp)	Consensus size x Copy number	Start-End	Location
1	32	16 x 2	3450-3482	*atpA*/*atpF* (IGS)
2	284	142 x 2	13170-13454	*rpoC1* (Intron)
3	60	30 x 2	13496-13557	*rpoC1* (Intron)
4	30	15 x 2	46625-46653	*trnR*-CCG/*accD* (IGS)
5	90	30 x 3	47533-47619	*accD* (CDS)
6	42	21 x 2	48149-48192	*accD* (CDS)
7	52	26 x 2	57988-58043	*rps18* (CDS)
8	32	16 x 2	61875-61905	*rpl2*/*rps19* (IGS)
9	54	18 x 3	62177-62237	*rps19* (CDS)
10	63	21 x 3	66568-66630	*rps11* (CDS)
11	32	16 x 2	75172-75203	*clpP*/*ycf1* (IGS)
12	104	52 x 2	75412-75529	*clpP*/*ycf1* (IGS)
13	36	18 x 2	79255-79292	*ycf1* (CDS)
14	162	52 x 3	79351-79504	*ycf1* (CDS)
15	162	81 x 2	79362-79519	*ycf1* (CDS)
16	108	27 x 4	79401-79519	*ycf1* (CDS)
17	132	33 x 4	80478-80619	*ycf1* (CDS)
18	96	24 x 4	80732-80820	*ycf1* (CDS)
19	273	21 x 13	81305-81571	*ycf1* (CDS)
20	96	48 x 2	82408-82528	*ycf1* (CDS)
21	30	15 x 2	89787-89817	*ndhE*/*psaC* (IGS)
22	126	42 x 3	93843-93963	*rpl32* (CDS)
23	64	32 x 2	97838-97902	*trnR*-ACG/*rrn5* (IGS)
24	300	60 x 5	109209-109531	*rps12*/*rps7* (IGS)
25	36	12 x 3	116515-116547	*ycf2* (CDS)
26	60	20 x 3	119998-120055	*ycf2*/*trn1*-CAU (IGS)
27	128	64 x 2	131733-131853	*trnQ*-UUG/*psbK* (IGS)
28	26	13 x 2	132530-132556	*psbK*/*psb1* (IGS)

CDS, coding sequences; IGS, intergenic spacers.

The complete cp genome sequence of *P. lambertii* revealed significant structural changes occurring in the cp genome, even in species from the same genus. These results reinforce the apparently loss of *rps16* gene in Podocarpaceae cp genome. In addition, several SSRs in the *P. lambertii* cp genome are likely intraspecific polymorphism sites which may allow highly sensitive phylogeographic and population structure studies, as well as phylogenetic studies, of species of this genus.

Financial Disclosure

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 funders had no role in

study design, data collection and analysis, decision to publish, or preparation of the manuscript.

CONSIDERAÇÕES FINAIS E PERSPECTIVAS FUTURAS

Os resultados do presente trabalho apresentam avanços expressivos na área de sequenciamento de genomas plastidiais em coníferas. Primeiramente, foi desenvolvido um protocolo para isolamento de cloroplasto com alta qualidade em coníferas. A partir disso, possibilitou-se o isolamento de DNA plastidial com alta qualidade, podendo ser empregado para sequenciamento em sequenciadores de nova geração. Com o DNA plastidial isolado, aumenta-se significativamente o número de *reads* obtidos que realmente são plastidiais. No caso do *P. lambertii*, estima-se que 76% dos *reads* obtidos foram de DNA plastidial, e não de possíveis contaminações com DNA mitocondrial ou nuclear. Esse é um importante fator facilitador no momento da montagem do genoma, devido à alta cobertura média do genoma, onde com o *P. lambertii*, por exemplo, obteve-se uma cobertura média de 1200x no genoma.

Além disso, é possível que esse protocolo possa ser aplicado para outras espécies nativas de difícil isolamento do cloroplasto por seu alto conteúdo endógeno de polifenóis, polissacarídeos e outros contaminantes. Os tampões e reagentes usados nesse protocolo contam com uma alta carga de antioxidantes, somado ao uso do Percoll que aumenta significativamente a pureza do cloroplasto isolado.

O cloroplasto isolado com alta qualidade também torna possível a aplicação para o isolamento de proteínas, permitindo também análises de proteômica em cloroplastos de coníferas. Esses estudos podem esclarecer importantes questões em proteínas diferenciais no cloroplasto sob diferentes estímulos, como estresse osmótico, intensidade luminosa e outros. O isolamento dessas proteínas permite um trabalho mais específico em relação às proteínas contidas no cloroplasto.

Adicionalmente, a partir dos dados obtidos com o sequenciamento do genoma plastidial do *P. lambertii* foram apontadas diversas regiões repetidas no genoma que podem ser verificadas para possíveis polimorfismos. A partir disso, pode ser realizado o desenho de iniciadores espécie-específicos para filogeografia dessa espécie e possivelmente de espécies relacionadas.

Além disso, foram geradas informações que servem de base para futuros trabalhos a serem realizados no Laboratório de Fisiologia do Desenvolvimento e Genética Vegetal (LFDGV) e em outros laboratórios do Programa de Pós-Graduação em Recursos Genéticos Vegetais. Essa nova linha de pesquisa aberta no LFDGV/UFSC pode dar continuidade com trabalhos de isolamento, sequenciamento de organelas e análises de bioinformática, as quais serão de grande importância para qualificação e

formação de recursos humanos e para o avanço do conhecimento científico nesta área.

REFERÊNCIAS

AHLERT, D.; STEGEMANN, S.; KAHLAU, S.; RUF, S.; BOCK, R. Insensitivity of chloroplast gene expression to DNA methylation. **Molecular Genetics and Genomics**, v. 282, p. 17–24, 2009.

ALKATIB, S.; SCHARFF, L. B.; ROGALSKI, M.; FLEISCHMANN, T. T.; MATTHES, A. et al. The contributions of wobbling and superwobbling to the reading of the genetic code. PLoS Genetics, 8(11):e1003076, 2012

AMBROGELLY, A.; PALIOURA, S.; SOLL, D. Natural expansion of the genetic code. **Nature Chemical Biology**, v. 3, p.29-35, 2007.

ARIF, I. A.; BAKIR, M. A.; KHAN, H. A.; AHAMED, A.; FARHAN, A. H. et al. (2010) A simple method for DNA extraction from mature date palm leaves: impact of sand grinding and composition of lysis buffer. **International Journal of Molecular Sciences**, v. 11, p. 3149-3157.

ATHERTON, R. A.; MCCOMISH, B. J.; SHEPHERD, L. D.; BERRY, L. A.; ALBERT, N. W. et al. Whole genome sequencing of enriched chloroplast DNA using the Illumina GAII platform. **Plant Methods**, v. 6, n. 22, 2010.

AYLIFFE, M. A.; TIMMIS, J. N. Tobacco nuclear DNA contains long tracts of homology to chloroplast DNA. **Theoretical and Applied Genetics**, v. 85, p. 229–238, 1992.

AYLIFFE, M. A.; SCOTT, N. S.; TIMMIS, J. N. Analysis of plastid DNA like sequences within the nuclear genomes of higher plants. **Molecular Biology Evolution**, v. 15, p. 738–745, 1998.

BENDICH, A. J. Why do chloroplasts and mitochondria contain so many copies of their genome? **BioEssays**, v. 6, p. 279-282, 1987.

BIFFIN, E.; CONRAN, J.; LOWE, A. Podocarp Evolution: A Molecular Phylogenetic Perspective. In: TURNER, B. L.; CERNUSAK, L. A. (Org.). **Ecology of the Podocarpaceae in Tropical Forests**. Washington: Smithsonian Institution Scholarly Press, 2011. p. 1–20.

BITTENCOURT, R. **Distribuição da diversidade genética e estratégias para conservação de quatro espécies na Floresta Ombrófila Mista ameaçadas no Estado de Santa Catarina**. 2011. 135 f. Tese (Doutorado em Recursos Genéticos Vegetais) - Centro de Ciências Agrárias, Universidade Federal de Santa Catarina, Florianópolis, 2011

BOCK, R. Transgenic plastids in basic research and plant biotechnology. **Journal of Molecular Biology**, v. 312, p. 425–438, 2001.

BOCK, R. Plastid biotechnology: prospects for herbicide and insect resistance, metabolic engineering and molecular farming. **Current Opinion in Biotechnology**, v. 18, p. 100–106, 2007.

BOCK, R. Strategies for metabolic pathway engineering with multiple transgenes. **Plant Molecular Biology**, v. 83, p. 21-31, 2013.

BOCK, R. Genetic engineering of the chloroplast: novel tools and new applications. **Current Opinion in Biotechnology**, v. 26, p. 7-13, 2014.

BOOKJANS, G.; STUMMANN, B. M.; HENNINGSEN, K. W. Preparation of chloroplast DNA from pea plastids isolated in a medium of high ionic strength. **Analytical Biochemestry**, v. 141, p. 244–247, 1984.

BRUNSFELD, S. J.; SOLTIS, P. S.; SOLTIS, D. E.; GADEK, P. A.; QUINN, C. J.; STRENGE, D. D.; RANKER, T. A. Phylogenetic relationships among the genera of Taxodiaceae and Cupressaceae: evidence from *rbcL* sequences. **Systematic Botany**, v. 19, p. 253-262, 1994.

CARVALHO, P. E. R. **Espécies florestais brasileiras**. Recomendações silviculturais, potencialidades e uso da madeira. Brasília: EMBRAPA, 1994.

CHUMLEY, T. W.; PALMER, J. D.; MOWER, J. P.; BOORE, J. L.; FOURCADE, H. M. et al. The complete chloroplast genome sequence of *Pelargonium* x *hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. **Molecular Biology Evolution**, v. 23, p. 1–16, 2006.

CLARKE, J. L.; DANIELL, H.; NUGENT, J. M. Chloroplast biotechnology, genomics and evolution: current status, challenges and future directions. **Plant Molecular Biology**, v. 76, p. 207-209, 2011

CRONN, R.; LISTON, A.; PARKS, M.; GERNANDT, D. S.; SHEN, R. et al. Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. **Nucleic Acids Research**, v. 36, n. 19:e122, 2008.

DIEKMANN, K.; HODKINSON, T. R.; FRICKE, E.; BARTH, S. An optimized chloroplast DNA extraction protocol for grasses (Poaceae) prove suitable for whole plastid genome sequencing and SNP detection. **PLoS ONE**, n.3: e2813, 2008.

FARJON, A. **World checklist and bibliography of conifers**. Kew: The Royal Botanical Gardens, 1988.

FARJON, A. **A natural history of conifers.** Portland: Timber Press, 2008.

FLEISCHMANN, T. T.; SCHARFF, L. B.; ALKATIB, S.; HASDORF, S.; SCHOTTLER, M. A. et al. Nonessential plastid-encoded ribosomal proteins in tobacco: a developmental role for plastid translation and implications for reductive genome evolution. **Plant Cell**, v. 23, p. 3137–3155, 2011.

FOJTOVA, M.; KOVARIK, A.; MATYASEK, R. Cytosine methylation of plastid genome in higher plants. fact or artefact? **Plant Science**, v. 160, p. 585–593, 2001.

GARCIA, L. C.; NOGUEIRA, A. C.; ALQUINI, Y. Aspectos morfo-anatômicos de sementes de *Podocarpus lambertii* Klotz. e *Podocarpus sellowii* Klotz. - Podocarpaceae. **Revista Brasileira de Sementes**, v. 28, n. 3, p. 129-134, 2006.

GARCIA, C. G.; NOGUEIRA, A. C. Resposta de sementes de *Podocarpus lambertii* e *Podocarpus sellowii* - (Podocarpaceae) à dessecação. **Ciência Florestal**, v. 18, n. 3, p. 347-352, 2008.

GARGANO, D.; MAPLE-GRØDEM, J.; REISINGER, V.; EICHACKER, L.A.; MØLLE, S.G. Analysis of the chloroplast proteome in arc mutants and identification of novel protein components associated with FtsZ2. **Plant Molecular Biology**, v. 81, p. 235–244, 2013.

GIELLY, L.; TABERLET, P. The use of chloroplast DNA to resolve plant phylogenies: noncoding versus *rbcL* sequences. **Molecular Biology and Evolution**, v. 11, p. 769-777, 1994.

GOREMYKIN, V. V.; SALAMINI, F.; VELASCO, R.; VIOLA, R. Mitochondrial DNA of *Vitis vinifera* and the issue of rampant horizontal gene transfer. **Molecular Biology Evolution**, v. 26, p. 99–110, 2009.

GREINER, S.; BOCK, R. Tuning a ménage à trois: Co-evolution and co-adaptation of nuclear and organellar genomes in plants. **Bioessays**, v. 35, p. 354–365, 2013.

HAGEMANN, R. The sexual inheritance of plant organelles. In: DANIELL, H.; CHASE, C. (Org,). **Molecular Biology and Biotechnology of Plant Organelles**. Heidelberg: Springer, 2004, p. 93-113.

HASEBE, M.; KOFUGI, R.; ITO, M.; KATO, M.; IWATSUKI, K.; UEDA, K. Phylogeny of gymnosperms inferred from *rbcL* gene sequences. **The Botanical. Magazine**, Tokyo, v. 105, p. 673-679, 1992.

HE, Z. H.; LI, H. L.; SHEN, Y.; LI, Z. S.; MI, H. Comparative analysis of the chloroplast proteomes of a wheat (*Triticum aestivum* L.) single seed descent line and its parents. **Journal of Plant Physiology**, v. 13, p. 1139-1147, 2013.

HILL, R. S.; POLE, M. S. Leaf and Shoot Morphology of Extant Afrocarpus, Nageia and Retrophyllum (Podocarpaceae) Species, and Species with Similar Leaf Arrangement, from Tertiary Sediments in Australasia. **Australian Systematic Botany,** v. 5, p. 337-358, 1992.

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

HIRAO, T.; WATANABE, A.; KURITA, M.; KONDO, T.; TAKATA, K. 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, M.; FRISO, G.; NISHIMURA, K.; QU, X.; OLINARES, P. D.; MAJERAN, W.; SUN, Q.; VAN WIJK, K. J. Construction of plastid reference proteomes for maize and *Arabidopsis* and evaluation of their orthologous relationships; the concept of orthoproteomics. **Journal of Proteome Research**, v. 12, p. 491–504, 2013.

JANSEN, R. K.; RAUBESON, L. A.; BOORE, J. L.; DEPAMPHILIS, C. W.; CHUMLEY, T. W. et al. Methods for obtaining and analyzing whole chloroplast genome sequences. **Methods of Enzymology**, v. 395, p. 348-384, 2005.

JANSEN, R. K.; KAITTANIS, C.; SASKI, C.; LEE, S.; TOMKINS, J.: ALVERSON, A. J.; DANIELL, H. Phylogenetic analyses of *Vitis* (Vitaceae) based on complete chloroplast genome sequences: effects of taxon sampling and phylogenetic methods on resolving relationships among rosids. **BMC Evolutionary Biology**, v. 6, p. 1-14, 2006.

JANSEN, R. K., CAI, Z.; RAUBESON, L. A.; DANIELL, H.; DEPAMPHILIS, C. W. 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 USA**, v. 104, p. 19369–19374, 2007.

KANEKO, T.; SATO, S.; KOTANI, H.; TANAKA, A.; ASAMIZU, E. et al. Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803: II. Sequence determination of the entire genome and assignment of potential protein-coding regions. **DNA Research**, v. 3, p. 109–136, 1996.

KEELING, C. I.; BOHLMANN, J. Diterpene resin acids in conifers. **Phytochemistry**, 67: 2415–2423, 2006.

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

KELCH, D. G. Phylogenetic assessment of the monotypic genera Sundacarpus and Manoao (Coniferales: Podocarpaceae) utilising evidence from 18S rDNA sequences. **Australian Systematic Botany**, v. 15, p. 29–35, 2002.

KIKUCHI, S.; BÉDARD, J.; HIRANO, M.; HIRABAYASHI, Y.; OISHI, M. et al. Uncovering the protein translocon at the chloroplast inner envelope membrane. **Science**, 339: 571–574, 2013.

KUBIS, S. E.; LILLEY, K. S.; JARVIS, P. Isolation and preparation of chloroplast from *Arabidopsis thaliana* plants. In: Posch A, editor. **2D-PAGE: sample preparation and fractionation**: Humana Press. pp. 171-183, 2008.

de LAUBENFELS, D. J. D. A taxonomic revision of the genus *Podocarpus*. **Blumea**, v. 30, p. 251–278, 1985.

LEDRU, M.; SALATINO, M. L. F.; CECCANTINI, G.; SALATINO, A.; PINHEIRO, F. et al. Regional assessment of the impact of climatic change on the distribution of a tropical conifer in the lowlands of South America. **Diversity and Distributions**, v. 13, p. 761–771, 2007.

LIN, C. P.; HUANG, J. P.; WU, C. S.; HSU, C. Y.; CHAW, S. M. Comparative chloroplast genomics reveals the evolution of Pinaceae genera and subfamilies. **Genome Biology Evolution**, v. 2, p. 504–517, 2010.

LISBOA P. L. B.; ALMEIDA, S. S. Vida e obra de João Murça Pires (1917-1994). **Acta Botanica Brasilica,** v. 9, p. 303-314, 1995.

LOGACHEVA, M. D.; PENIN, A. A.; SAMIGULLIN, T. H.; VALLEJO-ROMAN, C. M.; ANTONOV, A. S. Phylogeny of flowering plants by the chloroplast genome sequences: in search of a “Lucky Gene”. **Biochemistry**, v. 72, p. 1324-1330, 2007.

LOHSE, M.; DRECHSEL, O.; KAHLAU, S.; BOCK, R. 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. W575-81, 2013.

LONGHI, S. J.; BRENA, D. A.; RIBEIRO, S. B.; GRACIOLI, C. R.; LONGHI, R. V., et al. Fatores ecológicos deteminantes na ocorrência de *Araucaria angustifolia* e *Podocarpus lambertii* na floresta Ombrófila mista da FLONA de São Francisco de Paula, RS, Brasil. **Ciência Rural**, Santa Maria, v. 40, p. 57-63, 2010.

KURTZ, S.; PHILLIPPY, A.; DELCHER, A. L.; SMOOT, M.; SHUMWAY, M.; et al. Versatile and open software for comparing large genomes. **Genome Biology**, v. 5, n. R12, 2004.

LIN, C.; HUANG, J.; WU, C.; HSU, C.; CHAW, S. Comparative chloroplast genomics reveals the evolution of Pinaceae genera and subfamilies. **Genome Biology Evolution**, v. 2, p. 504–517, 2010.

MAIER, R. M.; NECKERMANN, K.; IGLOI, G. L.; KÖSSEL, H. Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. **Journal Molecular Biology**, v. 251, p. 614-628, 1995.

MALIGA, P.; BOCK, R. Plastid biotechnology: food, fuel and medicine for the 21st century. **Plant Physiology**, v. 155, p. 1501–1510, 2011.

MARÉCHAL, A.; BRISSON, N. Recombination and the maintenance of plant organelle genome stability. **New Phytologist**, v. 186, p. 299-317, 2010.

MARTIN, W.; DEUSCH, O.; STAWSKI, N.; GRUNHEIT, N.; GOREMYKIN, V. Chloroplast genome phylogenetics: why we need independent approaches to plant molecular evolution. **Trends in Plant Science**, v. 10, p. 203-209, 2005.

MASTROBERTI, A. A.; MARIATH, J. E. A. Leaf anatomy of *Araucaria angustifolia* (Bertol.) Kuntze (Araucariacea). **Revista Brasileira de Botânica**, v. 26, p. 343-353, 2003.

MATSUOKA, Y.; YAMAZAKI, Y.; OGIHARA, Y.; TSUNEWAKI, K. Whole chloroplast genome comparison of rice, maize, and wheat: implications for chloroplast gene diversification and phylogeny of cereals. **Molecular Biology and Evolution**, v. 19, p. 2084–2091, 2002.

MOORE, M. J.; BELL, C.D.; SOLTIS, P.S.; SOLTIS, D.E. Using plastid genomic-scale data to resolve enigmatic relationships among basal angiosperms. **Proceedings of the National Academy of Sciences USA**, v. 104, p. 19363 – 19368, 2007.

MOORE, M. J.; SOLTIS, P.S.; BELL, C. D.; BURLEIGH, J. G.; SOLTIS, D. E. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. **Proceedings of the National Academy of Sciences USA**, v. 107, p. 4623–4628, 2010.

MORLEY, R. J. Dispersal and paleoecology of tropical podocarps. In: TURNER, B. L.; CERNUSAK, L. A.; editors. **Ecology of the Podocarpaceae in tropical forests**. Washington: Smithsonian Institute Scholarly Press. pp. 21-42, 2011.

MURRAY, B. G. Karyotype variation and evolution in gymnosperms. In LEITCH, I. J.; GREILHUBER, J.; DOLEZEL, J.; WENDEL, J. F. editors. **Plant genome diversity 2: physical structure, behaviour and evolution of plant genomes**. Springer. pp 231-243, 2013.

MYERS, N.; MITTERMEIER, R. A.; MITTERMEIER, C. G.; FONSECA, G. A. B.; KENT, J. Biodiversity hotspots for conservation priorities. **Nature**, v. 403, p. 853-858, 2000.

NEALE, D. B.; MARSHALL, K. A.; SEDEROFF, R. R. Chloroplast and mitochondrial DNA are paternally inherited in *Sequoia sempervirens* D. Don Endl. **Proceedings of the National Academy of Sciences**, v. 86, p. 9347-9349, 1989.

PAGE, C. N. Coniferophytina. In: KRAMER, K. U.; GREEN, P.S. editors. **The families and genera of vascular plants, Pteridophytous and Gymnosperms**: Springer. pp. 332-346, 1990.

PALMER, J. D.; THOMPSON, W. F. Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. **Cell**, v. 29, p. 537-550, 1982.

PALMER, J. D. Chloroplast DNA and molecular phylogeny. **BioEssays**, v. 2, p. 263-267, 1985.

PALMER, J. D.; SEIN, D. B. Conservation of chloroplast genome structure among vascular plants. **Current Genetics**, v. 11, p. 823-834, 1986.

PALMER, J. D.; JANSEN, R. K.; MICHAELS, H. J.; CHASE, M. W.; MANHART, J.R. Chloroplast DNA variation and plant phylogeny. **Annals Missouri Botany Garden**, v. 75, p. 1180-1206, 1988.

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

PROVAN, J.; CORBETT, G.; MCNICOL, J. W.; POWELL, W. Chloroplast variability in wild and cultivated rice (*Oryza* spp.) revealed by polymorphic chloroplast simple sequence repeats. **Genome**, v. 40, p. 104–110, 1997.

PROVAN, J.; RUSSELL, J. R.; BOOTH, A.; POWELL, W. Polymorphic chloroplast simple-sequence repeat primers for systematic and population studies in the genus *Hordeum*. **Molecular Ecology**, v. 8, p. 505-511, 1999.

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

ROGALSKI, M.; SCHOETTLER, M. A.; THIELE, W.; SCHULZE, W. X.; BOCK, R. Rpl33, a nonessential plastid encoded ribosomal protein in tobacco, is required under cold stress conditions. **Plant Cell**, v. 20, p. 2221–2237, 2008.

ROGALSKI, M.; CARRER, H. Engineering plastid fatty acid biosynthesis to improve food quality and biofuel production in higher plants. **Plant Biotechnology Journal**, v. 9, p. 554–564, 2011.

ROUSSEAU-GUEUTIN, M.; AYLIFFE, M. A.; TIMMIS, J. N. Conservation of plastid sequences in the plant nuclear genome for millions of years facilitates endosymbiotic evolution. **Plant Physiology**, v. 157, p. 2181–2193, 2011.

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. W686-689, 2005.

SCHMIDT, M.; SCHNEIDER-POETSCH, H. A. The evolution of gymnosperms redrawn by phytochrome genes: the Gnetatae appear at the base of the gymnosperms. **Journal of Molecular Evolution**, v. 54, p. 715-724, 2002.

SETOGUCHI, H.; OSAWA, T. A.; PINTAUD, J.; JAFFRÉ, T.; VEILLON, J. Phylogenetic relationships within Araucariaceae based on *rbcL* gene sequences. **American Journal of Botany**, v. 85, p. 1507–1516, 1998.

SHAW, J.; LICKEY, E. B.; SCHILLING, E. E.; SMALL, R. L. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and hare III. **American Journal of Botany**, v. 94, p. 275–288. 2007.

SHI, C.; HU, N.; HUANG, H.; GAO, J.; ZHAO, Y. et al. An improved chloroplast DNA extraction procedure for whole plastid genome sequencing. **PLoS ONE**, v. 7, n. e31468, 2012.

SOLL, J.; SCHLEIFF, E. Protein import into chloroplasts. **Nature Reviews**, v. 5, p. 198–208, 2004.

STRAUSS, S. H.; PALMER, J. D.; HOWE, G. T.; DOERKSEN, A. H. Chloroplast genomes of two conifers lack a large inverted repeat and are extensively rearranged. **Proceedings of the National Academy of Sciences of USA**, v. 85, p. 3898-3902, 1988.

SUGIURA, C.; SUGITA, M. Plastid transformation reveals that moss *tRNA*(Arg)-CCG is not essential for plastid function. **The Plant Journal**, v. 40, p. 314-321, 2004.

TERAKAMI, S.; MATSUMURA. Y.; KURITA, K.; KANAMORI, H.; KATAYOSE, Y. et al. Complete sequence of the chloroplast genome from pear (*Pyrus pyrifolia*): genome structure and comparative analysis. **Tree Genetics and Genomes**, v. 8, p. 841–854, 2012.

TEWARI, K. K.; WILDMAN, S. G. Chloroplast DNA from tobacco leaves. **Science**, v. 153, p. 1269-1271, 1966.

TIMMIS, J. N.; AYLIFFE, M. A.; HUANG, C. Y.; MARTIN, W. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. **Nature Reviews Genetics**, v. 5, p. 123–135, 2004.

TSUDZUKI, J.; NAKASHIMA, K.; TSUDZUKI, T.; HIRATSUKA, J.; SHIBATA, M. 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 and General Genetics**, v. 232, p. 206–214, 1992.

YAMAMOTO, E. S.; OTTO, A.; SIMONEIT, B. R. T. Lignans and resin of *Araucaria angustifolia* by gas chromatography/mass spectrometry. **Journal of Mass Spectrometry**, v. 39, p. 1337-1347, 2004.

YI, X.; GAO, L.; WANG, B.; SU, Y.; WANG, T. 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 Evolution**, v. 5, p. 688–698, 2013.

VANYUSHIN, B.F. **DNA methylation in plants**. Berlin: Springer-Verlag, 2006.

VIEIRA, L. N.; FAORO, H.; FRAGA, H. P. F.; ROGALSKI, M.; SOUZA, E. M. et al. (2014) An improved protocol for intact chloroplasts and cpDNA isolation in conifers. **PLoS ONE**, v. 9, n. e84792, 2014.

WAKASUGI, T.; TSUDZUKI, J.; ITO, S.; NAKASHIMA, K.; TSUDZUKI, 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 USA**, v. 91, p. 9794–9798, 1994.

WALKER, D. A.; CEROVIC, Z. E. G.; ROBINSON, S. P. Isolation of intact chloroplasts: general principles and criteria of integrity. **Methods Enzymology**, v. 148, p. 145-157, 1987.

WEIXLBAUMER, A.; MURPHY, F. V.; DZIERGOWSKA, A.; MALKIEWICZ, A.; VENDEIX, F. A. et al. Mechanism for expanding the decoding capacity of transfer RNAs by modification of uridines. **Nature structural and molecular biology**, v. 14, p. 498-502, 2007.

WERNER, O.; PATINO, J.; GONZÁLES-MANCEBO, J. M.; GABRIEL, R. M. D.; ROS, R. M. The taxonomic status and the geographical relationships of the Macaronesian endemic moss *Fissidens luisieri* (Fissidentaceae) based on DNA sequence data. **Bryologist**, v. 112, p. 315–324, 2009.

WICKE, S.; SCHNEEWEISS, G. M.; DEPAMPHILIS, C. W.; MÜLLER, K. F.; QUANDT, D. Evolution of the plastid chromosome in land plants: gene content, gene order, gene function. **Plant Molecular Biology**, v, 76, p. 273–297, 2011.

WOLF, P. G.; ROPER, J. M.; DUFFY, A. M. The evolution of chloroplast genome structure in ferns. **Genome**, v. 53, p. 731–738, 2010.

WOLFE, K. H.; LI, W. H.; SHARP, P. M. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. **Proceedings of the National Academy of Sciences**, v. 84, p. 9054-9058, 1987.

WOLFE, K. H.; MORDEN, C. W.; PALMER, J. D. Function and evolution of a minimal plastid genome from a non-photosynthetic parasitic plant.

Proceedings of the National Academy of Sciences USA, v. 89, p. 10648-10652, 1992.

WU, C. S.; WANG, Y. N.; LIU, S. M.; CHAW, S. M. 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 Evolution**, v. 24, p. 1366–1379, 2007.

WU, C. S.; LAI, Y. T.; LIN, C. P.; WANG, Y. N.; CHAW, S. M. Evolution of reduced and compact chloroplast genomes (cpDNAs) in gnetophytes: selection towards a lower cost strategy. **Molecular Phylogenetics and Evolution**, v. 52, p. 115–124, 2009.

WU, C. S.; WANG, Y. N.; HSU, C. Y.; LIN, C. P.; CHAW, S. M. 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 Evolution**, v. 3, p. 1284-1295, 2011a.

WU, C. S.; LIN, C. P.; HSU, C. Y.; WANG, R. J.; CHAW, S. M. Comparative chloroplast genomes of Pinaceae: insights into the mechanism of diversified genomic organizations. **Genome Biology and Evolution**, v. 3, p. 309-319, 2011b.

WU, C. S.; CHAW, S. M. Highly rearranged and size-variable chloroplast genomes in conifer II clade (cupressophytes): evolution towards shorter intergenic spacers. **Plant Biotechnology Journal**, doi: 10.1111/pbi.12141, 2013.

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

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