



Universidade Federal de Goiás
Instituto de Ciências Biológicas



Programa de Pós-graduação em Genética e Biologia Molecular

LEONARDO CARLOS JERONIMO CORVALAN

**Genoma cloroplastidial de *Serjania erecta* Raldk: Variação no número
de genes e análise de seleção de genes de plastomas da família
Sapindaceae**

Orientador: Dr. Rhewter Nunes

Coorientador: Dr. José Alexandre Felizola Diniz-Filho

GOIÂNIA - GO

2022



UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO (TECA) PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TESES

E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a [Lei 9.610/98](#), o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

O conteúdo das Teses e Dissertações disponibilizado na BDTD/UFG é de responsabilidade exclusiva do autor. Ao encaminhar o produto final, o autor(a) e o(a) orientador(a) firmam o compromisso de que o trabalho não contém nenhuma violação de quaisquer direitos autorais ou outro direito de terceiros.

1. Identificação do material bibliográfico

Dissertação Tese

2. Nome completo do autor

Leonardo Carlos Jeronimo Corvalan

3. Título do trabalho

Genoma cloroplastidial de *Serjania erecta* Raldk: Variação no número de genes e análise de seleção de genes de plastomas da família Sapindaceae

4. Informações de acesso ao documento (este campo deve ser preenchido pelo orientador)

Concorda com a liberação total do documento SIM NÃO¹

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. Após esse período, a possível disponibilização ocorrerá apenas mediante:

a) consulta ao(à) autor(a) e ao(à) orientador(a);

b) novo Termo de Ciência e de Autorização (TECA) assinado e inserido no arquivo da tese ou dissertação.

O documento não será disponibilizado durante o período de embargo.

Casos de embargo:

- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;
- Publicação da dissertação/tese em livro.

Obs. Este termo deverá ser assinado no SEI pelo orientador e pelo autor.



Documento assinado eletronicamente por **LEONARDO CARLOS JERONIMO CORVALAN, Discente**, em 22/03/2022, às 14:40, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **RHEWTER NUNES, Usuário Externo**, em 22/03/2022, às 14:40, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **2775774** e o código CRC **95084A0A**.



UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO (TECA) PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TESES

E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a [Lei 9.610/98](#), o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

O conteúdo das Teses e Dissertações disponibilizado na BDTD/UFG é de responsabilidade exclusiva do autor. Ao encaminhar o produto final, o autor(a) e o(a) orientador(a) firmam o compromisso de que o trabalho não contém nenhuma violação de quaisquer direitos autorais ou outro direito de terceiros.

1. Identificação do material bibliográfico

Dissertação Tese

2. Nome completo do autor

Leonardo Carlos Jeronimo Corvalan

3. Título do trabalho

Genoma cloroplastidial de *Serjania erecta* Raldk: Variação no número de genes e análise de seleção de genes de plastomas da família Sapindaceae

4. Informações de acesso ao documento (este campo deve ser preenchido pelo orientador)

Concorda com a liberação total do documento SIM NÃO¹

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. Após esse período, a possível disponibilização ocorrerá apenas mediante:

a) consulta ao(à) autor(a) e ao(à) orientador(a);

b) novo Termo de Ciência e de Autorização (TECA) assinado e inserido no arquivo da tese ou dissertação. O documento não será disponibilizado durante o período de embargo.

Casos de embargo:

- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;
- Publicação da dissertação/tese em livro.

Obs. Este termo deverá ser assinado no SEI pelo orientador e pelo autor.



Documento assinado eletronicamente por **Leonardo Carlos Jeronimo Corvalán, Discente**, em 01/03/2024, às 15:33, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site

https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0,
informando o código verificador **4421215** e o código CRC **0F3B76D6**.

Referência: Processo nº 23070.010495/2022-71

SEI nº 4421215

Leonardo Carlos Jeronimo Corvalan

Genoma cloroplastidial de *Serjania erecta* Raldk: Variação no número de genes e análise de seleção de genes de plastomas da família Sapindaceae

Dissertação apresentada ao Programa de Pós-Graduação em Genética e Biologia Molecular, da Universidade Federal de Goiás, como parte dos requisitos da obtenção do título de Mestre em Genética e Biologia Molecular.

Área de concentração: Genética e Biologia Molecular

Linha de pesquisa: Genômica funcional, estrutural e proteômica

Orientador: Dr. Rhewter Nunes

Coorientadora: Dr. José Alexandre F. Diniz-Filho

Goiânia - GO

2022

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Corvalan, Leonardo Carlos Jeronimo

Genoma cloroplastidial de *Serjania erecta* Raldk: Variação no número de genes e análise de seleção de genes de plastomas da família Sapindaceae [manuscrito] / Leonardo Carlos Jeronimo Corvalan. - 2022.

61 f.: il.

Orientador: Prof. Dr. Rhewter Nunes; co-orientador Dr. José Alexandre Felizola Diniz-Filho.

Dissertação (Mestrado) - Universidade Federal de Goiás, Instituto de Ciências Biológicas (ICB), Programa de Pós-Graduação em Genética e Biologia Molecular, Goiânia, 2022.

Bibliografia. Anexos.

Inclui siglas, gráfico, tabelas.

1. Análise comparativa. 2. Genoma organelar. 3. Seleção negativa.
I. Nunes, Rhewter, orient. II. Título.

CDU 57



UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
ATA DE DEFESA DE DISSERTAÇÃO

Ata nº **101** da sessão de Defesa de Dissertação de **Leonardo Carlos Jeronimo Corvalan**, que confere o título de Mestre(a) em **Genética e Biologia Molecular**, na área de concentração em **Genética e Biologia Molecular**

Ao/s **quatro dias do mês de março de 2022**, a partir da(s) **08h00**, via webconferência, realizou-se a sessão pública de Defesa de Dissertação intitulada “**Genoma cloroplastidial de Serjania erecta Raldk: Variação no número de genes e análise de seleção de genes de plastomas da família Sapindaceae**”. Os trabalhos foram instalados pelo(a) Orientador(a), Professor(a) Doutor(a) **Rhewter Nunes (UFG)** com a participação dos demais membros da Banca Examinadora: Professor(a) Doutor(a) **Mariane Brom Sobreiro (UFG)**, membro titular externo; Professor(a) Doutor(a) **Renata de Oliveira Dias (ICB/UFG)**, membro titular interno. Durante a arguição os membros da banca **não fizeram** sugestão de alteração do título do trabalho. A Banca Examinadora reuniu-se em sessão secreta a fim de concluir o julgamento da Dissertação, tendo sido(a) o(a) candidato(a) **aprovado(a)** pelos seus membros. Proclamados os resultados pelo(a) Professor(a) Doutor(a) **Rhewter Nunes (UFG)**, Presidente da Banca Examinadora, foram encerrados os trabalhos e, para constar, lavrou-se a presente ata que é assinada pelos Membros da Banca Examinadora, ao(s) **quatro dias do mês de março de 2022**.

TÍTULO SUGERIDO PELA BANCA



Documento assinado eletronicamente por **RHEWTER NUNES, Usuário Externo**, em 04/03/2022, às 15:11, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Renata De Oliveira Dias, Professora do Magistério Superior**, em 04/03/2022, às 16:38, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **MARIANE BROM SOBREIRO, Usuário Externo**, em 05/03/2022, às 15:09, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **2735406** e o código CRC **08C1A140**.

AGRADECIMENTOS

Primeiramente, agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) que me concedeu a bolsa de nível mestrado permitindo minha estadia em Goiânia. Também agradeço ao Instituto Nacional de Ciência e Tecnologia em Ecologia, Evolução e Conservação da Biodiversidade (INCT – EECBio), no contexto do grupo de trabalho de Genética e Genômica Evolutiva, subsidiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq – processo 465610/20145) e Fundação de Amparo à Pesquisa do Estado de Goiás (FAPEG – processo 201810267000023), que garantiu toda infraestrutura necessária para o desenvolvimento desse trabalho.

Agradeço a todos os colaboradores desse trabalho, Mariane B. Sobreiro, Larissa Carvalho, Cinti P. Targueta, Ramilla dos S. B. Braga-Ferreira, José A. F. Diniz-filho, Mariana P. de C. Telles e Rhewter Nunes, por toda a dedicação e paciência em me ensinar. Sem vocês a realização desse trabalho seria impossível.

Obrigado a todos os professores e funcionários do programa de Pós-graduação em Genética e Biologia Molecular da Universidade federal de Goiás, em especial a professora Renata de Oliveira Dias pela orientação no estágio de docência e por toda ajuda na bioinformática. Agradeço imensamente ao meu orientador Rhewter Nunes por me guiar nessa jornada do mestrado, pela paciência em me ensinar e por todo conhecimento compartilhado. Gostaria também de agradecer aos meus coorientadores José A. F. Diniz-filho e Mariana P. de C. Telles por compartilharem toda sua experiencia no delineamento desse trabalho.

Aos meus amigos do Laboratório de Genética & Biodiversidade e aos amigos de aula online, Victoria Albernaz, Victor Felipe, Larissa Carvalho, Aline Menegat, Laís Souza Santos e Sheilline Oliveira, que mesmo com distanciamento social sempre me ajudaram durante todo o mestrado.

Agradeço a minha mãe Rosa Maria Jeronimo por sempre me apoiar e motivar todos os dias. Agradeço ao meu pai Carlos Alonso Corvalan Soto, aos meus avôs Américo Hissa Satuf e Carlos Zorobabel Corvalan Gajardo e minhas avós Ildeia Navara Satuf, Maria Francisca Gerônimo e Maria Soto Lucero., meus padrinhos Eduardo Satuf e Mariangela Satuf por todo apoio aos meus estudos.

Não poderia deixar de agradecer a minha família de Goiânia, Abel Vieira de Melo Bisneto, Andrea Maciel e meu irmão Rafael Alonso Jeronimo Corvalan, por sempre estarem presente e todo auxilio emocional.

SUMÁRIO

RESUMO	10
ABSTRACT	11
1. INTRODUÇÃO GERAL	12
2. OBJETIVOS	17
2.1 Objetivo geral	17
2.2 Objetivos específicos	17
APENDICE I – MANUSCRITO	18
CONCLUSÕES FINAIS	57
REREFÊNCIAS	58

RESUMO

Serjania erecta é uma planta da família Sapindaceae utilizada na medicina popular e que apresenta potencial uso no tratamento da doença de Alzheimer e algumas doenças gástricas. Embora seu potencial medicinal seja reconhecido, pouco se sabe sobre aspectos genéticos e evolutivos da espécie. Neste âmbito, o objetivo do presente trabalho foi sequenciar e montar o genoma cloroplastidial de *S. erecta*, bem como, realizar análises comparativas com plastomas de outras espécies da família Sapindaceae. Para isso foi sequenciado o DNA total de um indivíduo de *S. erecta* coletado em Araxá (MG - Brasil), utilizando plataforma MiSeq (Illumina). O plastoma foi montado utilizando o programa NOVOPlasty v3.2 e anotado utilizando o CHLOROBOX. Para as análises comparativas, foram utilizados genomas cloroplastidiais de 11 espécies de gêneros diferentes da família Sapindaceae (*Acer buergerianum*, *Aesculus wangii*, *Dimocarpus longan*, *Dipteronia dyeriana*, *Dodonaea viscosa*, *Eurycorymbus cavaleriei*, *Koelreuteria paniculata*, *Litchi chinensis*, *Pometia tomentosa*, *Sapindus mukorossi* e *Xanthoceras sorbifolium*). O plastoma de *S. erecta* apresentou um tamanho de 159,297 bp com 132 genes, destes 87 são genes codificadores de proteínas, 37 de tRNAs e 8 rRNAs. A estrutura e ordem dos genes nos genomas cloroplastidiais da ordem Sapindales foi relativamente conservada variando de 128 genes a 132. Foi possível identificar três fatores que causam a variação no número de genes: (1) o aumento das regiões invertidas repetidas (IR) que causam a duplicação dos genes *rpl22*, *rps3* e *rps19*; (2) a pseudogenização do gene *rps2*; (3) variação no número de genes codificadores de tRNAs. A análise filogenética utilizando sequências codificadoras de proteínas do cloroplasto apresentou nós bem suportados dentro de Sapindaceae e *Serjania* formou um clado com os gêneros *Sapindus*, *Litchi*, *Dimocarpus* e *Pometia*. Apenas duas regiões genicas (*ycf1* e *ndhF*) apresentaram elevada diversidade nucleotídica e nenhum gene está sobre seleção positiva ($ka/ks > 1$). Os resultados obtidos neste estudo fornecem a montagem e anotação do genoma do cloroplasto de *S. erecta*, a primeira anotação de uma espécie desse gênero e fornece uma ideia de como os genomas cloroplastidiais evoluíram na família Sapindaceae.

Palavras chaves: Análise comparativa; Genoma organelar; seleção negativa.

ABSTRACT

Serjania erecta from the Sapindaceae family is a plant with medicinal properties. Studies indicate potential for use in the treatment of Alzheimer's disease, gastric diseases and anti-inflammatory use. Although, little is known about its genetic and evolutionary aspects. The goal of this study was to assemble the chloroplast genome of *S. erecta*, and use it in a comparative analysis with other plastomas of the Sapindaceae family. For this, we sequenced a single specimen of *S. erecta* from Araxá (MG - Brasil) using Illumina Miseq. The chloroplast genome was assembled using NOVOPlasty v3.2 and annotated using the CHLOROBOX platform. For comparative analysis was used eleven chloroplast genomes from different Sapindaceae family species (*Acer buergerianum*, *Aesculus wangii*, *Dimocarpus longan*, *Dipteronia dyeriana*, *Dodonaea v iscosa*, *Eurycorymbus cavaleriei*, *Koelreuteria paniculata*, *Litchi chinensis*, *Pometia tomentosa*, *Sapindus mukorossi*, *Xanthoceras sorbifolium*). The chloroplast genome of *S. erecta* has a size of 159,297 bp with 132 genes, including 87 are protein-coding genes, 37 are tRNAs and 8 rRNAs. Among twelve chloroplast genomes avalieded, *S. erecta* has the lowest amount of complex repeats and microsatellites. The structure and order of genes in chloroplast genomes of the order Sapindales was extremely conserved. The variation in numbers of genes was from the 132 genes to 128 genes in the Sapindaceae family. We suggest that three factors cause variation in the number of genes in the family: (1) Inverted repeat region (IR) expansion events cause the duplication of the *rpl22*, *rps3* and *rps19* genes; (2) the pseudogenization of the *rps2* gene; (3) variation in the number of genes encoding tRNAs. The phylogenetic tree had well supported nodes within Sapindaceae and *Serjania* formed a clade with *Sapindus*, *Litchi*, *Dimocarpus*, and *Pometia* genera. Only two genomic regions (*ycf1* and *ndhF*) showed high nucleotide diversity and no one gene is under positive selection ($ka/ks > 1$). The results obtained in this study provide the assembly and annotation of the chloroplast genome of *S. erecta*, the first annotation of a species of the genus. It also provides an idea of how chloroplast genomes evolved in the Sapindaceae family.

Keywords: Comparative analysis; negative selection; organellar genome.

1. INTRODUÇÃO GERAL

A família Sapindaceae é a segunda família mais diversa entre as nove famílias da ordem Sapindales, com cerca de 1900 espécies distribuídas em 142 gêneros (ACEVEDO-RODRÍGUEZ et al., 2011). Estima-se que a família surgiu cerca de 87 Ma durante o período Cretáceo superior em regiões da eurásia, dispersando-se pelo hemisfério sul durante o Paleoceno (BUERKI et al., 2013; MUELLNER-RIEHL et al., 2016). Atualmente, mais de 80% de biodiversidade da família se localiza em regiões tropicais e subtropicais do hemisfério sul, em especial 436 espécies são encontradas no Brasil e 191 são espécies endêmicas do Brasil (BUERKI et al., 2021; REFLORA, 2020a).

Essa família também é conhecida como família *soapberry* (fruto-sabão) e abriga diversas plantas de importância alimentícia tais como guaraná (*Paullinia cupana* Kunth), lichia (*Litchi chinensis* Sonn), longan (*Dimocarpus longan* Lour.), pitomba (*Talisia esculenta* Radlk.) e rambutan (*Nephelium lappa* ceum L.); e também espécies de importância madeireira como buckeyes (gênero *Aesculus*) e Fijian logan (*Pometia pinnata* JRForst. & G.Forst). Além disso, também são encontradas nessa família espécies com potencial uso medicinal como em plantas do gênero *Serjania*, *Sapindus* e *Paullinia* (BUERKI et al., 2021; GOMIG et al., 2008; GOYAL, 2014; LIDILHONE; GENISE; NEUSA, 2013).

Descrita por Radlkofer (1933), a família Sapindaceae foi revisada com a utilização de marcadores cloroplastidiais (*matK* e *rbcL*) e passou a incluir as antigas famílias Aceraceae e Hippocastanaceae e o gênero *Xanthoceras* (HARRINGTON et al., 2005). As filogenias moleculares mais recentes com a utilização de mais marcadores e inclusão de marcadores nucleares, confirmam a topologia proposta por Harrington (2005) (BUERKI et al., 2009, 2021; MUELLNER-RIEHL et al., 2016). Atualmente, são descritas 20 tribos distribuídas em quatro subfamílias, 2 tribos na subfamília Hippocastanoideae, 2 tribos na Dodonaeoideae e 16 tribos na Sapindoideae. Na subfamília Xanthoceratoideae não foram classificados tribos até então (BUERKI et al., 2021).

Xanthoceratoideae é subfamília mais basal da família Sapindaceae, sendo composta por apenas um gênero (*Xanthoceras*) com áreas de ocorrência na China e Coreias (BUERKI et al., 2021). A subfamília Hippocastanoideae é composta pelas antigas famílias Aceraceae e Hippocastanaceae, apresentando 5 gêneros distribuídos predominantemente em regiões de clima temperado com algumas linhagens em zonas tropicais da América do Sul (BUERKI et al., 2021; HARRINGTON et al., 2005). A

subfamília Dodonaeoideae é composta por 140 espécies e 24 gêneros. A subfamília com maior número de espécies é a Sapindoideae com 107 gêneros e aproximadamente 1400 espécies distribuídas em regiões tropicas e subtropicais, nesta subfamília destaca-se a tribo Paullinieae que representa mais de 50% das Sapindaceae do novo mundo (ACEVEDO-RODRÍGUEZ et al., 2011; GUIMARÃES; FORNI-MARTINS, 2021).

Pertencente, à tribo Paullinieae, o gênero *Serjania* é composto por 230 espécies que são majoritariamente trepadeiras lenhosas perenes e que são caracterizadas pela presença de frutos esquizocárpicos separados em três mericarpos samarídeos com um lóculo distal (ACEVEDO-RODRIGUEZ, 1990; ACEVEDO-RODRÍGUEZ, 1993). Quanto à morfologia das sementes das espécies pertencentes ao gênero apresentam mericarpo altamente variável, sendo dispersas pelo vento e ou por correntes de água, fazendo com que espécies do gênero sejam encontradas desde o sudoeste dos Estados Unidos até o norte da Argentina (ACEVEDO-RODRIGUEZ, 1990; VAN DER HAM; TOMLIK, 1994). Todas as espécies do gênero com cariótipo conhecido são espécies diploides com 24 cromossomos ($2n=24$), os quais podem apresentar variações na simetria de seus cromossomos formando grupos discordantes da divisão taxonômica, proposta por Acevedo-Rodríguez (1993) (GUIMARÃES; FORNI-MARTINS, 2021; URDAMPILLETA; FERRUCCI; VANZELA, 2012).

Estudos de filogenia molecular indicam que o gênero *Serjania* é parafilético formando um clado com espécies dos gêneros *Balsas*, *Chimborazoa* e *Houssayanthus* (ACEVEDO-RODRÍGUEZ et al., 2017). Já a classificação taxonômica clássica das espécies do gênero é dificultada devido a morfologia de seu pólen não apresentar grandes variações, tornando complexo estabelecer relações dentro do gênero (FERRUCCI; ACEVEDO-RODRÍGUEZ, 2005; VAN DER HAM; TOMLIK, 1994). Os atuais dados moleculares disponíveis para gênero não conseguem reconstruir a história evolutiva das espécies resultando em extensas politomias, fazendo necessário ampliar o conhecimento genômico do gênero (ACEVEDO-RODRÍGUEZ et al., 2017).

Serjania erecta Ralk é popularmente conhecida como cipó cinco folhas, cinco folhas ou retrato de teiú, devido a morfologia suas folhas, com 5 folíolos estípulas triangulares e pecíolo alado sempre mais comprido que a raque (REFLORA, 2020b). Também é conhecida como timbó-bravo, timbó e tururi, nomes de origem indígena usado para designar cipós de Fabaceae e Sapindaceae que seu consumo pode causar intoxicação (PEREIRA; FAMADAS, 2004).

Tradicionalmente as folhas de *S. erecta* são utilizadas no tratamento de inflamação, dor de estômago e de úlceras. Já as cascas são tipicamente utilizadas no tratamento de hipertensão (GUARIM NETO, 2009; GUARIM NETO; SANTANA; DA SILVA, 2000). Os aspectos medicinais dessa espécie vem sendo avaliado, indicando potencial utilização no tratamento da doença de Alzheimer, efeito antibacteriano, ação gastroprotetora e efeito antiofídico, à peçonha de *Bothrops jararacussu* e *Bothrops neuwiedi* (ARRUDA, 2011; BROGGINI et al., 2010; CARDOSO et al., 2013; FERNANDES et al., 2011; GUIMARÃES et al., 2015). Acredita-se que esta sua utilização medicinal está associado a metabólicos secundários, como saponinas, terpenos, flavonóides e taninos, contidos nas folhas e nas raízes do vegetal (ARRUDA, 2011; BROGGINI et al., 2010; CARDOSO et al., 2013; FERNANDES et al., 2011; GUIMARÃES et al., 2015). Além desta utilização medicinal, *S. erecta* pode ser utilizado como inseticida botânico no controle de *Chrysodeixis includens* (Lepidoptera) em plantações de soja, podendo o extrato de suas folhas ser uma alternativa ao uso de agrotóxicos (FREITAS, 2018).

As flores de zigomorfas dispostas em inflorescências paniculadas de *S. erecta* tem um tamanho médio de 10 mm de diâmetro (LENZA; FERREIRA, 2000). A espécie apresenta dioiccia temporal, com flores masculinas brevestioladas e flores femininas longistioladas podendo ser encontrado indivíduos com ambas as flores ou com apenas um tipo de flor (LENZA; FERREIRA, 2000). São relatados como visitantes florais: Dípteros, Himenópteros, Lepidópteros e Hemípteros, os quais são recompensados com néctar nas flores masculinas e sementes nas flores femininas (LENZA; FERREIRA, 2000). A espécie apresenta frutos samaroides e suas sementes são aladas, sendo dispersas pelo vento (PERES, 2016). Poucos estudos ecológicos de *S. erecta* foram desenvolvidos nos últimos anos e nenhum estudo de genética molecular da espécie foi desenvolvido até então.

Em geral, mesmo com a grande biodiversidade e importância econômica da família Sapindaceae, apenas quatro espécies (*Acer yangbiense*, *Dimocarpus longan*, *Litchi chinensis*, *Xanthoceras sorbifolium*) tem seus genomas disponibilizados em bancos de dados públicos (<https://www.ncbi.nlm.nih.gov/genome/browse#!/overview/Sapindaceae>; Acesso em 9 de novembro de 2021). Os dois primeiros genomas da espécie da família *Xanthoceras sorbifolium* Bunge e *Acer yangbiense* Y. S. Chen & Q. E. Yang foram obtidos combinando sequenciamento *Pacific Biosciences Single-molecule Real-time* e

sequenciamento Illumina e foram ambos publicados em 2019 (LIANG et al., 2019; YANG et al., 2019). Já para genomas cloroplastidiais 94 genomas são encontrados no banco de dados *Genome* do NCBI (*National Center for Biotechnology Information*) para família Sapindaceae, representando 4,95% da diversidade de espécies da família, ainda desses 77 genomas cloroplastidiais são de espécies pertencentes ao gênero *Acer* e apenas 13 gêneros tem ao menos um genoma disponível (<https://www.ncbi.nlm.nih.gov/genome/browse#!/organelles/Sapindaceae>; Acesso em 9 de novembro de 2021).

Os genomas cloroplastidiais vêm sendo amplamente utilizados com sucesso para resolução filogenética de alguns táxons e apresentam regiões com potencial utilização como marcador *barcoding* (DONG et al., 2015; GONÇALVES et al., 2019; MOORE et al., 2010). A obtenção desses genomas é facilitada devido ao seu tamanho pequeno, quando comparado ao genoma nuclear e apresentarem milhares de cópias em uma única célula da folha, sendo uma importante ferramenta nos estudos de genética e biologia molecular (TWYFORD; NESS, 2017). Os primeiros genomas cloroplastidiais foram sequenciados ainda em 1986, utilizando dados obtidos por sequenciamento de Sanger e de Maxam e Gilbert (OHYAMA et al., 1986; SHINOZAKI et al., 1986). Atualmente com surgimento e o relativo barateamento das técnicas de sequenciamento de DNA de alto desempenho, juntamente com o desenvolvimento de novos *pipelines* para montagem de genomas cloroplastidiais, tem permitido a obtenção de plastomas até então não sequenciados (TWYFORD; NESS, 2017).

Em geral, o DNA cloroplastidial (cpDNA) tem tamanho que variam de 120,000 bp a 160,000 bp em plantas terrestres, entretanto em plantas parasitas o cpDNA pode ser significativamente reduzido, visto à não necessidade de realizar fotossíntese (MOHANTA et al., 2020; RAVIN et al., 2016; TWYFORD; NESS, 2017; WICKE et al., 2011). O cpDNA em plantas com sementes apresenta uma estrutura relativamente conservada com duas regiões repetidas invertidas (IR) que dividem uma região grande de cópia única (LSC) e uma região pequena de cópia única (SSC) (DUGAS et al., 2015; WICKE et al., 2011; XIAO-MING et al., 2017). Nesses genomas são descritos entre 70 e 88 genes codificadores de proteínas, e 33 e 35 genes codificadores de RNAs funcionais e aproximadamente 110 genes diferentes (WICKE et al., 2011).

Os genomas cloroplastidiais tem potencial utilização em estudos evolutivos podendo ser uma ferramenta para esclarecer as relações filogenéticas do gênero parafilético *Serjania* (ACEVEDO-RODRÍGUEZ et al., 2017; DONG et al., 2015;

GONÇALVES et al., 2019; MOORE et al., 2010). Entre o gênero destaca-se *Serjania erecta* uma planta com potencial utilização farmacológica e como inseticidas naturais, o qual ainda não existe recursos genômicos para esta espécie (ARRUDA, 2011; BROGGINI et al., 2010; CARDOSO et al., 2013; FERNANDES et al., 2011; FREITAS, 2018; GUIMARÃES et al., 2015). Portanto, neste estudo, objetivamos montar o primeiro genoma cloroplastidial completo de *S. erecta* e realizar análises comparativas e evolutivas com plastomas de outras Sapindaceae. As análises comparativas buscaram responderem três perguntas: (1) Quais fatores podem levar à variação no número de genes de plastídios da família Sapindaceae? (2) Quais genes potencialmente estão sobre efeito da seleção natural em genomas cloroplastidiais? (3) Quais regiões do genoma cloroplastidial apresentam maior potencial para uso como marcador molecular?

2. OBJETIVOS

2.1 Objetivo geral

Realizar a montagem e anotação do genoma cloroplastidial de *Serjania erecta* Ralck e utilizá-lo para realizar análises comparativas e evolutivas entre plastomas da família Sapindaceae.

2.2 Objetivos específicos

- I.** Sequenciar, montar e anotar o genoma cloroplastidial de *Serjania erecta* Ralck;
- II.** Realizar análises comparativas entre as estruturas de genomas cloroplastidiais da família Sapindaceae;
- III.** Identificar variação no número de genes em genomas cloroplastidiais da família Sapindaceae;
- IV.** Calcular a relação entre as mutações não sinônimas e sinônimas entre genes homólogos do genoma cloroplastidial de Sapindaceae, bem como, identificar as regiões que são *hotspots* mutacionais nesses genomas;
- V.** Identificar regiões com potencial uso para estudos de filogenia molecular.

APENDICE I – MANUSCRITO

Genoma cloroplastidial de *Serjania erecta* Raldk: Variação no número de genes e análise de seleção de genes de plastomas da família Sapindaceae

** Este manuscrito segue as normas da revista Planta.

Genoma cloroplastidial de *Serjania erecta* Raldk: Variação no número de genes e análise de seleção de genes de plastomas da família Sapindaceae

Leonardo C. J. Corvalan^a, Mariane B. Sobreiro^{a,c}, Larissa Carvalho^a, Cintia P. Targueta^{a,b}, Ramilla dos S. B. Braga-Ferreira^{a,c}, José A. F. Diniz-filho^{a,d}, Mariana P. de C. Telles^{a,e}, Rhewter Nunes^{a,c,*}

^a PPG em Genética e Biologia Molecular – Universidade Federal de Goiás, Goiânia, Goiás, Brasil

^b PPG em Biotecnologia e Biodiversidade – Rede Pró-Centro-Oeste – Universidade Federal de Goiás.

^c Instituto Nacional de Ciência e Tecnologia – Ecologia, Evolução e Conservação da Biodiversidade (INCT - EECBio).

^d PGG em Ecologia e Evolução - Universidade Federal de Goiás, Goiânia, Goiás, Brasil

^e Pontifícia Universidade Católica de Goiás, Goiânia, Goiás, Brasil

Conclusão principal:

Os genomas cloroplastidiais da família Sapindaceae apresentam estrutura bem conservada com variações no número de genes codificadores de proteínas associadas ao tamanho da região invertida repetida.

Resumo

Serjania erecta Raldk é uma planta com potencial de uso no tratamento de Alzheimer, doenças gástricas e como agente anti-inflamatório. Entretanto, aspectos genéticos e evolutivos da espécie permanecem desconhecidos. Assim, neste estudo, sequenciamos e montamos o genoma do cloroplasto de *S. erecta* e o comparamos com genomas de onze gêneros diferentes da família Sapindaceae. Foi recuperado um genoma cloroplastidial para *S. erecta* de 159,297 bp, dividido em uma região longa de cópia única (LSC) de 84,556 bp, uma região curta de cópia única (SSC) de 18,057 bp e duas regiões invertidas repetidas (IRA e IRb) de 18,057 bp. Entre os 12 genomas cloroplastidiais, *S. erecta* destaca-se como a espécie com menor quantidade de repetições complexas e microssatélites. A estrutura dos genomas na família é relativamente conservada e o número de genes varia de 128 a 132 genes, tal variação está associada a três fatores: (1) Eventos expansão e retração no tamanho das IRs; (2) Pseudogenização do gene *rps2* e (3) Ausência ou duplicação de genes codificadores de tRNAs. As análises de diversidade nucleotídica identificaram apenas duas regiões gênicas (*ndhF* e *ycf1*), apresentando potencial como marcadores moleculares para estudos do gênero *Serjania*. Foi identificado seleção neutra apenas para o gene *psaI* e nos demais genes foi identificado seleção negativa, demonstrando uma grande conservação das regiões gênicas nesta família. Este trabalho fornece o primeiro genoma cloroplastidial de um membro da tribo Paullinieae e contribui com o entendimento da evolução molecular de Sapindaceae.

Palavras chaves:

Análise comparativa; Genoma organelar; Seleção negativa; Marcador molecular.

1. Introdução

Os cloroplastos são organelas responsáveis pela fotossíntese, que devido sua origem na endossimbiose de cianobactérias e células eucariotas primitivas, tem um DNA próprio (Margulis and Bermudes 1985; Rodríguez-Ezpeleta et al. 2005; Sato 2020). Nas espermatófitas, geralmente é observado um padrão de genoma circular subdividido em quatro regiões, duas regiões copiadas invertidas repetidas (IRs – *Inverted repeats*), uma região curta de cópia única (SSC – *small single copy*) e uma região longa de cópia única (LSC- *Large single copy*) (Jansen and Ruhlman 2012; Xiao-Ming et al. 2017). Os tamanhos desses genomas podem variar de 19,400 pb (*Cytinus hypocistis*) a 242,500 pb (*Pelargonium transvaalense*) em Spermatophyta (*National Center for Biotechnology Information, 2021*) sendo amplamente utilizados com sucesso para avaliar a história evolutiva das espécies (Leebens-Mack et al. 2005; Moore et al. 2010).

Divulgado em 1986, o genoma de *Nicotiana tabacum* e *Marchantia polymorpha* foram os primeiros genomas cloroplastidiais associando dados de sequenciamento de Sanger e de Maxam e Gilbert (Ohyama et al. 1986; Shinozaki et al. 1986). Atualmente, com os avanços nas tecnologias de sequenciamento de alto desempenho (HTS), o número de genomas de cloroplastos sequenciados e disponibilizados em bancos de dados públicos tem crescido, chegando a 4.915 genomas cloroplastidiais depositados no banco de dados *genome* do NCBI (*National Center for Biotechnology Information*) (Ohyama et al. 1986; Shinozaki et al. 1986; Twyford and Ness 2017).

Em plantas que realizam fotossíntese o número de genes varia entre 100 a 140. Em algumas famílias, como Sapindaceae, Poaceae, Leguminosae e Cactaceae, eventos de expansão e retração das IR influenciam significativamente no número total de genes (Wang et al. 2017; Xiao-Ming et al. 2017; Souza et al. 2019; Köhler et al. 2020; Dong et al. 2021). Embora exista certa variação no número de genes, a ordem deles em geral é conservada em Angiospermas, apresentando potencial para serem utilizados em estudos comparativos evolutivos (Wicke et al. 2011).

A família Sapindaceae, também conhecida como *soapberry family* (fruto-sabão), abriga cerca 1900 espécies de 144 gêneros, com aproximadamente 80% da sua biodiversidade contida em regiões tropicais e subtropicais do hemisfério sul (ACEVEDO-RODRÍGUEZ et al. 2010; Muellner-Riehl et al. 2016; Acevedo-Rodríguez et al. 2017; Buerki et al. 2021). Abrigando árvores, arbustos, lianas (plantas trepadeiras

lenhosas) e trepadeiras herbáceas, esta família apresenta espécies de elevada importância econômica, devido seus usos madeireiros, frutíferos e medicinais (Ferrucci and Acevedo-Rodríguez 2005; Buerki et al. 2009; ACEVEDO-RODRÍGUEZ et al. 2010; Muellner-Riehl et al. 2016; Dong et al. 2021). Ainda que apresentem grande parte da biodiversidade da família Sapindaceae, não existem dados de genoma cloroplastidial para mais de 50% das tribos da família e apenas 13 gêneros contêm pelo menos um genoma cloroplastidial sequenciado (*National Center for Biotechnology Information*, 2021).

Data-se que a família Sapindaceae surgiu no início do Cretáceo superior (105 Ma) e começaram a divergir aproximadamente 87 Ma, na Eurásia, se dispersando pelo Hemisfério sul, durante o Paleoceno. Atualmente, essa família é subdividida em quatro subfamílias, (1) Xanthoceroideae e (2) Hippocastanoideae, de ocorrência predominante em regiões temperadas; (3) Dodonaeoideae, distribuída em regiões tropicais e subtropicais e (4) Sapindoideae, com ampla distribuição em regiões tropicais e subtropicais é a mais diversa entre as quatro subfamílias (Buerki et al. 2009). A distribuição em tribos nesta família ainda é discutida, a revisão mais recente propõe 20 tribos, das quais 16 pertencem a subfamília Sapindoideae, 2 tribos a subfamília Dodonaeoideae, 2 tribos a subfamília Hippocastanoideae e na subfamília Xanthoceroideae não foram classificadas tribos (Buerki et al. 2021).

Serjania erecta Raldk incluída na subfamília Sapindoideae (tribo Paullinieae), é uma planta tradicionalmente utilizada para tratamento de úlceras e hipertensão com potencial uso no tratamento de doença de Alzheimer, doenças gástricas e uso anti-inflamatório (Gomig et al. 2008; Buerki et al. 2009; Hiruma-Lima et al. 2009; Guimarães et al. 2015). O gênero *Serjania* apresenta aproximadamente 230 espécies distribuídas desde os Estados Unidos até a Argentina, entretanto as relações filogenéticas entre espécies do gênero não são bem definidas, apresentando politomias extensas (Acevedo-Rodríguez 1990; Acevedo-Rodríguez et al. 2017; Buerki et al. 2021). Todas as espécies do gênero *Serjania* com cariótipo conhecido são diploides ($2n=24$) (Urdampilleta et al. 2012) e seus dados genômicos disponíveis em bancos públicos são escassos, não apresentando rascunhos de genomas ou genomas organelares disponibilizados em bancos de dados.

Neste estudo, foi montado o genoma cloroplastidial de *Serjania erecta*, primeiro genoma cloroplastidial da tribo Paullinieae (Sapindaceae) e também foram realizadas análises comparativas evolutivas com espécies da família Sapindaceae. As análises

comparativas tiveram como principal objetivo identificar: (1) Variações no número de genes em plastomas da família Sapindaceae; (2) Seleção em genes em genomas cloroplastidiais da família Sapindaceae; (3) Regiões com potencial utilização como marcadores para o gênero *Serjania*.

2. Material e métodos

2.1. Extração e sequenciamento de DNA

Foram coletadas folhas de um indivíduo de *Serjania erecta* na Reserva Particular do Patrimônio Natural Ecocerrado Brasil em Araxá (Minas Gerais) para a extração de DNA. Utilizando o método CTAB (Doyle and Doyle 1987), o DNA foi extraído e quantificado utilizando eletroforese horizontal em gel de agarose (1%). A biblioteca para sequenciamento foi construída utilizando o kit SureSelectQXT (Agilent Technologies) e validação da qualidade da biblioteca foi feita utilizando o Bioanalyzer 2100 (Agilent). Posteriormente, a biblioteca foi sequenciada na plataforma MiSeq (Illumina) no modo *paired-end* (2x300), utilizando o kit MiSeq V3 600 ciclos (Illumina). Todas as etapas laboratoriais foram conduzidas no Laboratório de Genética & Biodiversidade – LGBio, da Universidade Federal de Goiás, em Goiânia (GO) - Brasil.

2.2. Montagem e anotação do genoma cloroplastidial

A montagem do genoma de *S. erecta* foi conduzida utilizando como *seed* o gene *rbcL* de *Serjania polyphylla* (NCBI acesso: GU935455.1) no programa NOVOPlasty v3.2 (Dierckxsens et al. 2017). Para isso foi utilizado 8,8 milhões de *paired-end reads* gerados no tópico anterior.

A anotação do genoma cloroplastidial foi conduzida utilizando o *webservice software GeSeq* (Tillich et al. 2017). Neste programa a predição de genes codificadores de proteínas e RNAs ribossomais (rRNAs) foi feita nos programas *HMMER profile Search* (Wheeler and Eddy 2013) e *BLAT search* (Kent 2002) e a predição de RNAs de transferências (tRNAs) utilizando ARAGORN v1.2.38 (Laslett and Canback 2004). Em seguida, as anotações sobrepostas e genes mal anotados foram removidas ou reanotadas manualmente, utilizando o programa *Ugene* (Okonechnikov et al. 2012). O mapa gráfico do genoma cloroplastidial foi feito usando o *Draw Organelle Genome Maps (OGDRAW)* (Lohse et al. 2013). A frequência de cada códon foi identificada utilizando

webservice Sequence Manipulation Suite: codon usage
(https://www.bioinformatics.org/sms2/codon_usage).

2.3. Análise comparativa estrutural e de regiões repetitivas

O genoma cloroplastidial de *Serjania erecta* foi comparado ao de 11 espécies pertencentes a família Sapindaceae (Tabela 1). Estes genomas foram obtidos no banco de dados *Genome (RefSeq)* do *National Center for Biotechnology Information* (NCBI), e representam todos os gêneros de Sapindaceae depositados no banco de dados até março de 2021 (números de acesso do NCBI na Tabela S1). Posteriormente, todos genomas foram reanotados através do mesmo protocolo utilizado para *S. erecta*.

Para identificar o tamanho de cada região dos genomas cloroplastidial (*small single copy, large single copy e inverted repeat*) e os genes localizados em suas extremidades foi utilizado o *Geneiousprime* (Kearse et al. 2012). A identificação de possíveis eventos de rearranjo e inversão nos genomas cloroplastidiais foi feita utilizando o programa *Mauve* (Darling et al. 2004), um alinhador múltiplo de genomas.

A identificação de estruturas de repetições genômicas complexas (*forward, reverse, palindromic e complement*) foram conduzidas no programa *REPuter* (Kurtz et al. 2001), definindo o tamanho mínimo da repetição de 30 bp e uma distância de *Hamming* de 3 bp. Já a identificação das regiões microssatélites (SSR) foi feita utilizando a ferramenta web MISA (Beier et al. 2017). Para isso foram definidos os números de repetições mínimas 10, 5, 4, 3, 3, 3 para as repetições mononucleotídicas, dinucleotídicas, trinucleotídicas, tetranucleotídicas, pentanucleotídicas e hexanucleotídicas, respectivamente.

2.4. Reconstrução filogenética

Ao total foram conduzidas três reconstruções filogenéticas: Uma para a família Sapindaceae, a fim de se compreender a relação filogenética de *S. erecta* com outras genoma cloroplastidiais da família; as outras duas filogenias, tiveram como objetivo testar se os de marcadores moleculares de genomas cloroplastidiais disponíveis juntamente com os dados se *S. erecta* atualmente disponíveis são capazes esclarecer as relações filogenética do gênero *Serjania*.

Para a reconstrução filogenética dos genomas cloroplastidiais da família Sapindaceae, foram selecionadas 23 espécies (Tabela S1), das quais 22 espécies eram

pertencentes á ordem Sapindales e uma espécie da ordem Malvales (*Gossypium hirsutum*) que foi utilizada como *outgroup* (grupo externo) no enraizamento da árvore. Entre as espécies da ordem Sapindales, 16 pertencem à família Sapindaceae (*Serjania erecta*, *Acer buergerianum*, *Acer truncatum*, *Acer longipes*, *Aesculus chinensis*, *Aesculus wangii*, *Dimocarpus longan*, *Dipteronia sinensis*, *Dipteronia dyeriana*, *Dodonaea viscosa*, *Eurycorymbus cavaleriei*, *Koelreuteria paniculata*, *Litchi chinensis*, *Pometia tomentosa*, *Sapindus mukorossi* e *Xanthoceras sorbifolium*), duas espécies pertencentes à família Meliaceae (*Khaya senegalensis* e *Cedrela odorata*), duas espécies da família Rutaceae (*Citrus sinensis*, *Ruta graveolens*). Para as famílias Anacardiaceae e Burseraceae foram selecionadas as espécies *Mangifera indica* e *Boswellia sacra*, respectivamente. As CDS destas espécies foram obtidas do *GenBank* (seus números de referência se localizam na Tabela 1S).

As CDS foram alinhadas utilizando o *MAFFT version 7* (Katoh and Standley 2013) e concatenadas utilizando *Sequence Matrix v.1.7.8* (Vaidya et al. 2011). As regiões mais informativas para filogenia molecular foram selecionadas, utilizando o webserver Gblocks (Talavera and Castresana 2007). A reconstrução filogenética foi realizada utilizando o método de máxima verossimilhança (ML), no programa IQ-TREE *version 1.6.12* (Nguyen et al. 2015). Foram utilizados 63979 pb na construção da árvore filogenética e o melhor modelo foi selecionado utilizando o *ModelFinder* (Kalyaanamoorthy et al. 2017) implementado no IQ-TREE (Nguyen et al. 2015). Selecionou-se o modelo GTR+F+R3 e os valores de suporte de nó das árvores foram avaliados utilizando o método de bootstrapping com 1000 réplicas.

As outras duas árvores filogenéticas para o gênero *Serjania* foram construídas baseadas em dados de marcadores previamente publicados (Buerki et al. 2009, 2010; Acevedo-Rodríguez et al. 2017). Estas duas filogenias foram construídas utilizando os dados dos marcadores moleculares, a primeira utilizando os dados dos marcadores *matK*, *rpoB*, *trnD-trnT*, *trnK-matK*, *trnL*(intron) e *trnL-trnF* e a segunda outras apenas informações do *trnL* (íntron) (tabela S2). Para *S. erecta*, tais regiões foram extraídas do genoma cloroplastidial. Nessas filogenias seguimos o mesmo *pipeline* utilizado na árvore de ML das CDS dos genomas cloroplastidiais. O melhor modelo para ambas árvores foi o K3Pu+F.

2.5. Diversidade nucleotídica e relação de mutações sinônimas (ks) e não sinônimas (ka)

Nas análises de diversidade nucleotídica foram utilizadas as mesmas doze espécies utilizadas nas análises comparativas (tabela 1), separando-as em dois grupos: (1) formado pelas 12 espécies da família Sapindaceae e (2) subgrupo de 6 espécies pertencente a subfamília Sapindoideae. Os genomas dos dois grupos foram alinhados utilizando o programa MAFFT (Kato and Standley 2013) e posteriormente, o cálculo da diversidade nucleotídica (π) foi realizado, utilizando o programa DnaSP v6 definindo janelas de 600 bp e passos de 200 bp, (Rozas et al. 2017).

A fim de entender o processo de evolução dos genes dos cloroplastos, calculamos as relações entre mutações não sinônimas e sinônimas (ka/ks), que quando superior a 1, indica processo de seleção positiva e quando inferior a 1, indica processo de seleção negativa e quando igual 1 ausência de seleção (NEI and KUMAR 2000). O cálculo foi realizado utilizando os 77 CDS em comum entre os genomas dos cloroplastos das 12 espécies de Sapindaceae, os quais foram alinhados no programa MAFFT (Kato and Standley 2013). O valor de ka/ks foi estimado considerando dois modelos neutros, um considerando uma árvore filogenética (*runmode* = 0; *modelo* = 0; *NSsites* = 0) e outro apenas a relação par a par (*runmode* = - 2; *model* = 0; *NSsites* = 0) e estabelecendo um número mínimo de quatro mutações sinônimas, através do programa PAML 4.9 (Yang 2007).

Foi investigado a presença de locus sobre seleção positiva ($ka/ks > 1$) em todos os 77 genes, para isso foi utilizando o contraste entre modelos de seleção e modelos neutros: (1) M2a (modelo de seleção positiva) x M1a (modelo neutro); (2) M8 (modelo de seleção positiva com distribuição beta) x M7 (modelo neutro com distribuição beta); (3) M8 (modelo de seleção positiva com distribuição beta) x M8a (modelo de hipótese nula). Afim de se testar a significância dos contrastes calculamos Testes de Razão de Verossimilhança (LRT) e a Taxa de Falsa Descoberta (FDR), ambos em linguagem R(R Core Team 2020).

3. Resultados

3.1. Montagem e anotação do genoma cloroplastidial de *Serjania erecta*

O genoma cloroplastidial de *S. erecta* possui um tamanho de 159,297 bp e conteúdo de guanina e citosina (GC%) de 37,9% (Figura 1 e Tabela 1). Ele apresenta a típica estrutura quadripartida de cloroplastos de embriófitos com uma região longa de cópia única (LSC - *large single copy*) de 84,556 bp, uma região curta de cópia única (SSC - *small single copy*) de 18,057 bp e duas regiões invertidas repetidas (IRa e IRb - *inverted repeat*) de 28,342 bp separando as regiões de cópia única (Xiao-Ming et al. 2017) (Figura 1 e Tabela 1).

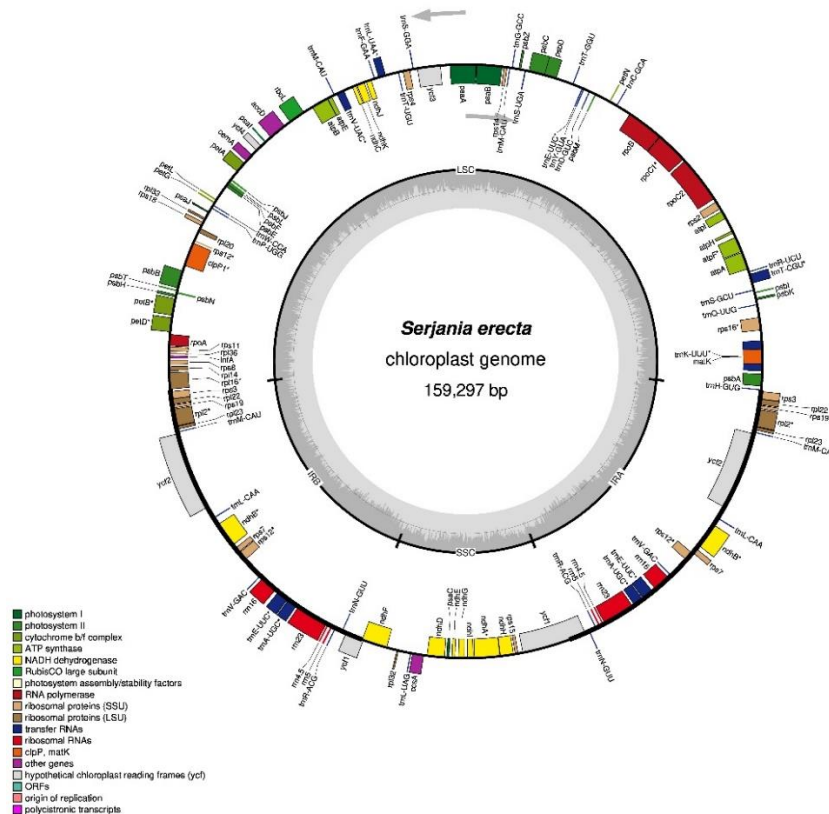


Figura 1. Mapa genômico do cloroplasto de *Serjania erecta* Radlk (Sapindaceae: Sapindoideae). Na circunferência mais externa estão representados os genes coloridos de acordo com o grupo funcional, os genes externos à circunferência são transcritos no sentido horário e os genes internos são transcritos no sentido anti-horário. A circunferência interna representa as regiões invertidas repetidas (IRa e IRb - *inverted repeat regions*), região longa de cópia única (LSC - *large single copy*), região curta de cópia única (SSC - *small single copy*). O gráfico de barras corresponde ao conteúdo GC. * representa genes que tem íntrons.

Tabela 1. Descrição dos doze genomas cloroplastidiais da família Sapindaceae utilizados na análise comparativa.

Species	GS	LSC	IR	SSC	GC (%)	Protein	tRNA	rRNA	Gene
<i>Serjania erecta</i>	159,297	84,556	28,342	18,057	37.90	87	37	8	132
<i>Acer buergerianum</i>	156,477	86,246	26,080	18,071	37.88	83	37	8	128
<i>Aesculus wangii</i>	155,871	84,882	26,390	18,210	37.95	83	37	8	128
<i>Dimocarpus longan</i>	160,833	85,709	28,427	18,270	37.79	87	37	8	132
<i>Dipteronia dyeriana</i>	157,071	85,530	26,723	18,095	37.97	85	37	8	130
<i>Dodonaea viscosa</i>	159,375	87,205	27,099	17,972	37.86	85	36	8	129
<i>Eurycorymbus cavaleriei</i>	158,777	86,941	26,922	17,992	37.92	85	36	8	129
<i>Koelreuteria paniculata</i>	163,258	90,237	27,376	18,269	37.30	85	37	8	130
<i>Litchi chinensis</i>	162,524	85,751	30,102	16,569	37.80	87	37	8	132
<i>Pometia tomentosa</i>	160,818	85,667	28,395	18,361	37.87	87	37	8	132
<i>Sapindus mukorossi</i>	160,481	85,650	27,979	18,873	37.66	87	37	8	132
<i>Xanthoceras sorbifolium</i>	161,231	85,300	28,619	18,693	37.69	86	38	8	132
Average	159,667.75	86,139.50	27,704.50	18,119.33	37.8	85.58	36.92	8	132.92
Standard deviation	2,319.68	1,493.70	1,147.41	560.61	0.18	1.51	0.51	0	1.88
Coefficient of variation (CV)	0.01	0.02	0.04	0.03	0.00	0.02	0.01	0.00	0.01

Nota: GS: tamanho do genoma cloroplastidial (*genome size*); LSC: tamanho da região longa de cópia única; SSC: tamanho da região curta de cópia única; IR: tamanho da região invertida repetida (*inverted repeat regions*); GC (%): conteúdo GC em porcentagem; Protein: número de genes codificadores de proteínas; Gene: número total de genes; *Average*: média; *Standard deviation*: desvio padrão; *Coefficient of variation*: coeficiente de variação.

Foram anotados 132 genes e 2 pseudogenes (*infA* e *ycf1*) do genoma cloroplastidial de *S. erecta*, destes 87 genes são codificadores de proteínas e 45 RNAs funcionais (Tabela 1 e S3). Entre os genes codificadores de proteínas 9 genes estão duplicados e todos localizados nas IR (*rpl2*, *rpl22*, *rpl23*, *rps3*, *rps7*, *rps12*, *rps19*, *ndhB*, *ycf2*), 9 genes apresentam um íntron (*rpl2*, *rpl16*, *rps16*, *rpoC1*, *ndhA*, *ndhB*, *petB*, *petD*, *ndhF*) e os genes *clpP*, *rps12* e *ycf3* apresentam dois íntrons (Tabela S3). Já nos genes codificadores de RNAs, 37 genes codificam tRNAs. Destes destacando-se o gene *trnM-CAU* com 4 cópias, outros 6 genes duplicados (*trnA-UGC*, *trnE-UUC*, *trnL-CAA*, *trnN-GUU*, *trnR-ACG*, *trnV-GAC*) e 5 genes contém um íntron (*trnA-UGC*, *trnK-UUU*, *trnL-UAA*, *trnT-CGU*, *trnV-UAC*) e os 4 genes codificadores rRNA (*rrn4.5*, *rrn5*, *rrn16*, *rrn23*) se localizam nas IR, portanto, duplicados (Tabela S3).

Ao total foram identificados 26,991 códons (Tabela S4), destes códons os mais utilizados foram o AAA e ATT, representando 1136 códons (4,02% do total de códons) e 1131 códons (4,02% do total de códons). Sendo que, AAA codifica o aminoácido lisina (LYS) e ATT codifica o aminoácido isoleucina (ILE). Entre os aminoácidos, o mais frequente é a leucina com 2835 aminoácidos (10,32% do total de aminoácidos) seguido da isoleucina (ILE) e lisina (LYS) com 2280 aminoácidos (8,45% do total de aminoácidos) e 2068 aminoácidos (7,66% do total de aminoácidos), respectivamente.

3.2. Análise comparativa estrutural e de regiões repetitivas

O genoma cloroplastidial de *S. erecta* apresenta estrutura e tamanho genômico semelhante ao das outras Sapindaceae avaliadas (tabela 1), destacando-se apenas como a espécie com a menor tamanho da LSC, com 84,556 bp, a qual variou até um tamanho máximo de 90,237 pb (*Koelreuteria paniculata*). O tamanho total dos genomas cloroplastidiais variou entre 163,258 bp (*K. paniculata*) e 155,871 bp (*Aesculus wangii*), variando 7387 bp. O conteúdo GC médio entre as doze Sapindaceae é de 37.80% com desvio padrão de 0.18 (Tabela 1).

O número total de genes varia entre 132 e 128 genes. O maior número de genes é descrito em 5 das 6 espécies pertencentes a subfamília Sapindoideae (*Dimocarpus longan*, *Litchi chinensis*, *Pometia tomentosa*, *Sapindus mukorossi* e *S. erecta*) e na espécie *Xanthoceras sorbifolium* (subfamília Xanthoceroideae). O menor número de genes foi observado nas espécies *A. buergerianum* e *Aesculus wangii*, ambas pertencente subfamília Hippocastanoideae (figura 2). Genomas cloroplastidiais com menor número de genes também apresentaram menores tamanhos de IR (tabela 1), a qual a região variou entre 26,080 bp com 128 genes na espécie *Acer buergerianum* a 30,102 bp com 132 genes para espécie *Litchi chinensis*.

Entre os genes codificadores de rRNAs todas as espécies apresentam os mesmos 4 genes duplicados, assim como esperado para genomas cloroplastidial de angiospermas. Já os genes codificadores de tRNA variaram entre 38 genes em *X. sorbifolium* e 36 genes identificado nas duas espécies da subfamília Dodonaeoideae (*Dodonaea viscosa* e *Eurycorymbus cavaleriei*). Destacando-se, os genes *trnH-GUG* duplicado no genoma de *X. sorbifolium* e o gene *trnT-CGU* ausente na subfamília Dodonaeoideae (figura 2). Não houve diferenças quanto a quantidade de íntrons entre as 12 espécies da família Sapindaceae (tabela S5).

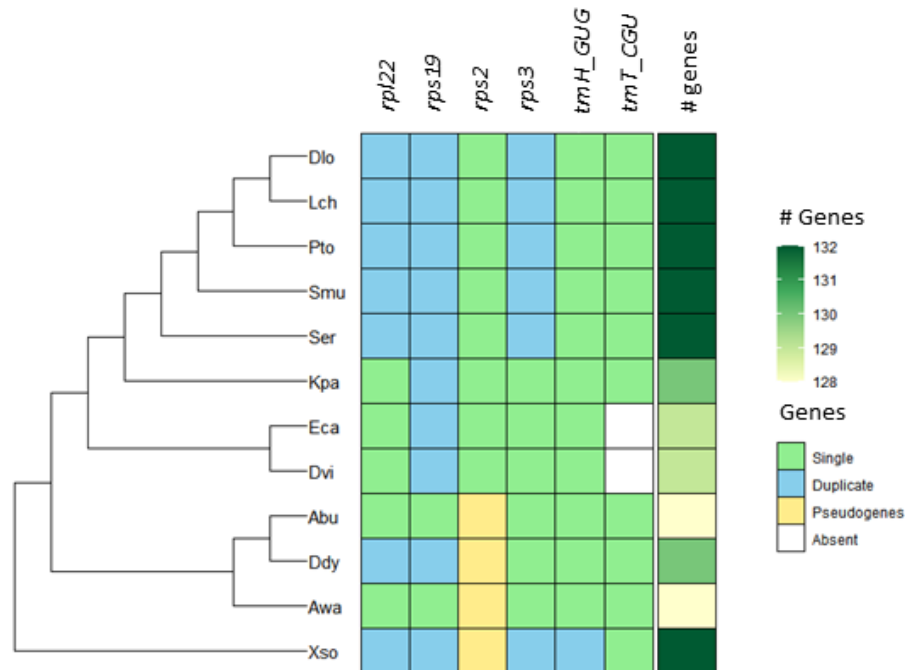


Figura 2. Genes que apresentam variação em sua quantidade em genomas cloroplastidiais entre as espécies *Acer buergerianum* (Abu), *Aesculus wangii* (Awa), *Dimocarpus longan* (Dlo), *Dipteronia dyeriana* (Ddy), *Dodonaea viscosa* (Dvi), *Eurycorymbus cavaleriei* (Eca), *Koelreuteria paniculata* (Kpa), *Litchi chinensis* (Lch), *Pometia tomentosa* (Pto), *Sapindus mukorossi* (Smu), *Serjania erecta* (Ser) e *Xanthoceras sorbifolium* (Xso). “# genes” indica o número total de genes identificados.

Os genes que flanqueiam as regiões de transição das regiões dos genomas cloroplastidial foram identificados com a finalidade de observar padrões relacionados com a quantidade e a posição de genes (Figura 3). Entre as LCS e as IR, existe uma variação nos genes que flanqueiam essa região. Foram identificados quatro padrões na região de transição entre a LSC e a IRb, o primeiro (*Standard 1* (LSC-IRb)) flanqueado pelos genes *rpl16* e *rps3*, nas espécies *D. longan*, *L. chinensis*, *P. tomentosa*, *S. mukorossi*, *S. erecta* e *X. sorbifolium*, todas elas espécies com 132 genes. O segundo padrão (*Standard 2* (LSC-IRb)) flanqueado pelos genes *rpl22* e *rps19*, foi identificado nas espécies *D. viscosa*, *E. cavaleriei* e *K. paniculatai*. O terceiro padrão (*Standard 3* (LSC-IRb)) flanqueado pelos genes *rps19* e *rpl2*, foi identificado em *A. buergerianum* e *A. wangii*, e o quarto padrão (*Standard 4* (LSC-IRb)) em uma única espécie *D. dyeriana*, flanqueandos pelos genes *rps3* e *rpl22*.

Já na região de transição entre as LSC e IRa foram identificados 5 padrões similares aos padrões de transição entre LSC e a IRb. No primeiro padrão (*Standard 1*

(LSC-IRa)) a região é flanqueada pelos *rps3* e *trnH*, formado pelas mesmas espécies que também apresentam o padrão 1(LSC-IRb) (*Standard 1* (LSC-IRb)), com exceção a espécie *X. sorbifolium* que é a única flanqueada por *rps3* e *psbA*(LSC-IRa) (*Standard 5* (LSC-IRa)). Os padrões 2, 3 e 4 (IRa-LSC) contém as mesmas espécies dos padrões 2, 3 e 4 da IRb-LSC (figura 3). Já nas regiões de transição entre as IR e SSC foram flanqueadas pelos mesmos genes em todas as espécies, entre IRb e SSC pelos *ycf1* pseudogenizado e o *ndhF* e, entre SSC e Ira, o gene *ycf1* (figura 3).

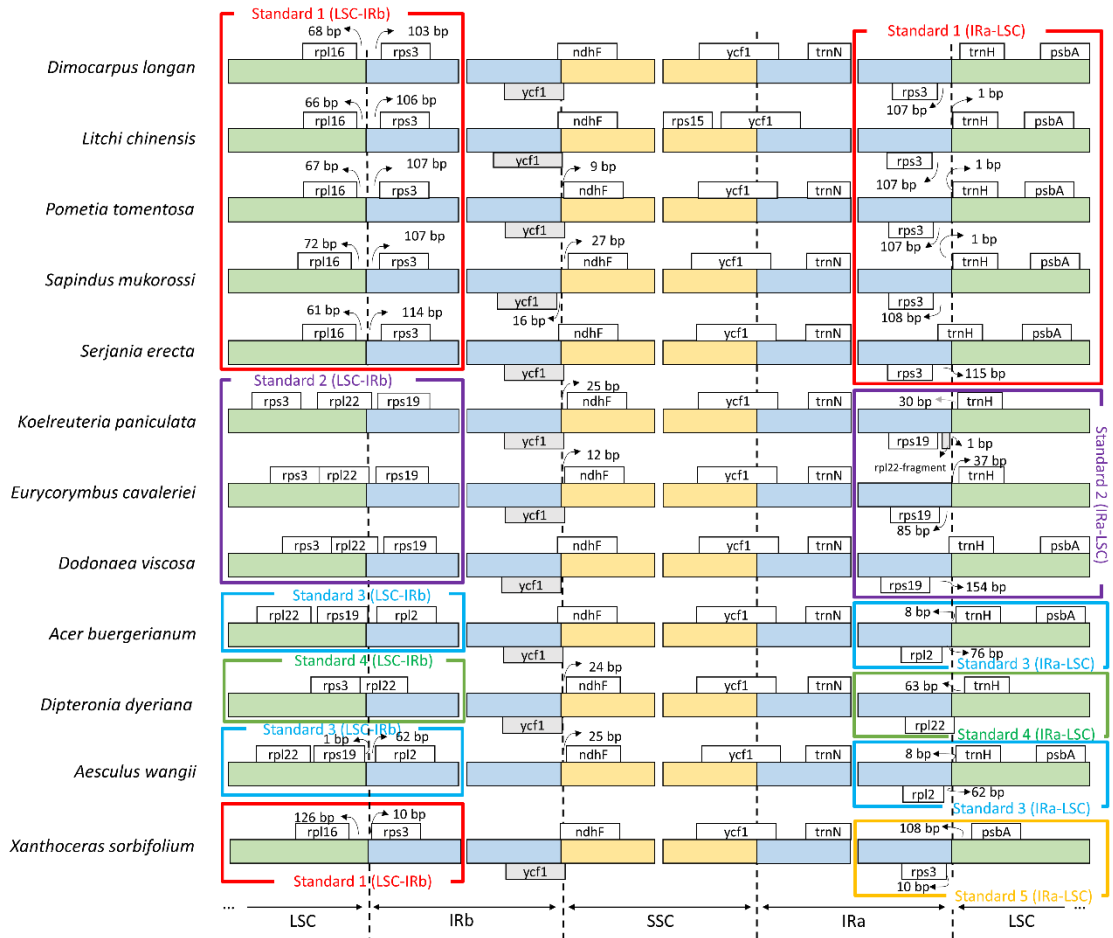


Figura 3. Comparação entre as regiões que delimitam dos genomas cloroplastidial. Regiões invertida repetida (IRa e IRb - *inverted repeat regions*) em azul, região longa de cópia única (LSC - *large single copy*) em verde, região curta de cópia única (SSC - *small single copy*) em amarelo. As caixas em branco acima representam os genes transcritos no sentido anti-horário e as abaixo os genes transcritos no sentido horários e as caixas cinzas representam pseudogenes.

Identificamos apenas um evento de rearranjo dos genomas cloroplastidiais para as doze espécies, sendo este evento único para *K. paniculatai* (figura S1). Ao total identificamos três blocos sintênicos, que se dividem apenas devido a uma região

localizada na IRa de *K. paniculatai*, o qual está na orientação complementar reversa comparada as outras espécies (bloco em verde figura S1).

O número de repetições complexas entre as 12 espécies variou entre 24 repetições (*S. erecta*) e 119 repetições (*K. paniculatai*) (figura 4A). Na família, foi possível encontrar os quatro tipos de repetições em apenas 4 espécies e os tipos mais frequentes foram os *forward* e *palindromic*. Estas repetições são encontradas em sua maioria nas regiões da LSC, seguido das IR e minoritariamente SSC em todas espécies, exceto *D. viscosa* (figura 4B).

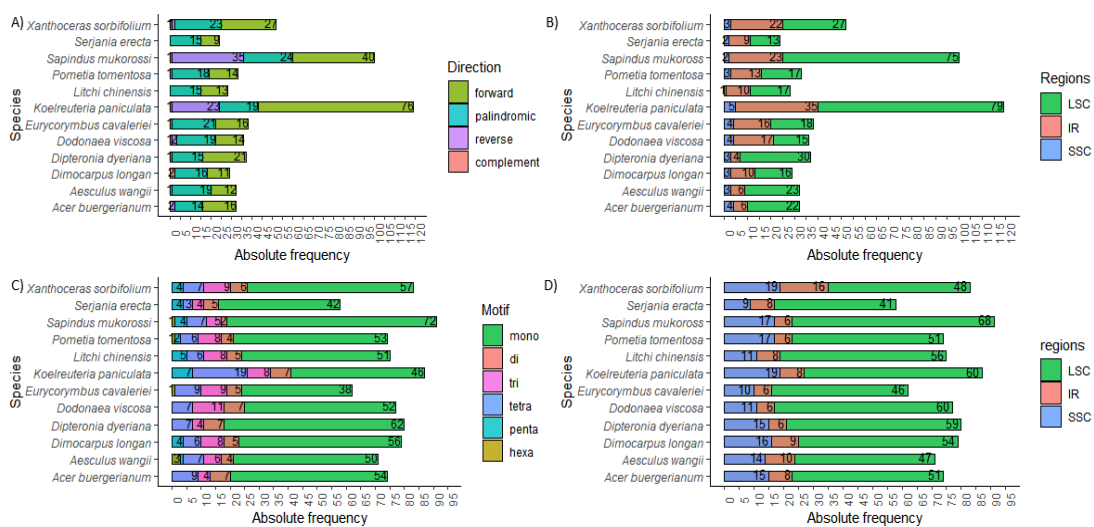


Figura 4. Distribuição de repetições nos genomas cloroplastidiais de doze espécies da família Sapindaceae. A) Distribuição das repetições complexas por tipo. B) Distribuição das repetições complexas de acordo com sua região no genoma cloroplastidial. C) Distribuição das SSR por seu motivo de repetição (Motif). D) Distribuição das SRR de acordo com sua região no genoma cloroplastidial. *Absolute frequency*: Frequência absoluta; *Species*: espécies.

S. erecta também obteve o menor número SSR (58 repetições) e *S. mukorossi* o maior número, com 91 repetições SRR (figura 4C), no qual a repetição mononucleotídica foi a mais abundante em todas as espécies. As SSR se localizam majoritariamente nas LSC e com menor frequência nas IR em todas as espécies (figura 4D).

3.3. Evolução do genoma cloroplastidiais de Sapindaceae

Para a árvore filogenética da família Sapindaceae foram utilizados ao total 68,917 bp da porção codificadora de proteínas dos genomas cloroplastidiais para a

reconstrução filogenética, resultando em uma árvore filogenética com alto valor de suporte para os nós, exceto para a ramificação entre a família Sapindaceae e o clado formado pelas famílias Rutaceae e Meliaceae (valor de suporte bootstrap=55) (figura 5). Na família Sapindaceae, foi possível identificar os 4 clados monofiléticos formado pelas 4 subfamílias, concordando com as topologias propostas por Bueki et al., (2009), Acevedo-rodríguez et al., (2017) e *Angiosperm Phylogeny Group* (APG) IV (Chase et al. 2016). A filogenia para o gênero *Serjania* apresentou extensas politomias e recuperou o grupo parafilético como descrito por Acevedo-rodríguez et al., (2017), tanto para as árvores com a utilização de dados dos 5 marcadores (figura S2a) quanto na que utilizou apenas do íntron do *trnL* (figura S2b).

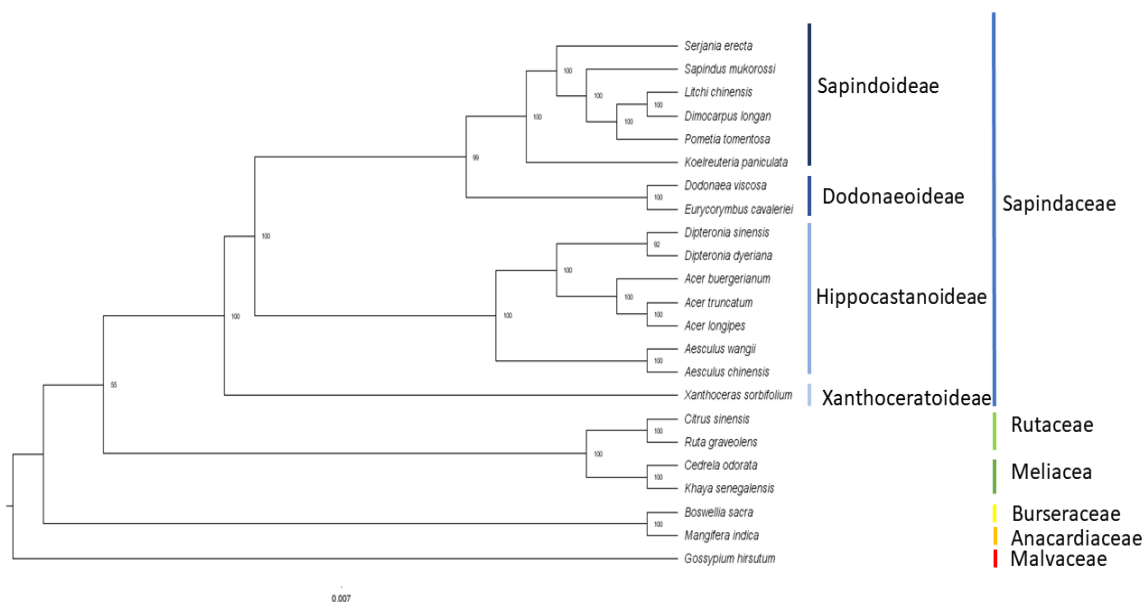


Figura 5. Árvores filogenética obtida pelo método de máxima verossimilhança (ML) da família Sapindaceae construída utilizando 74 genes codificadores de proteínas dos genomas cloroplastidial de 23 espécies e 1000 Bootstraps no programa IQ-Tree.

A diversidade nucleotídica (π) variou entre 0.00078 e 0.14732 para família Sapindaceae com mediana de 0.03572 (figura 6A e Tabela S6) e para subfamília Sapindoideae a variação da diversidade nucleotídica foi de 0 a 0.09744 com mediana de 0.02267 (figura 6B e Tabela S7). Os *hotspots* de diversidade nucleotídica foram definidos como os sítios com valores acima do dobro da mediana, no qual para a família Sapindaceae identificamos 10 *hotspots* de diversidade nucleotídica (*tRNA-Lys – rps16*, *atpI – rps2*, *rpoB – tRNA-asp*, *trna-Tyr - psbD*, *psbZ*, *ndhC – tRNA-Val*, *petA – psbJ*, *ndhF*, *rpl32 – cssA* e *ycf1*) (figura 6A) e para subfamília indenticamos 11 hotspots de diversidade nucleotídica (*tRNA-Lys-rps16*, *rps16 – tRNA-Gln*, *tRNA-Ser – tRNA-Gly*,

tRNA-Cys – *psbM*, *ndhC* – *tRNA-val*, *ycf4* – *cemA*, *petA* - *psbJ*, *psbE* – *petL*, *ndhF*, *rpl32* – *cssA* e *ycf1*(figura 6B). Destes as regiões *tRNA-Lys-rps16*, *ndhC-tRNA-val*, *petA-psbJ*, *ndhF*, *rpl32-ccsA* e *ycf1* são *hotspots* tanto para família quanto para subfamília.

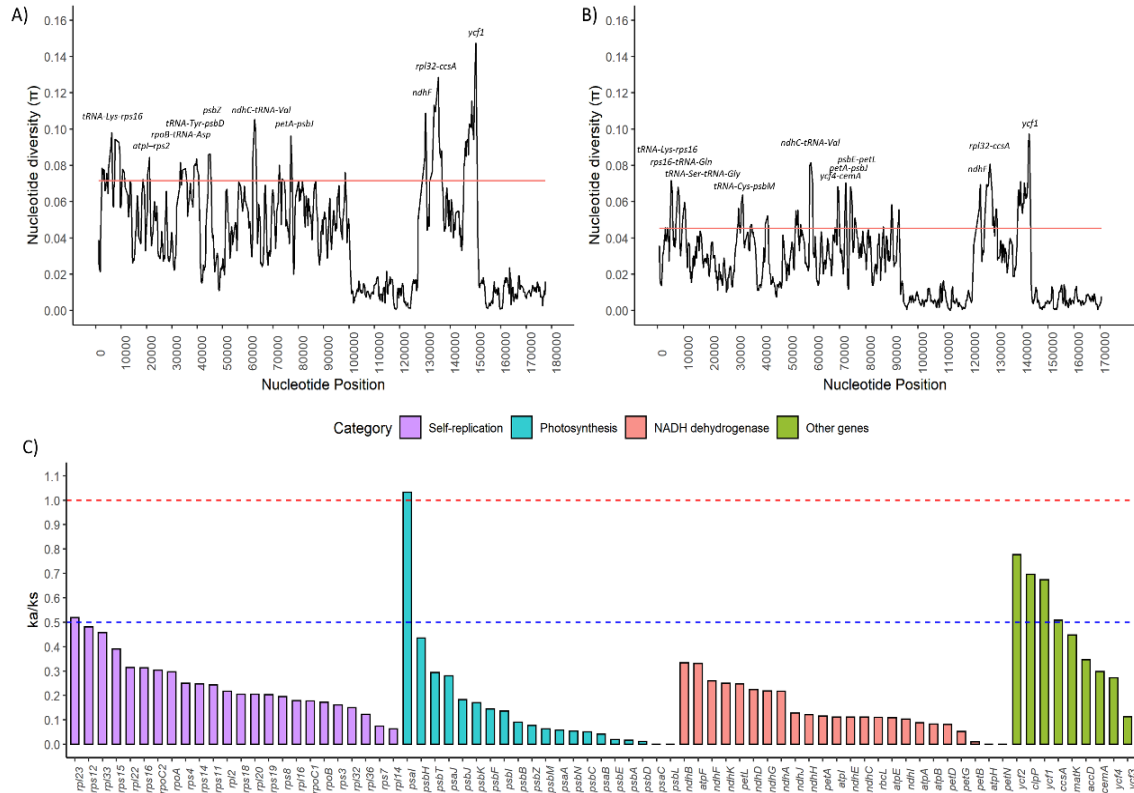


Figura 6. Diversidade nucleotídica (π) e relação das mutações não sinônimas (k_a) e sinônimas (k_s). A) Distribuição da diversidade nucleotídica ao longo dos genomas cloroplastidial de doze espécies da família Sapindaceae. B) Distribuição da diversidade nucleotídica ao longo dos genomas cloroplastidiais de seis espécies da subfamília Sapindoideae. C) Relação das mutações não sinônimas e mutações sinônimas (k_a/k_s) para os 77 genes codificadores de proteínas em comum nos genomas cloroplastidiais de doze espécies da família Sapindaceae. A linha vermelha contínua nos gráficos A e B indica o dobro da mediana dos valores da diversidade nucleotídica (π) no genoma cloroplastidial. A linha vermelha tracejada indica o valor em que a relação k_a/k_s é igual 1 e a linha azul tracejada indica o valor em que a relação k_a/k_s é igual 0.5.

Os valores médios entre as razões das mutações não sinônimas e mutações sinônimas (k_a/k_s) foram de 0.2160 (SD = 0.19), variando entre 0.001 nos genes *atpH*, *petN*, *psaC* e *psbL* e 1.032 para o gene *psal*. Não identificamos nenhum gene com valores k_a/k_s significativamente superior a 1 e apenas o gene *psal* apresentou tal relação próximo a 1, demonstrando que dos 77 genes em comum dos genomas cloroplastidiais, 76 genes estão sobre seleção negativa.

A fim de entender o processo de seleção natural para os genes par-a-par, foram selecionados os genes com maiores valores de ka/ks em cada um dos grupos funcionais de genes (Figura 7). Para os genes *rpl23* e *psaI* (Figura 7a e Figura 7b) foram identificados *outliers* (NA), relações entre genes que não apresentaram a quantidade de mutações sinônimas mínimas definida. Os genes *ndhB* e *ycf2* (figura 6) obtiveram os maiores valores na relação ka/ks para as categorias da NADH desidrogenase (*NADH dehydrogenase*) e outros genes (*Other genes*), demonstrando relações de seleção positiva pontuais (Figura 7c e Figura 7d). Em especial, *S. erecta* apresentou o gene *psaI* com valores ka/ks superiores a 1 para 7 das 11 espécies a qual foi comparado par-a-par (figura 7b).

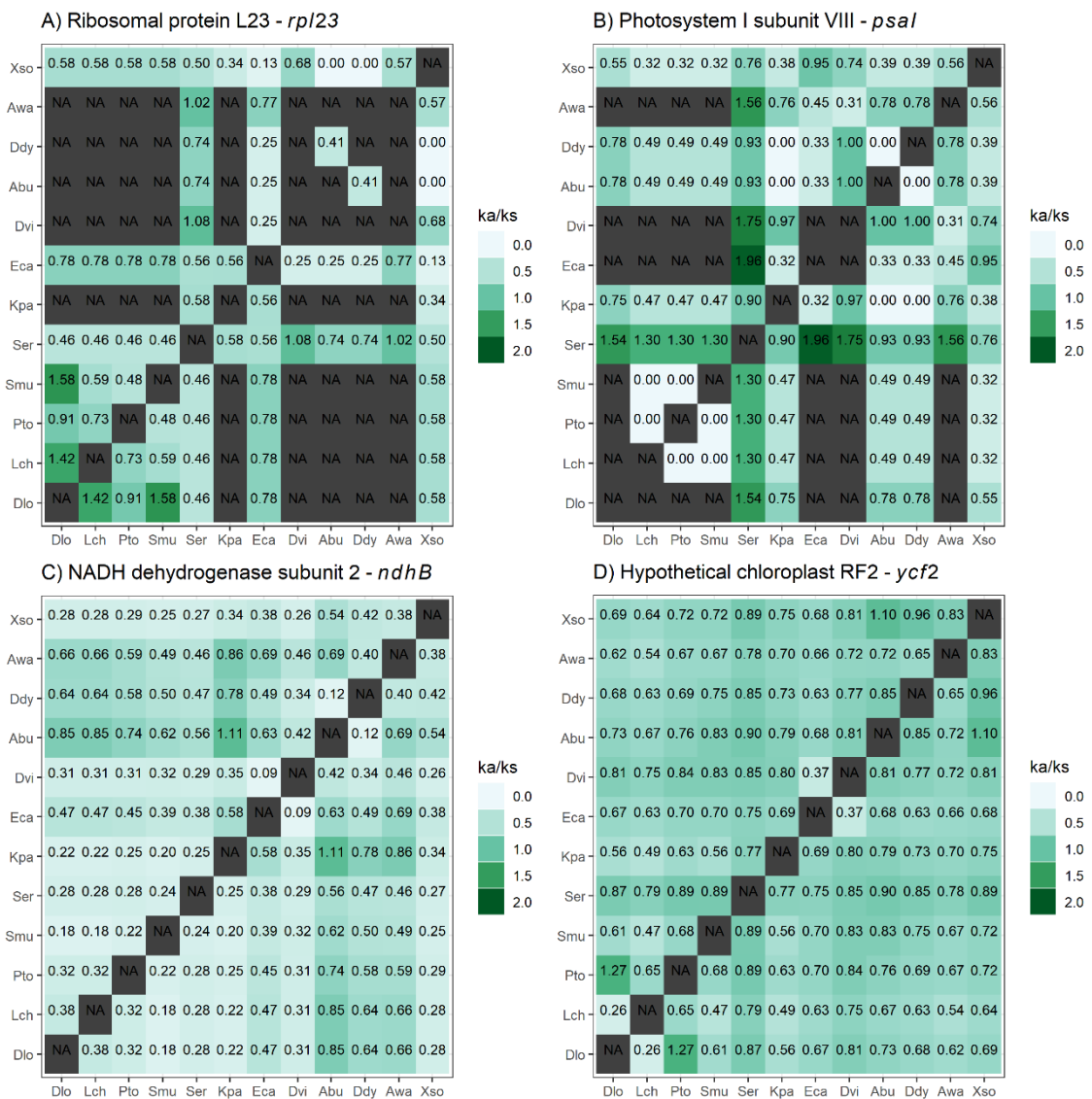


Figura 7. Relações entre as mutações não sinônimas e mutações sinônimas (ka/ks) par-a-par, referente genes do cloroplasto de 12 Sapindaceaes. A) Mapa de calor para o gene *rpl23* (Self-replication). B) Mapa de calor para o gene *psaI* (photosynthesis). C) Mapa de calor para o gene

ndhB (NADH dehydrogenase). D) Mapa de calor para o gene *ycf2* (Other genes). Valores iguais a NA representam outliers, genes com menor quantidades de mutações sinônimas que necessário para o cálculo de *ka/ks*.

Foi possível identificar seleção positiva local para os três contrastes testados evidenciados com base em LTR e FDR (M2a x M1a; (2) M8 x M7; (3) M8 x M8a) (Tabela S8). Entre estes contrastes, o contraste entre os modelos M8 x M7 apresentaram o maior número de genes sobre seleção positiva local com 12 genes (*atpA*, *clpP*, *matK*, *ndhA*, *ndhF*, *petD*, *psaB*, *rpl32*, *rpoB*, *rpoC2*, *ycf1*, *ycf2*). Já para os contrastes entre os modelos M2a x M1a e M8 x M8a evidenciamos seleção local positivas para 4 genes (*ndhF*, *rpoC2*, *ycf1*, *ycf2*). O gene *psaI* único gene sobre ausência de seleção e apresentou locus sob seleção positivas significativa apenas para o LDR.

4. Discussão

4.1 Variações estruturais e no números de genes

Os genomas cloroplastidiais da família Sapindaceae apresentaram uma estrutura relativamente conservada para as doze espécies avaliadas, com pouca variação no tamanho do genoma e conteúdo GC. Entre as regiões dos genomas dos cloroplastos a que maior apresentou variação foram as IRs, apresentando variações nos genes que flanqueiam a transição entre a LSC e IR estando associados a quantidades de genes. Revelando que os eventos de expansão e retração das IR são uma importante fonte na variação do número de genes na família, existindo uma possível relação filogenética. Também foi identificado que dos 77 genes codificadores de proteínas em comum da família, 76 estão sobre processo de seleção negativa.

O genoma *S. erecta* apresentou tamanho genômico, estrutura, número de gene e conteúdo GC dentro do esperado para genomas cloroplastidiais da família Sapindaceae. O valor do conteúdo GC para esta espécie são os mesmos descritos para *Dodonaea viscosa* e *Acer truncatum* e entre as doze espécies analisadas o conteúdo GC se mostrou extremamente conservado assim como descrito até então para família (Saina et al. 2018; Ma et al. 2019; Dong et al. 2021). Foi identificada ligeira variação no tamanho das quatro regiões dos genomas cloroplastidiais (LSC, SCCL e IR), considerando o coeficiente de variação (CV) os tamanhos das IRs (CV=0.04) e da SCC (CV= 0.03) variaram mais que LSC (CV=0.01), padrão esperado para plantas terrestres (Xiao-Ming et al. 2017). Esta variação pode estar associada a eventos de expansão e retração no

tamanho das IRs, eventos já descritos em membros da família Sapindaceae e outras angiospermas como um importante mecanismo evolutivo das plantas (Dugas et al. 2015; Li et al. 2017; Saina et al. 2018; Wang et al. 2018; Ma et al. 2019; Dong et al. 2021).

O número de genes em Sapindaceae concordam com estudos previamente publicados (Dong et al. 2021), sugerimos que esta variabilidade no número de genes se deve a três fatores: a expansão e retração no tamanho da IRs, o processo de pseudogenização do gene *rps2* e a mobilidade de genes codificadores de tRNAs. Embora o número de genes sejam concordantes com descritos na literatura, as anotações dos genomas cloroplastidiais depositadas no NCBI se diferenciam-se do descrito, portanto reanotamos todos os genomas cloroplastidiais usados na análise comparativa (Saina et al. 2018; Dong et al. 2021).

Os processos de expansões e retrações nas IR podem ser evidenciados pelos genes que flanqueiam a transição desta região com as regiões de cópia única (Li et al. 2017). Então, os padrões identificados evidenciam uma possível relação filogenética entre os eventos de expansão e redução das IR com o número de genes. Os genes *rpl22*, *rps19* e *rps3*, localizados na IR na subfamília Sapindoideae exceto na espécie mais basal *Koelreuteria paniculata* (Figura 2), fazem parte de um conjunto de genes codificadores de proteínas ribossomais (Christopher et al. 1988). Entre estes três genes o gene *rps19* é comumente na IR em Dicotiledôneas, ainda que em outras famílias como Euphorbiaceae o gene possa ser encontrado tanto na IR quanto na LSC (Li et al. 2017). Já o gene *rpl22* encontrado em genomas cloroplastidiais de Angiospermas, exceto na família Fabaceae, quando localizado na IR na família Rutaceae (Ordem Sapindales) evidencia eventos de expansão da IR (Gantt et al. 1991; Bausher et al. 2006). Estudos anteriores para família Sapindaceae também evidenciaram processos de expansão e retração na região LSC-IRb, ocorrendo até mesmo dentro de gêneros da família Sapindaceae, ressaltando que os eventos de expansão e retração das IR são uma importante fonte de variabilidade na quantidade de genes em Angiospermas (Li et al. 2017; Xiao-Ming et al. 2017; Saina et al. 2018; Ma et al. 2019; Köhler et al. 2020)

O gene *rps2* produz a proteína ribossomal 2S e é encontrado pseudogenizado nas subfamílias Hippocastanoideae e Xanthoceroideae (Zhou et al. 2016; Chen et al. 2019). Tal gene comumente é encontrado sob seleção positiva em outras famílias de Angiospermas e se localiza próximo a um *hotspot* de diversidade nucleotídica (atpI-

rps2), podendo então o acúmulo de mutações não sinônimas ter levado a pseudogenização deste gene (Wu et al. 2018; Li et al. 2020a, b; Sobreiro et al. 2020).

O gene *trnH-GUG* é exclusivamente localizado em regiões internas das IR em *X. sorbifolium*, então possivelmente esta duplicação não está associada a eventos de expansão das IRs. Além disso, a ausência dos genes do *trnT-CGU* em *E. cavaleriei* já havia sido descrita (Dong et al. 2021). Foi possível observar que *D. viscosa* apresenta a mesma ausência, o que pode indicar uma sinapormofia da subfamília Dodonaeoideae (Dong et al. 2021). Essa ausência completa de genes pode ser um caso de transferência de genes do cloroplasto para o genoma nuclear ou mitocondrial (Xiong et al. 2009).

Os genes se mostraram extremamente conservados e todos apresentaram mesmo número de íntrons. Os genes *clpP*, *rps12* e *ycf3* apresentaram dois íntrons e foram observados em todos as espécies anotadas, diferenciando-se de algumas anotações que não identificaram o gene *rps12*. O gene *rps12* apresenta 2 dois éxons 3' na IR e um éxon 5' na LSC, o gene *trans-splice* é relatado em diversas Viridiplantae e em alguns grupos taxonômicos, como as samambaias, há exclusão de um íntron (Hildebrand et al. 1988; Ma et al. 2019; Liu et al. 2020; Sobreiro et al. 2020).

Dois pseudogenes *infA* e uma das cópias de *ycf1* foram identificados em todas as espécies Sapindaceae. O gene *infA* é caracterizado por ser um gene móvel, o qual passou por diversos eventos de transferência do cloroplasto para o núcleo principalmente em linhagens do clado Rosids (Millen et al. 2001). Já a pseudogenização de uma das cópias de *ycf1* comumente encontrada em genomas cloroplastidiais de Angiospermas ocorre devido sua localização na zona de transição entre as IR e a SSC (figura 2)(Li et al. 2017).

Entre os aminoácidos, a Leucina e Citosina são os que mais e menos, respectivamente, compõem as proteínas do genomas cloroplastidial, assim como descrito em outras Sapindaceae (Muellner-Riehl et al. 2016; Ma et al. 2019). O uso predominante de alguns aminoácidos no cloroplastos pode estar associado a conservação dos genes codificadores de tRNAs, uma vez que a composição e número de tRNAs afetam as composições dos códons (Novoa et al. 2012).

Entre as espécies que identificamos regiões repetitivas, *K. paniculata* e *S. erecta* apresentaram as maiores e menores quantidades de repetições complexas, respectivamente e *S. mukorossi* e *S. erecta* as maiores e menores quantidades de repetições microssatélites, respectivamente, corroborando com estudo anterior que destaca *S. mukorossi* como uma espécie rica em regiões repetitivas em seu cloroplasto

(Dong et al. 2021). Tais regiões repetitivas são muito informativas quanto à diversidade genética dentro de espécies, sendo comumente utilizadas como marcadores moleculares em estudos de biogeografia e genética de populações (Varshney et al. 2005).

4.2 Análise de seleção de genes

Entre todos os genes avaliados não encontramos evidências de seleção positiva para nenhum desses genes ($ka/ks > 1$) e apenas o gene *psaI* não está sobre seleção natural, indicando uma grande conservação das sequências de aminoácidos produzidos pelos genes dos cloroplastos em Sapindales (NEI and KUMAR 2000). Estudos prévios em genomas cloroplastidiais da família Sapindaceae, também identificaram seleção negativa em quase todos os genes, apenas sendo encontrada seleção positiva em genes em análises de espécies par a par, assim como demonstrado para os genes *ndhB* e *ycf2* (Saina et al. 2018; Dong et al. 2021).

O único gene que identificamos sobre evolução neutra (*psaI*) compõem a subunidade VIII do centro de reação do fotossistema I, também apresentou locus sob seleção positiva para o FDR (Plöchinger et al. 2016; Schöttler et al. 2017). Teste anteriores em *Nicotiana tabacum*, induzindo mutações nesse gene não causaram alterações no crescimento das plantas, apenas reposta ao estresse de alta luz e frio durante a senescências das folhas (Schöttler et al. 2017). Embora *psaI* seja o único gene sob evolução neutra, diversas das relações ka/ks não apresentaram mutações sinônimas (ka) e o valores mais altos das relações estão em sua maioria relacionados com genes de *S. erecta*. A relação de *S. erecta* com as demais espécies para o gene *psaI* demonstra acúmulo de mutações não sinônimas na espécie, evidenciando possível seleção positiva para a espécie, tornando necessário comparações deste gene com genes ortólogos de espécies mais próximas para avaliar o efeito evolutivo destas mutações não sinônimas.

Os contrastes entre os modelos M7 x M8 identificou 12 genes (*atpA*, *clpP*, *matK*, *ndhA*, *ndhF*, *petD*, *psaB*, *rpl32*, *rpoB*, *rpoC2*, *ycf1*, *ycf2*) com locus sob seleção positiva, diferenciando-se dos contrastes entre os modelos M2a x M1a e M8 x M8a que são mais rigorosos na identificação e identificaram 4 genes (*ndhF*, *rpoC2*, *ycf1* e *ycf2*) com loci sobre seleção positiva (Yang and Swanson 2002; Swanson et al. 2003). O gene *ndhF* codifica a NADH desidrogenase F e está presente na maioria das plantas vasculares, e a seleção positiva deste gene está associada adaptação à maior incidência luminosa em locais de alta altitude em *Silene vulgaris* (Caryophyllaceae) e ao estresse salino em espécies do gênero *Limonium* (Plumbaginaceae), podendo então estes locus sobre

seleção positiva estar relacionados a adaptação em Sapindaceas (Krüger et al. 2019; Darshetkar et al. 2021). O gene *rpoC2* codifica uma subunidade do complexo proteico RNA polimerase codificada pelo plastídio (Zhang et al. 2015). Já os genes *ycf1* e *ycf2* em estudos prévios também demonstraram locus sob seleção positiva na ordem Dipsacales estando relacionados a adaptação ao ambiente (Fan et al. 2018).

4.3 Evolução molecular de Sapindaceae

A reconstrução filogenética baseada em CDS de genomas cloroplastidiais recuperou grupos monofiléticos tanto para as quatro subfamílias (Dodonoideae, Hippocastanoideae, Sapindoideae, Xanthoceroideae) quanto para as famílias (Sapindaceae, Rutaceae, Meliaceae, Burceraceae e Anacardiaceae) da ordem Sapindales, com uma topologia concordante com as descritas na última década (Buerki et al. 2009, 2010, 2021; Muellner-Riehl et al. 2016; Acevedo-Rodríguez et al. 2017). A utilização de dados genômicos do cloroplasto (*matK* e *rbcL*) na família Sapindaceae foram importantes na inclusão das antigas famílias Aceraceae e Hippocastanaceae e do gênero *Xanthoceras* à família, cuja classificação é aceita até hoje (Harrington et al. 2005). A classificação em tribos mais recente da família Sapindaceae identificou 20 tribos, das quais 16 pertencem à subfamília Sapindoideae a qual menos de 50% das tribos apresentam genomas cloroplastidiais disponíveis em banco de dados públicos (Buerki et al. 2021). Desta forma, fornecemos a primeira filogenia utilizando uma larga escala de genes cloroplastidiais que incluem um membro da supertribo Paullinioidae (tribos: Athyaneae, Bridgesieae, Thouinieae e Paullinieae; subfamília Sapindoideae) (Acevedo-Rodríguez et al. 2017; Buerki et al. 2021).

Ambas as filogenias para o gênero *Serjania*, utilizando dados de marcadores moleculares, indicam a formação de um gênero parafilético com inclusão de espécies dos gêneros *Balsas*, *Chimborazoa* e *Houssayanthus*, como anteriormente descrito por Acevedo-Rodríguez et al., (2017). Estima-se que os eventos de especiação dentro do gênero ocorreram durante o Neogeno, resultando em mais de 230 espécies distribuídas pelo Neotrópico. A grande diversificação pode estar associada à capacidade de dispersão de sementes pelo vento (Muellner-Riehl et al. 2016; Acevedo-Rodríguez et al. 2017; Buerki et al. 2021). Entretanto, os dados provenientes dos marcadores moleculares utilizados não são assertivos quanto à determinação de grupos dentro do gênero, formando uma grande politomia, no qual o *S. erecta* está inserida (Buerki et al. 2009, 2010; Acevedo-Rodríguez et al. 2017). A disponibilização de genomas

cloroplastidiais para do gênero pode auxiliar no desenvolvimento de futuros marcadores moleculares próprio para este gênero, podendo definir relações interespecíficas do gênero. Entretanto, é necessário maior quantidade de genomas de espécies filogeneticamente próximas.

Investigando a diversidade nucleotídica entre a família e a subfamília foi possível identificar *hotspots* de diversidade genética, que podem ter utilização como marcador molecular, resolvendo relações intraespecíficas e como marcadores *barcoding*. Em geral, os espaços intergênicos apresentam maior diversidade nucleotídica, enquanto as regiões gênicas são mais conservadas em espécies da família Sapindaceae (Ma et al. 2019; Dong et al. 2021). Os resultados corroboram com estudos anteriores dessa família, que caracterizaram as regiões intergênicas *rps16-tRNA-Gln*, *tRNA-Ser-tRNA-Gly*, *petA-psbJ* e *rpl32-cssA*, como sítios de grande diversidade nucleotídica (Zhou et al. 2016; Ma et al. 2019; Dong et al. 2021). Apenas os genes *ndhF* e *ycf1* apresentaram elevada diversidade nucleotídica, os quais já são descritos e utilizados como um marcador molecular capaz de distinguir espécies de angiospermas e com grande potencial para família deste estudo (Dong et al. 2015, 2021; Amar 2020; Ramírez-Barahona et al. 2020). Embora apenas um genoma cloroplastidial do gênero *Serjania* tenha sido descrito, as duas regiões gênicas têm potencial como marcadores para estudos filogenéticos podendo ser utilizados para esclarecer questões evolutivas do gênero.

Apenas os genes *psaI*, *ycf2*, *clpP*, *ycf1*, *rpl23* e *ccsA* obtiveram relações *ka/ks* superiores a 0.5 apresentando maior potencial para utilização como marcadores moleculares (Dong et al. 2021). Entre quatro genes que apresentam códons sobre seleção positiva (*ndhF*, *rpoC2*, *ycf1* e *ycf2*), os genes *ycf1* e *ndhF* são descritos como regiões *barcoding* promissoras para gênero *Prunus* e *ycf1* e *rpoC2* são regiões filogeneticamente informativas para tribos da família Asteraceae, demonstrando que estes genes também tem potencial utilização na família Sapindaceae (Amar 2020; Shen et al. 2020). Em especial o gene *ycf1* e *ndhF*, o qual também apresentaram elevada diversidade nucleotídica, podem ser uma alternativa ao uso dos marcadores tradicionais, como *matK* e *rbcL* (Dong et al. 2015, 2021; Amar 2020).

5. Conclusão

A realização deste estudo forneceu a primeira anotação e montagem de um genoma cloroplastidial de uma espécie da tribo Paullinieae (Sapindaceae), o genoma

cloroplastidial de *Serjania erecta*. A análise comparativa da família indica uma estrutura conservada com variação no número de genes, devido a eventos de retração e expansão nas IR, pseudogenização do gene *rps2* e ausência e duplicação de genes codificadores de tRNAs. Os genes codificadores de proteínas são extremamente conservados na família Sapindaceae e apenas o gene *psaI* não está sob seleção negativa. Este trabalho traz indícios que os genes *ycf1e ndhF* possam ser os mais indicados como marcadores para estudos filogenéticos da subfamília Sapindoideae e potencialmente contribuir para a resolução de incertezas taxonômicas nesse grupo.

Author contributions

RN, MPCT and JAFD conceived and supervised the Project. CPT and RSBB Collected plant samples, extracted and sequencing total DNA. RN performed chloroplast assembly. LCJC, MBS and LC performed all the remaining analyses. LCJC and RN wrote the manuscript.

Acknowledgements

This work was developed in the context of Instituto Nacional de Ciência e Tecnologia em Ecologia, Evolução e Conservação da Biodiversidade (INCT – EECBio), supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq – process 465610/20145) and Fundação de Amparo à Pesquisa do Estado de Goiás (FAPEG – process 201810267000023). LCJC received a master's fellowship from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). C.P.T. was supported by a PNPd scholarship from CAPES. RN, RSBB and MBS was supported by a PNPd scholarship from FAPEG. LC was supported by a doctoral scholarship CNPq (435477/2018-8). M.P.C.T. and JAFD has been continuously supported by a productivity grant from CNPq.

Declaration of Competing Interest

The authors declare no conflict of interests.

Referencias

- Acevedo-Rodríguez P (1990) Distributional patterns in Brazilian *Serjania* (Sapindaceae). *Acta Bot Brasilica* 4:69–82. <https://doi.org/10.1590/s0102-33061990000100005>
- ACEVEDO-RODRÍGUEZ P, WELZEN PC VAN, ADEMA F, HAM RWJM VAN DER (2010) Sapindaceae. *Fam genera Vasc plants Eudicots Sapindales, Cucurbitales, Myrtaceae*
- Acevedo-Rodríguez P, Wurdack KJ, Ferrucci MS, et al (2017) Generic Relationships and Classification of Tribe Paullinieae (Sapindaceae) with a New Concept of Supertribe Paullinioidae. *Syst Bot* 42:96–114. <https://doi.org/10.1600/036364417X694926>
- Amar MH (2020) *ycf1-ndhF* genes, the most promising plastid genomic barcode, sheds light on phylogeny at low taxonomic levels in *Prunus persica*. *J Genet Eng Biotechnol* 18:.

<https://doi.org/10.1186/s43141-020-00057-3>

- Bausher MG, Singh ND, Lee SB, et al (2006) The complete chloroplast genome sequence of *Citrus sinensis* (L.) Osbeck var “Ridge Pineapple”: Organization and phylogenetic relationships to other angiosperms. *BMC Plant Biol* 6:1–11. <https://doi.org/10.1186/1471-2229-6-21>
- Beier S, Thiel T, Münch T, et al (2017) MISA-web: a web server for microsatellite prediction. *Bioinformatics* 33:2583–2585. <https://doi.org/10.1093/bioinformatics/btx198>
- Buerki S, Callmander MW, Acevedo-Rodríguez P, et al (2021) An updated infra-familial classification of Sapindaceae based on targeted enrichment data. *Am J Bot* 108:1234–1251. <https://doi.org/10.1002/ajb2.1693>
- Buerki S, Forest F, Acevedo-Rodríguez P, et al (2009) Plastid and nuclear DNA markers reveal intricate relationships at subfamilial and tribal levels in the soapberry family (Sapindaceae). *Mol Phylogenet Evol* 51:238–258. <https://doi.org/10.1016/j.ympev.2009.01.012>
- Buerki S, Lowry PP, Alvarez N, et al (2010) Phylogeny and circumscription of Sapindaceae revisited: Molecular sequence data, morphology and biogeography support recognition of a new family, Xanthoceraceae. *Plant Ecol Evol* 143:148–159. <https://doi.org/10.5091/plecevo.2010.437>
- Chase MW, Christenhusz MJM, Fay MF, et al (2016) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot J Linn Soc* 181:1–20. <https://doi.org/10.1111/boj.12385>
- Chen M, Zhang H, Jiang M (2019) The complete chloroplast genome sequence of *Acer cinnamomifolium* (Aceraceae), a plant species endemic to China. *Mitochondrial DNA Part B Resour* 4:3450–3451. <https://doi.org/10.1080/23802359.2019.1674211>
- Christopher DA, Cushman JC, Price CA, Hallick RB (1988) Organization of ribosomal protein genes *rp123*, *rpl2*, *rps19*, *rpl22* and *rps3* on the *Euglena gracilis* chloroplast genome. *Curr Genet* 14:275–285
- Darling ACE, Mau B, Blattner FR, Perna NT (2004) Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 14:1394–1403. <https://doi.org/10.1101/gr.2289704>
- Darshetkar AM, Maurya S, Lee C, et al (2021) Plastome analysis unveils Inverted Repeat (IR) expansion and positive selection in Sea Lavenders (*Limonium*, Plumbaginaceae, Limonioideae, Limonieae). *PhytoKeys* 175:89–107. <https://doi.org/10.3897/phytokeys.175.61054>
- Dierckxsens N, Mardulyn P, Smits G (2017) NOVOPlasty: De novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res* 45:. <https://doi.org/10.1093/nar/gkw955>
- Dong F, Lin Z, Lin J, et al (2021) Chloroplast genome of rambutan and comparative analyses in sapindaceae. *Plants* 10:1–15. <https://doi.org/10.3390/plants10020283>
- Dong W, Xu C, Li C, et al (2015) *ycf1*, the most promising plastid DNA barcode of land plants. *Sci Rep* 5:8348. <https://doi.org/10.1038/srep08348>
- Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull* 19:11–15
- Dugas D V., Hernandez D, Koenen EJM, et al (2015) Mimosoid legume plastome evolution: IR expansion, tandem repeat expansions, and accelerated rate of evolution in *clpP*. *Sci Rep*

- 5:1–13. <https://doi.org/10.1038/srep16958>
- Fan WB, Wu Y, Yang J, et al (2018) Comparative chloroplast genomics of dipsacales species: Insights into sequence variation, adaptive evolution, and phylogenetic relationships. *Front Plant Sci* 9:1–13. <https://doi.org/10.3389/fpls.2018.00689>
- Ferrucci MS, Acevedo-Rodríguez P (2005) Three new species of *Serjania* (Sapindaceae) from south America. *Syst Bot* 30:153–162. <https://doi.org/10.1600/0363644053661904>
- Gantt JS, Baldauf SL, Calie PJ, et al (1991) Transfer of rpl22 to the nucleus greatly preceded its loss from the chloroplast and involved the gain of an intron. *EMBO J* 10:3073–3078. <https://doi.org/10.1002/j.1460-2075.1991.tb07859.x>
- Gomig F, Pietrovski EF, Guedes A, et al (2008) Topical anti-inflammatory activity of *Serjania erecta* Radlk (Sapindaceae) extracts. *J Ethnopharmacol* 118:220–224. <https://doi.org/10.1016/j.jep.2008.03.017>
- Guimarães CC, Oliveira DD, Valdevite M, et al (2015) The glycosylated flavonoids vitexin, isovitexin, and quercetrin isolated from *Serjania erecta* Radlk (Sapindaceae) leaves protect PC12 cells against amyloid- β 25–35 peptide-induced toxicity. *Food Chem Toxicol* 86:88–94. <https://doi.org/10.1016/j.fct.2015.09.002>
- Harrington MG, Edwards KJ, Johnson SA, et al (2005) Phylogenetic inference in sapindaceae sensu lato using plastid matK and rbcL DNA sequences. *Syst Bot* 30:366–382. <https://doi.org/10.1600/0363644054223549>
- Hildebrand M, Hallick RB, Passavant CW, Bourque DP (1988) Trans-splicing in chloroplasts: the rps 12 loci of *Nicotiana tabacum*. *Proc Natl Acad Sci U S A* 85:372–376. <https://doi.org/10.1073/pnas.85.2.372>
- Hiruma-Lima CA, Castelo APC, Arruda BN, et al (2009) Gastroprotective effect of *Serjania erecta* Radlk (Sapindaceae): Involvement of sensory neurons, endogenous nonprotein sulfhydryls, and nitric oxide. *J Med Food* 12:1411–1415. <https://doi.org/10.1089/jmf.2008.0269>
- Jansen RK, Ruhlman TA (2012) Plastid Genomes of Seed Plants. *Adv Photosynth Respir* 35:103–126. <https://doi.org/10.1007/978-94-007-2920-9>
- Kalyaanamoorthy S, Minh BQ, Wong TKF, et al (2017) ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat Methods* 14:587–589. <https://doi.org/10.1038/nmeth.4285>
- Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol* 30:772–780. <https://doi.org/10.1093/molbev/mst010>
- Kearse M, Moir R, Wilson A, et al (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647–1649. <https://doi.org/10.1093/bioinformatics/bts199>
- Kent WJ (2002) BLAT---The BLAST-Like Alignment Tool. *Genome Res* 12:656–664. <https://doi.org/10.1101/gr.229202>
- Köhler M, Reginato M, Souza-chies TT, Majure LC (2020) Insights Into Chloroplast Genome Evolution Across Opuntioideae (Cactaceae) Reveals Robust Yet Sometimes Conflicting Phylogenetic Topologies. 11:1–20. <https://doi.org/10.3389/fpls.2020.00729>
- Krüger M, Abeyawardana OAJ, Juříček M, et al (2019) Variation in plastid genomes in the gynodioecious species *Silene vulgaris*. *BMC Plant Biol* 19:1–15. <https://doi.org/10.1186/s12870-019-2193-0>

- Kurtz S, Choudhuri J V., Ohlebusch E, et al (2001) REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res* 29:4633–4642. <https://doi.org/10.1093/nar/29.22.4633>
- Laslett D, Canback B (2004) ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res* 32:11–16. <https://doi.org/10.1093/nar/gkh152>
- Leebens-Mack J, Raubeson LA, Cui L, et al (2005) Identifying the basal angiosperm node in chloroplast genome phylogenies: Sampling one’s way out of the Felsenstein zone. *Mol Biol Evol* 22:1948–1963. <https://doi.org/10.1093/molbev/msi191>
- Li C, Zhao Y, Xu Z, et al (2020a) Initial Characterization of the Chloroplast Genome of *Vicia sepium*, an Important Wild Resource Plant, and Related Inferences About Its Evolution. *Front Genet* 11:1–15. <https://doi.org/10.3389/fgene.2020.00073>
- Li P, Lou G, Cai X, et al (2020b) Comparison of the complete plastomes and the phylogenetic analysis of *Paulownia* species. *Sci Rep* 10:1–9. <https://doi.org/10.1038/s41598-020-59204-y>
- Li Z, Long H, Zhang L, et al (2017) The complete chloroplast genome sequence of tung tree (*Vernicia fordii*): Organization and phylogenetic relationships with other angiosperms. *Sci Rep* 7:1–11. <https://doi.org/10.1038/s41598-017-02076-6>
- Liu S, Wang Z, Wang H, et al (2020) Patterns and Rates of Plastid rps12 Gene Evolution Inferred in a Phylogenetic Context using Plastomic Data of Ferns. *Sci Rep* 10:1–12. <https://doi.org/10.1038/s41598-020-66219-y>
- Lohse M, Drechsel O, Kahlau S, Bock R (2013) OrganellarGenomeDRAW--a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Res* 41:575–581. <https://doi.org/10.1093/nar/gkt289>
- Ma Q, Wang Y, Zhu L, et al (2019) Characterization of the complete chloroplast genome of *acer truncatum bunge* (Sapindales: Aceraceae): A new woody oil tree species producing nervonic acid. *Biomed Res Int* 2019:. <https://doi.org/10.1155/2019/7417239>
- Margulis L, Bermudes D (1985) Symbiosis as a mechanism of evolution: status of cell symbiosis theory. *Symbiosis* 1:101–124
- Millen RS, Olmstead RG, Adams KL, et al (2001) Many parallel losses of infa from chloroplast dna during angiosperm evolution with multiple independent transfers to the nucleus. *Plant Cell* 13:645–658. <https://doi.org/10.1105/tpc.13.3.645>
- Moore MJ, Soltis PS, Bell CD, et al (2010) Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proc Natl Acad Sci U S A* 107:4623–4628. <https://doi.org/10.1073/pnas.0907801107>
- Muellner-Riehl AN, Weeks A, Clayton JW, et al (2016) Molecular phylogenetics and molecular clock dating of Sapindales based on plastid rbcL, atpB and trnL-trnF DNA sequences. *Taxon* 65:1019–1036. <https://doi.org/10.12705/655.5>
- NEI M, KUMAR S (2000) *Molecular evolution and phylogenetics*. Oxford University Press
- Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ (2015) IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32:268–274. <https://doi.org/10.1093/molbev/msu300>
- Novoa EM, Pavon-Eternod M, Pan T, Ribas De Pouplana L (2012) A role for tRNA modifications in genome structure and codon usage. *Cell* 149:202–213. <https://doi.org/10.1016/j.cell.2012.01.050>
- Ohyama K, Fukuzawa H, Kohchi T, et al (1986) Chloroplast gene organization deduced from

- complete sequence of liverwort *marchantia polymorpha* chloroplast DNA. *Nature* 322:572–574. <https://doi.org/10.1038/322572a0>
- Okonechnikov K, Golosova O, Fursov M, et al (2012) Unipro UGENE: A unified bioinformatics toolkit. *Bioinformatics* 28:1166–1167. <https://doi.org/10.1093/bioinformatics/bts091>
- Plöchlinger M, Torabi S, Rantala M, et al (2016) The low molecular weight protein psai stabilizes the light-harvesting complex II docking site of photosystem I. *Plant Physiol* 172:450–463. <https://doi.org/10.1104/pp.16.00647>
- R Core Team (2020) R: A language and environment for statistical computing. R Found Stat Comput
- Ramírez-Barahona S, Sauquet H, Magallón S (2020) The delayed and geographically heterogeneous diversification of flowering plant families. *Nat Ecol Evol* 4:1232–1238. <https://doi.org/10.1038/s41559-020-1241-3>
- Rodríguez-Ezpeleta N, Brinkmann H, Burey SC, et al (2005) Monophyly of primary photosynthetic eukaryotes: Green plants, red algae, and glaucophytes. *Curr Biol* 15:1325–1330. <https://doi.org/10.1016/j.cub.2005.06.040>
- Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC, et al (2017) DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol Biol Evol* 34:3299–3302. <https://doi.org/10.1093/molbev/msx248>
- Saina JK, Gichira AW, Li ZZ, et al (2018) The complete chloroplast genome sequence of *Dodonaea viscosa*: comparative and phylogenetic analyses. *Genetica* 146:101–113. <https://doi.org/10.1007/s10709-017-0003-x>
- Sato N (2020) Complex origins of chloroplast membranes with photosynthetic machineries: multiple transfers of genes from divergent organisms at different times or a single endosymbiotic event? *J Plant Res* 133:15–33. <https://doi.org/10.1007/s10265-019-01157-z>
- Schöttler MA, Thiele W, Belkuis K, et al (2017) The plastid-encoded Psal subunit stabilizes photosystem i during leaf senescence in tobacco. *J Exp Bot* 68:1137–1155. <https://doi.org/10.1093/jxb/erx009>
- Shen J, Zhang X, Landis JB, et al (2020) Plastome Evolution in *Dolomiaea* (Asteraceae, Cardueae) Using Phylogenomic and Comparative Analyses. *Front Plant Sci* 11:1–15. <https://doi.org/10.3389/fpls.2020.00376>
- Shinozaki K, Ohme M, Tanaka M, et al (1986) The complete nucleotide sequence of the tobacco chloroplast genome: Its gene organization and expression. *EMBO J* 5:2043–2049. <https://doi.org/10.1002/j.1460-2075.1986.tb04464.x>
- Sobreiro MB, Vieira LD, Nunes R, et al (2020) Chloroplast genome assembly of *Handroanthus impetiginosus*: comparative analysis and molecular evolution in Bignoniaceae. *Planta* 252:1–16. <https://doi.org/10.1007/s00425-020-03498-9>
- Souza UJB de, Nunes R, Targueta CP, et al (2019) The complete chloroplast genome of *Stryphnodendron adstringens* (Leguminosae - Caesalpinioideae): comparative analysis with related Mimosoid species. *Sci Rep* 9:1–12. <https://doi.org/10.1038/s41598-019-50620-3>
- Swanson WJ, Nielsen R, Yang Q (2003) Pervasive adaptive evolution in mammalian fertilization proteins. *Mol Biol Evol* 20:18–20. <https://doi.org/10.1093/oxfordjournals.molbev.a004233>
- Talavera G, Castresana J (2007) Improvement of phylogenies after removing divergent and

- ambiguously aligned blocks from protein sequence alignments. *Syst Biol* 56:564–577. <https://doi.org/10.1080/10635150701472164>
- Tillich M, Lehwark P, Pellizzer T, et al (2017) GeSeq - Versatile and accurate annotation of organelle genomes. *Nucleic Acids Res* 45:W6–W11. <https://doi.org/10.1093/nar/gkx391>
- Twyford AD, Ness RW (2017) Strategies for complete plastid genome sequencing. *Mol Ecol Resour* 17:858–868. <https://doi.org/10.1111/1755-0998.12626>
- Urdampilleta JD, Ferrucci MS, Vanzela ALL (2012) Cytogenetic studies in South American species of *Serjania* (Sapindaceae: Paullinieae). *Plant Biosyst* 146:835–846. <https://doi.org/10.1080/11263504.2012.705349>
- Vaidya G, Lohman DJ, Meier R (2011) SequenceMatrix: Concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics* 27:171–180. <https://doi.org/10.1111/j.1096-0031.2010.00329.x>
- Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: Features and applications. *Trends Biotechnol* 23:48–55. <https://doi.org/10.1016/j.tibtech.2004.11.005>
- Wang W, Chen S, Zhang X (2018) Whole-Genome comparison reveals divergent IR borders and mutation hotspots in chloroplast genomes of herbaceous bamboos (Bambusoideae: Olyreae). *Molecules* 23:1–20. <https://doi.org/10.3390/molecules23071537>
- Wang YH, Qu XJ, Chen SY, et al (2017) Plastomes of Mimosoideae: structural and size variation, sequence divergence, and phylogenetic implication. *Tree Genet Genomes* 13:. <https://doi.org/10.1007/s11295-017-1124-1>
- Wheeler TJ, Eddy SR (2013) Nhmmer: DNA homology search with profile HMMs. *Bioinformatics* 29:2487–2489. <https://doi.org/10.1093/bioinformatics/btt403>
- Wicke S, Schneeweiss GM, dePamphilis CW, et al (2011) The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. *Plant Mol Biol* 76:273–297. <https://doi.org/10.1007/s11103-011-9762-4>
- Wu Y, Liu F, Yang DG, et al (2018) Comparative chloroplast genomics of *Gossypium* species: Insights into repeat sequence variations and phylogeny. *Front Plant Sci* 9:1–14. <https://doi.org/10.3389/fpls.2018.00376>
- Xiao-Ming Z, Junrui W, Li F, et al (2017) Inferring the evolutionary mechanism of the chloroplast genome size by comparing whole-chloroplast genome sequences in seed plants. *Sci Rep* 7:1–10. <https://doi.org/10.1038/s41598-017-01518-5>
- Xiong AS, Peng RH, Zhuang J, et al (2009) Gene duplication, transfer, and evolution in the chloroplast genome. *Biotechnol Adv* 27:340–347. <https://doi.org/10.1016/j.biotechadv.2009.01.012>
- Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586–1591. <https://doi.org/10.1093/molbev/msm088>
- Yang Z, Swanson WJ (2002) Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. *Mol Biol Evol* 19:49–57. <https://doi.org/10.1093/oxfordjournals.molbev.a003981>
- Zhang J, Ruhlman TA, Sabir J, et al (2015) Coordinated rates of evolution between interacting plastid and nuclear genes in geraniaceae. *Plant Cell* 27:563–573. <https://doi.org/10.1105/tpc.114.134353>
- Zhou T, Chen C, Wei Y, et al (2016) Comparative transcriptome and chloroplast genome analyses of two related dipteronia species. *Front Plant Sci* 7:1–13.

Material Suplementar

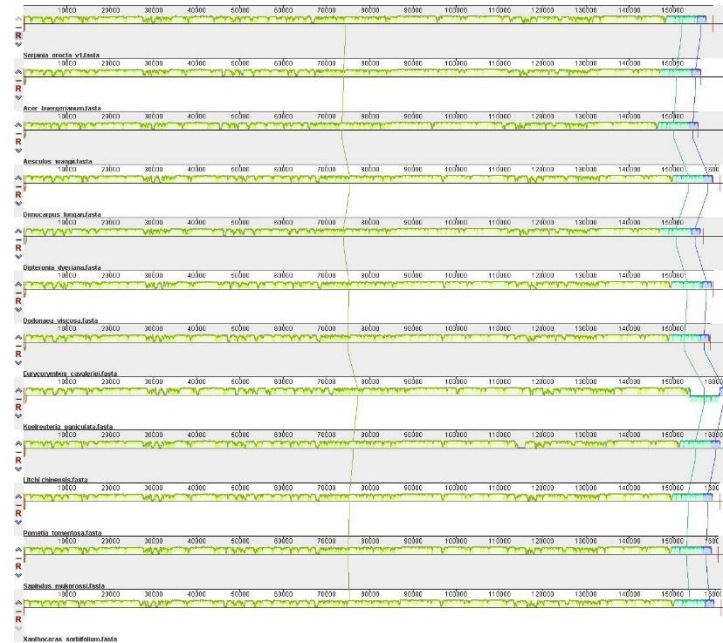


Figura S1. Alinhamento progressivo dos genomas cloroplastidiais das doze espécies da família Sapindaceae

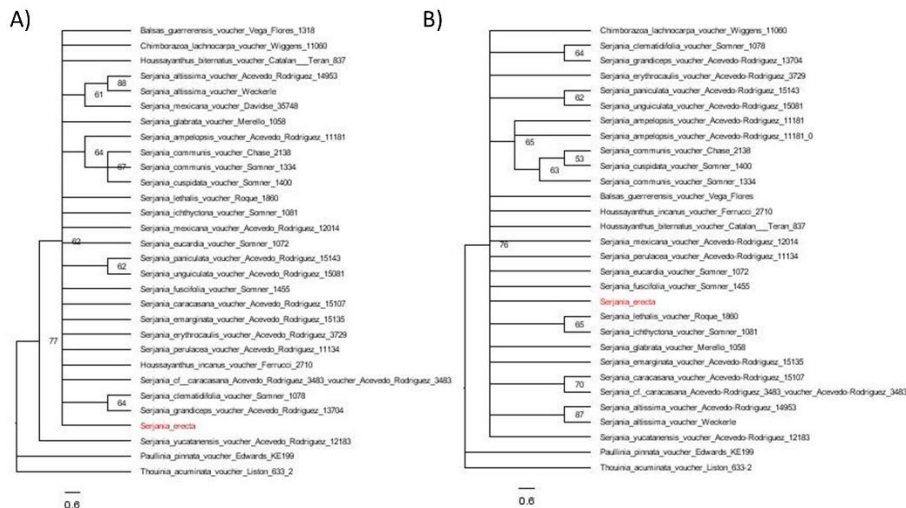


Figura S2. Árvore filogenética de máxima verossimilhança (ML) para o gênero *Serjania*, baseado em marcadores moleculares. Dados obtidos dos estudos de Buerki et al (2009), Buerki et al (2010) e Acevedo-Rodríguez et al (2017). A) Árvore filogenética ML utilizando os marcadores moleculares matK, rpoB, trnD-trnT, trnK-matK, trnL(intron) e trnL-trnF. B) Árvore filogenética ML utilizando o marcador trnL(intron).

Tabela S1. Número de referência no Genbank dos genomas cloroplastidiais utilizados na montagem da árvore filogenética utilizando sequências codificadoras de proteínas

<i>Species</i>	family	ID
<i>Serjania erecta</i>	Sapindaceae	In this work
<i>Acer buergerianum</i>	Sapindaceae	NC_034744.1
<i>Acer truncatum</i>	Sapindaceae	NC_037211.1
<i>Acer longipes</i>	Sapindaceae	NC_049126.1
<i>Aesculus chinensis</i>	Sapindaceae	NC_046788.1
<i>Aesculus wangii</i>	Sapindaceae	NC_035955.1
<i>Dimocarpus longan</i>	Sapindaceae	NC_037447.1
<i>Dipteronia sinensis</i>	Sapindaceae	NC_029338.1
<i>Dipteronia dyeriana</i>	Sapindaceae	NC_031899.1
<i>Dodonaea viscosa</i>	Sapindaceae	NC_036099.1
<i>Eurycorymbus cavaleriei</i>	Sapindaceae	NC_037443.1
<i>Koelreuteria paniculata</i>	Sapindaceae	NC_037176.1
<i>Litchi chinensis</i>	Sapindaceae	NC_035238.1
<i>Pometia tomentosa</i>	Sapindaceae	NC_048999.1
<i>Sapindus mukorossi</i>	Sapindaceae	NC_025554.1
<i>Xanthoceras sorbifolium</i>	Sapindaceae	NC_037448.1
<i>Khaya senegalensis</i>	Meliaceae	NC_037362.1
<i>Cedrela odorata</i>	Meliaceae	NC_037251.1
<i>Citrus sinensis</i>	Rutaceae	NC_008334.1
<i>Ruta graveolens</i>	Rutaceae	NC_045946.1
<i>Mangifera indica</i>	Anacardiaceae	NC_035239.1
<i>Boswellia sacra</i>	Burseraceae	NC_029420.1
<i>Gossypium hirsutum</i>	Malvaceae	NC_007944.1

Tabela S2. Números de referência do Genbank das sequências utilizadas para análise filogenética do gênero *Serjania*.

Species	referencia	matK	rpoB	trnD-trnT	trnK-matK	trnL intron	trnL-trnF
<i>Serjania altissima</i>	Buerki et al, 2009		EU720840			EU721269	EU721457
<i>Serjania communis</i>	Buerki et al, 2009	EU720640	EU720805	EU720969	EU721077	EU721241	EU721429
<i>Serjania glabrata</i>	Buerki et al, 2009	EU720703	EU720905	EU721010	EU721141	EU721327	EU721515
<i>Serjania mexicana</i>	Buerki et al, 2010	EU720704	EU720906		EU721142		
<i>Serjania altissim</i>	Acevedo-Rodríguez et al, 2017					KX585036	
<i>Serjania ampelopsis</i>	Acevedo-Rodríguez et al, 2017					KX585037	
<i>Serjania caracasana</i>	Acevedo-Rodríguez et al, 2017					KX585038	
<i>Serjania cf. caracasana</i>	Acevedo-Rodríguez et al, 2017					KX585039	
<i>Serjania clematidifolia</i>	Acevedo-Rodríguez et al, 2017					KX585040	
<i>Serjania communis</i>	Acevedo-Rodríguez et al, 2017					KX585041	
<i>Serjania cuspidata</i>	Acevedo-Rodríguez et al, 2017					KX585042	
<i>Serjania emarginata</i>	Acevedo-Rodríguez et al, 2017					KX585043	
<i>Serjania erythrocaulis</i>	Acevedo-Rodríguez et al, 2017					KX585044	
<i>Serjania eucardia</i>	Acevedo-Rodríguez et al, 2017					KX585045	
<i>Serjania fuscifolia</i>	Acevedo-Rodríguez et al, 2017					KX585046	
<i>Serjania grandiceps</i>	Acevedo-Rodríguez et al, 2017					KX585047	
<i>Serjania ichthyoctona</i>	Acevedo-Rodríguez et al, 2017					KX585048	
<i>Serjania lethalis</i>	Acevedo-Rodríguez et al, 2017					KX585049	
<i>Serjania marginata</i>	Acevedo-Rodríguez et al, 2017					KX585051	
<i>Serjania mexicana</i>	Acevedo-Rodríguez et al, 2017					KX585052	
<i>Serjania paniculata</i>	Acevedo-Rodríguez et al, 2017					KX585054	
<i>Serjania perulacea</i>	Acevedo-Rodríguez et al, 2017					KX585055	
<i>Serjania unguiculata</i>	Acevedo-Rodríguez et al, 2017					KX585056	
<i>Serjania yucatanensis</i>	Acevedo-Rodríguez et al, 2017					KX585057	
<i>Chimborazoa lachnocarpa</i>	Acevedo-Rodríguez et al, 2017					KX585014	
<i>Balsas guerrerensis</i>	Acevedo-Rodríguez et al, 2017					KX585005	
<i>Houssayanthus biternatus</i>	Acevedo-Rodríguez et al, 2017					KX585018	
<i>Houssayanthus incanus</i>	Acevedo-Rodríguez et al, 2017					KX585019	
<i>Paullinia pinnata</i>	Buerki et al, 2009	EU720575	EU720730	EU720926	EU721023	EU721167	EU721355
<i>Thouinia acuminata</i>	Buerki et al, 2009	EU720647	EU720814		EU721084	EU721249	EU721437

Tabela S3. Categorização dos genes anotados no genoma cloroplastidial de *Serjania erecta*.

Category	Gene groups	Gene names
Self-replication	Large subunit of ribosomal proteins	<i>rpl2^{1*}, rpl14, rpl16¹, rpl20, rpl22*, rpl23*, rpl32, rpl33, rpl36</i>
	Small subunit of ribosomal proteins	<i>rps2, rps3*, rps4, rps7*, rps8, rps11, rps12^{2*}, rps14, rps15, rps16¹, rps18, rps19*</i>
	DNA-dependent RNA polymerase	<i>rpoA, rpoB, rpoC1¹, rpoC2</i>
	Ribosomal RNA genes	<i>rrn4.5*, rrn5*, rrn16*, rrn23*</i>
	Transfer RNA genes	<i>trnA-UGC^{1*}, trnC-GCA, trnD-GUC, trnE-UUC, trnE-UUC*, trnF-GAA, trnG-GCC, trnH-GUG, trnK-UUU¹, trnL-CAA*, trnL-UAA¹, trnL-UAG, trnM-CAU**, trnN-GUU*, trnP-UGG, trnR-UCU, trnQ-UUG, trnR-ACG*, trnS-GCU, trnS-GGA, trnS-UGA, trnT-CGU¹, trnT-GGU, trnT-UGU, trnV-GAC*, trnV-UAC¹, trnW-CCA, trnY-GUA</i>
Photosynthesis	Photosystem I	<i>psaA, psaB, psaC, psaI, psaJ</i>
	Photosystem II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>
NADH dehydrogenase	NADH dehydrogenase	<i>ndhA¹, ndhB^{1*}, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>
	Cytochrome b/f complex	<i>petA, petB¹, petD¹, petG, petL, petN</i>
	ATP synthase	<i>atpA, atpB, atpE, atpF¹, atpH, atpI</i>
Other genes	RubisCo large subunit	<i>rbcL</i>
	Maturase K	<i>matK</i>
	Envelope membrane protein	<i>cemA</i>
	Subunit of acetyl-CoA carboxylase	<i>accD</i>
	C-type cytochrome synthesis gene	<i>ccsA</i>
	Protease	<i>clpP²</i>
	Conserved hypothetical chloroplast open reading frames	<i>ycf1, ycf2*, ycf3², ycf4</i>
Pseudogenes		<i>infA, ycf1</i>

Nota: *: genes duplicados; **: genes com 4 copias ¹: genes com um *intron*; ²: genes com dois *introns*; Category: categorias; Gene groups: grupo de genes; Gene names: Nome dos genes

Tabela S4. Uso de codons do genoma cloroplastidial de *Serjania erecta*.

Amino Acid	Codon	Number	Fraction	Amino Acid	Codon	Number	Fraction	
Ala	GCA	391	0.27	Leu	CTT	590	0.21	
	GCC	229	0.16		TTA	855	0.30	
	GCG	195	0.14		TTG	588	0.21	
	GCT	621	0.43	Lys	AAA	1136	0.75	
Arg	AGA	459	0.28		AAG	384	0.25	
	AGG	197	0.12	Met	ATG	628	1.00	
	CGA	373	0.23	Phe	TTC	544	0.34	
	CGC	141	0.09		TTT	1052	0.66	
	Asn	CGG	156	0.10	Pro	CCA	293	0.27
		CGT	309	0.19		CCC	250	0.23
AAC	307	0.23	CCG	173		0.16		
AAT	1032	0.77	CCT	384		0.35		
Asp	GAC	228	0.22	Ser	AGC	140	0.07	
	GAT	818	0.78		AGT	414	0.20	
Cys	TGC	91	0.29		TCA	428	0.21	
	TGT	227	0.71		TCC	330	0.16	
Gln	CAA	742	0.77		TCG	215	0.10	
	CAG	220	0.23		TCT	541	0.26	
Glu	GAA	1059	0.74	Thr	ACA	404	0.30	
	GAG	369	0.26		ACC	261	0.19	
Gly	GGA	718	0.39		ACG	169	0.13	
	GGC	188	0.10		ACT	511	0.38	
	GGG	358	0.19	Trp	TGG	454	1.00	
	GGT	575	0.31	Tyr	TAC	197	0.20	
His	CAC	169	0.27		TAT	778	0.80	
	CAT	467	0.73	Val	GTA	537	0.37	
Ile	ATA	720	0.32		GTC	181	0.12	
	ATC	429	0.19		GTG	198	0.14	
	ATT	1131	0.50		GTT	543	0.37	
Leu	CTA	388	0.14	End	TAA	53	0.58	
	CTC	217	0.08		TAG	22	0.24	
	CTG	197	0.07		TGA	17	0.18	

Tabela S5. Tamanho de íntrons e seus respectivos genes em genomas cloroplastidiais da família Sapindaceae

Region	LSC														IR					SCC
Species/Gene	<i>trnK-UUU</i>	<i>rps16</i>	<i>trnT-CGU</i>	<i>atpF</i>	<i>rpoC1</i>	<i>ycf3 1</i>	<i>ycf3 1</i>	<i>trnL-UAA</i>	<i>trnV-UAC</i>	<i>clpP 1</i>	<i>clpP 2</i>	<i>petB</i>	<i>petD</i>	<i>rpl16</i>	<i>rpl2</i>	<i>ndhB</i>	<i>rps12</i>	<i>trnE-UUC</i>	<i>trnA-UGC</i>	<i>ndhA</i>
<i>Serjania erecta</i>	2556	854	720	787	709	764	744	543	597	671	826	804	743	1056	663	607	541	956	802	1126
<i>Acer buergerianum</i>	2667	812	715	767	692	730	732	515	590	596	853	786	747	1053	664	682	545	957	807	1099
<i>Aesculus wangii</i>	2542	824	720	755	705	744	729	537	588	615	846	804	739	1051	664	681	537	958	840	1101
<i>Dimocarpus longan</i>	2511	839	725	754	717	745	730	536	601	655	874	798	766	1063	663	672	541	956	839	1082
<i>Dipteronia dyeriana</i>	2501	794	715	750	700	733	730	522	593	612	840	797	736	1073	664	682	546	957	807	1085
<i>Dodonaea viscosa</i>	2513	840	absent gene	758	719	824	732	553	591	662	875	816	733	1045	664	680	533	951	803	1106
<i>Eurycorymbus cavaleriei</i>	2498	834	absent gene	762	740	767	732	544	592	658	850	809	728	926	664	680	533	957	810	1089
<i>Koelreuteria paniculata</i>	2527	841	730	764	717	756	741	550	605	641	985	785	757	1074	663	672	541	963	841	522
<i>Litchi chinensis</i>	2511	837	725	756	714	745	735	536	600	655	867	798	812	1042	663	672	541	956	846	1110
<i>Pometia tomentosa</i>	2512	839	725	756	712	746	730	536	605	656	867	796	742	1028	663	672	541	957	840	1078
<i>Sapindus mukorossi</i>	2536	832	730	778	714	766	731	538	606	664	870	793	745	985	663	672	541	952	840	1102
<i>Xanthoceras sorbifolium</i>	2465	823	721	767	700	756	730	543	593	637	907	792	736	1055	664	681	536	961	840	1118

Tabela S6. Regiões hotspot de diversidade nucleotídica, valores médios e medianos da diversidade nucleotídica para 12 genomas cloroplastidiais da família Sapindaceae.

Regions	Nucleotide diversity (π)
<i>tRNA-Lys-rps16</i>	0.0981
<i>atpI-rps2</i>	0.0844
<i>rpoB-tRNA-Asp</i>	0.0817
<i>tRNA-Tyr-psbD</i>	0.0837
<i>psbZ</i>	0.0865
<i>ndhC-tRNA-Val</i>	0.1052
<i>petA-psbJ</i>	0.0963
<i>ndhF</i>	0.1089
<i>rpl32-ccsA</i>	0.1133
<i>ycf1</i>	0.1473
Median	0.0357
Average	0.0380
Standard deviation	0.0274

Tabela S7. Regiões hotspot de diversidade nucleotídica, valores médios e medianos da diversidade nucleotídica para 6 genomas cloroplastidiais da subfamília Sapindoideae.

Regions	Nucleotide diversity (π)
<i>tRNA-Lys-rps16</i>	0.07167
<i>rps16-tRNA-Gln</i>	0.068
<i>tRNA-Ser-tRNA-Gly</i>	0.05967
<i>tRNA-Cys-psbM</i>	0.06378
<i>ndhC-tRNA-Val</i>	0.08156
<i>ycf4-cemA</i>	0.06833
<i>petA-psbJ</i>	0.07044
<i>psbE-petL</i>	0.06811
<i>ndhF</i>	0.06956
<i>rpl32-ccsA</i>	0.08067
<i>ycf1</i>	0.09744
Median	0.0227
Average	0.0244
Standard deviation	0.0187

Tabela S8. Seleção local em genes codificadores de proteínas de genoma cloroplastidial de Sapindaceae.

Gene	ka/ks	LH1	NP1	LH2	NP2	LH7	NP7	LH8	NP8	LH8a	NP8a	M2xM1	P_M2xM1	FDR_M2xM1	M8xM7	P_M8xM7	FDR_M8xM7	M8axM8	P_M8axM8	FDR_M8axM8
accD	0.35	-4073.91	25	-4073.10	27	-4074.18	25	-4072.76	27	-4073.88	26	1.62	0.44	1.00	2.84	0.24	0.83	2.25	0.33	1.00
atpA	0.09	-3274.52	25	-3273.30	27	-3278.74	25	-3273.71	27	-3274.46	26	2.44	0.30	1.00	10.06	0.01	0.05	1.50	0.47	1.00
atpB	0.08	-3201.72	25	-3201.72	27	-3202.68	25	-3201.67	27	-3201.67	26	0.00	1.00	1.00	2.03	0.36	1.00	0.00	1.00	1.00
atpE	0.11	-823.44	25	-823.44	27	-823.83	25	-823.44	27	-823.46	26	0.00	1.00	1.00	0.78	0.68	1.00	0.04	0.98	1.00
atpF	0.33	-1297.96	25	-1297.96	27	-1297.97	25	-1297.96	27	-1297.96	26	0.00	1.00	1.00	0.03	0.99	1.00	0.00	1.00	1.00
atpH	0.00	-461.48	25	-461.48	27	-461.48	25	-461.48	27	-461.48	26	0.00	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00
atpI	0.11	-1638.28	25	-1638.28	27	-1638.28	25	-1638.28	27	-1638.28	26	0.00	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00
ccsA	0.51	-3275.11	25	-3275.11	27	-3275.13	25	-3271.92	27	-3276.28	26	0.00	1.00	1.00	6.43	0.04	0.19	8.71	0.01	0.14
cemA	0.30	-1739.06	25	-1739.06	27	-1738.76	25	-1738.76	27	-1738.76	26	0.00	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00
clpP	0.70	-1659.93	25	-1654.62	27	-1659.94	25	-1654.65	27	-1659.93	26	10.63	0.00	0.06	10.58	0.01	0.04	10.56	0.01	0.07
matK	0.45	-4928.26	25	-4925.52	27	-4929.94	25	-4925.13	27	-4928.28	26	5.49	0.06	0.47	9.61	0.01	0.05	6.29	0.04	0.30
ndhA	0.22	-2611.54	25	-2605.86	27	-2613.28	25	-2606.35	27	-2611.56	26	11.35	0.00	0.05	13.87	0.00	0.01	10.42	0.01	0.07
ndhB	0.33	-2389.78	25	-2389.61	27	-2389.92	25	-2389.60	27	-2389.78	26	0.35	0.84	1.00	0.65	0.72	1.00	0.36	0.83	1.00
ndhC	0.11	-711.31	25	-710.46	27	-712.68	25	-710.15	27	-711.31	26	1.69	0.43	1.00	5.06	0.08	0.34	2.32	0.31	1.00
ndhD	0.22	-3958.77	25	-3958.48	27	-3959.68	25	-3958.43	27	-3958.77	26	0.59	0.75	1.00	2.50	0.29	0.92	0.69	0.71	1.00
ndhE	0.11	-635.99	25	-635.99	27	-635.53	25	-635.53	27	-635.53	26	0.00	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00
ndhF	0.26	-7264.81	25	-7228.47	27	-7265.20	25	-7224.43	27	-7262.09	26	72.68	0.00	0.00	81.55	0.00	0.00	75.33	0.00	0.00
ndhG	0.22	-1239.29	25	-1239.29	27	-1239.31	25	-1239.22	27	-1239.26	26	0.00	1.00	1.00	0.17	0.92	1.00	0.07	0.96	1.00
ndhH	0.12	-2759.59	25	-2759.54	27	-2760.94	25	-2759.55	27	-2759.60	26	0.09	0.96	1.00	2.78	0.25	0.83	0.10	0.95	1.00
ndhI	0.10	-1215.19	25	-1215.19	27	-1215.09	25	-1214.38	27	-1214.82	26	0.00	1.00	1.00	1.43	0.49	1.00	0.89	0.64	1.00
ndhJ	0.13	-1113.51	25	-1113.51	27	-1113.38	25	-1113.38	27	-1113.38	26	0.00	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00
ndhK	0.25	-1576.85	25	-1575.39	27	-1576.96	25	-1575.40	27	-1577.23	26	2.92	0.23	1.00	3.12	0.21	0.81	3.67	0.16	0.82
petA	0.11	-2246.16	25	-2246.16	27	-2245.30	25	-2253.03	27	-2245.30	26	0.00	1.00	1.00	-15.46	1.00	1.00	-15.46	1.00	1.00
petB	0.01	-1261.55	25	-1261.54	27	-1261.55	25	-1261.56	27	-1261.55	26	0.00	1.00	1.00	-0.02	1.00	1.00	-0.02	1.00	1.00
petD	0.08	-964.79	25	-963.68	27	-968.68	25	-963.70	27	-964.79	26	2.22	0.33	1.00	9.94	0.01	0.05	2.17	0.34	1.00

petG	0.05	-222.82	25	-222.82	27	-223.79	25	-222.33	27	-222.33	26	0.00	1.00	1.00	2.91	0.23	0.83	0.00	1.00	1.00
petL	0.25	-199.26	25	-199.26	27	-199.27	25	-199.27	27	-199.27	26	0.00	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00
petN	0.00	-137.99	25	-137.99	27	-137.99	25	-137.99	27	-137.99	26	0.00	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00
psaA	0.06	-4330.07	25	-4330.07	27	-4332.01	25	-4330.05	27	-4330.05	26	0.00	1.00	1.00	3.92	0.14	0.57	0.00	1.00	1.00
psaB	0.04	-4349.55	25	-4348.24	27	-4357.64	25	-4348.28	27	-4349.58	26	2.63	0.27	1.00	18.72	0.00	0.00	2.61	0.27	1.00
psaC	0.00	-519.35	25	-519.35	27	-519.35	25	-519.35	27	-519.35	26	0.00	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00
psaI	1.03	-235.59	25	-232.07	27	-235.69	25	-232.07	27	-235.59	26	7.04	0.03	0.29	7.25	0.03	0.14	7.04	0.03	0.23
psaJ	0.28	-315.31	25	-315.31	27	-315.31	25	-315.31	27	-315.31	26	0.00	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00
psbA	0.02	-2035.98	25	-2035.98	27	-2036.68	25	-2035.97	27	-2035.97	26	0.00	1.00	1.00	1.42	0.49	1.00	0.00	1.00	1.00
psbB	0.09	-3230.47	25	-3230.47	27	-3230.41	25	-3230.41	27	-3230.41	26	0.00	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00
psbC	0.05	-2776.21	25	-2776.19	27	-2776.56	25	-2776.19	27	-2776.21	26	0.03	0.99	1.00	0.73	0.69	1.00	0.03	0.98	1.00
psbD	0.01	-1894.19	25	-1894.19	27	-1894.21	25	-1894.21	27	-1894.19	26	0.00	1.00	1.00	0.00	1.00	1.00	-0.03	1.00	1.00
psbE	0.02	-422.18	25	-422.18	27	-422.18	25	-422.18	27	-422.18	26	0.00	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00
psbF	0.14	-254.51	25	-254.51	27	-254.51	25	-254.51	27	-254.51	26	0.00	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00
psbH	0.43	-539.75	25	-539.23	27	-540.34	25	-539.23	27	-539.75	26	1.04	0.59	1.00	2.22	0.33	1.00	1.04	0.59	1.00
psbI	0.14	-256.65	25	-256.65	27	-256.66	25	-256.66	27	-256.66	26	0.00	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00
psbJ	0.18	-222.25	25	-222.23	27	-222.34	25	-222.23	27	-222.25	26	0.03	0.99	1.00	0.22	0.90	1.00	0.03	0.99	1.00
psbK	0.17	-421.83	25	-421.66	27	-422.17	25	-421.66	27	-421.84	26	0.35	0.84	1.00	1.03	0.60	1.00	0.36	0.84	1.00
psbL	0.00	-173.27	25	-173.27	27	-173.28	25	-173.27	27	-173.27	26	0.00	1.00	1.00	0.02	0.99	1.00	0.00	1.00	1.00
psbM	0.06	-177.87	25	-177.87	27	-177.88	25	-177.88	27	-177.88	26	0.00	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00
psbN	0.05	-259.86	25	-259.85	27	-260.06	25	-259.86	27	-259.86	26	0.00	1.00	1.00	0.40	0.82	1.00	0.00	1.00	1.00
psbT	0.29	-189.64	25	-189.55	27	-189.64	25	-189.55	27	-189.64	26	0.18	0.92	1.00	0.18	0.92	1.00	0.18	0.92	1.00
psbZ	0.08	-350.35	25	-350.35	27	-350.36	25	-350.36	27	-350.36	26	0.00	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00
rbcL	0.11	-3030.71	25	-3026.14	27	-3046.24	25	-3026.72	27	-3030.71	26	9.15	0.01	0.11	39.05	0.00	0.00	7.99	0.02	0.18
rpl14	0.06	-839.41	25	-839.41	27	-839.11	25	-839.11	27	-839.11	26	0.00	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00
rpl16	0.18	-915.98	25	-915.98	27	-915.96	25	-915.96	27	-915.96	26	0.00	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00
rpl20	0.20	-901.08	25	-900.94	27	-901.17	25	-900.84	27	-901.08	26	0.27	0.87	1.00	0.67	0.72	1.00	0.48	0.78	1.00
rpl22	0.31	-1323.99	25	-1323.99	27	-1323.89	25	-1323.76	27	-1324.06	26	0.00	1.00	1.00	0.24	0.88	1.00	0.59	0.74	1.00

rpl23	0.52	-463.85	25	-463.80	27	-463.85	25	-463.80	27	-463.85	26	0.09	0.95	1.00	0.10	0.95	1.00	0.09	0.95	1.00
rpl2	0.22	-1261.57	25	-1261.48	27	-1261.63	25	-1261.48	27	-1261.57	26	0.17	0.92	1.00	0.30	0.86	1.00	0.17	0.92	1.00
rpl32	0.15	-539.64	25	-536.69	27	-540.63	25	-535.81	27	-539.45	26	5.91	0.05	0.45	9.65	0.01	0.05	7.29	0.03	0.22
rpl33	0.46	-525.22	25	-524.61	27	-525.38	25	-524.59	27	-525.22	26	1.22	0.54	1.00	1.58	0.45	1.00	1.26	0.53	1.00
rpl36	0.12	-214.55	25	-214.12	27	-215.06	25	-214.27	27	-214.55	26	0.87	0.65	1.00	1.58	0.45	1.00	0.57	0.75	1.00
rpoA	0.30	-2667.83	25	-2667.74	27	-2668.15	25	-2667.61	27	-2667.83	26	0.17	0.92	1.00	1.09	0.58	1.00	0.44	0.80	1.00
rpoB	0.17	-7098.23	25	-7095.74	27	-7101.10	25	-7095.75	27	-7098.28	26	4.97	0.08	0.53	10.71	0.00	0.04	5.05	0.08	0.47
rpoC1	0.18	-4519.04	25	-4516.91	27	-4521.29	25	-4516.76	27	-4519.05	26	4.24	0.12	0.71	9.06	0.01	0.06	4.58	0.10	0.56
rpoC2	0.30	-10606.31	25	-10599.94	27	-10608.24	25	-10600.35	27	-10606.34	26	12.75	0.00	0.03	15.79	0.00	0.00	11.99	0.00	0.05
rps11	0.24	-938.01	25	-938.01	27	-937.56	25	-937.56	27	-937.56	26	0.00	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00
rps12	0.48	-566.04	25	-566.04	27	-566.04	25	-566.05	27	-566.04	26	0.00	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00
rps14	0.25	-692.80	25	-692.66	27	-692.82	25	-692.67	27	-692.89	26	0.27	0.87	1.00	0.30	0.86	1.00	0.45	0.80	1.00
rps15	0.39	-719.43	25	-719.10	27	-719.82	25	-719.11	27	-719.44	26	0.67	0.72	1.00	1.43	0.49	1.00	0.67	0.72	1.00
rps16	0.31	-688.38	25	-685.67	27	-688.69	25	-685.81	27	-688.38	26	5.41	0.07	0.47	5.76	0.06	0.25	5.16	0.08	0.47
rps18	0.20	-746.90	25	-746.30	27	-746.91	25	-746.18	27	-746.93	26	1.21	0.55	1.00	1.47	0.48	1.00	1.50	0.47	1.00
rps19	0.20	-534.97	25	-534.97	27	-534.98	25	-534.98	27	-534.98	26	0.00	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00
rps3	0.16	-1476.17	25	-1476.15	27	-1476.58	25	-1476.16	27	-1476.18	26	0.03	0.98	1.00	0.84	0.66	1.00	0.04	0.98	1.00
rps4	0.25	-1300.69	25	-1300.69	27	-1300.48	25	-1300.46	27	-1300.46	26	0.00	1.00	1.00	0.05	0.98	1.00	0.00	1.00	1.00
rps7	0.07	-638.79	25	-638.79	27	-638.79	25	-638.82	27	-638.82	26	0.00	1.00	1.00	-0.05	1.00	1.00	0.00	1.00	1.00
rps8	0.19	-1022.27	25	-1022.27	27	-1021.87	25	-1021.87	27	-1021.87	26	0.00	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00
ycf1	0.67	-20357.34	25	-20293.68	27	-20361.47	25	-20295.16	27	-20357.50	26	127.33	0.00	0.00	132.63	0.00	0.00	124.69	0.00	0.00
ycf2	0.78	-11906.36	25	-11888.09	27	-11907.54	25	-11887.62	27	-11906.35	26	36.53	0.00	0.00	39.84	0.00	0.00	37.47	0.00	0.00
ycf3	0.11	-979.20	25	-979.20	27	-979.29	25	-979.19	27	-979.19	26	0.00	1.00	1.00	0.20	0.91	1.00	0.00	1.00	1.00
ycf4	0.27	-1470.97	25	-1470.97	27	-1470.32	25	-1470.32	27	-1470.32	26	0.00	1.00	1.00	0.00	1.00	1.00	0.00	1.00	1.00

Nota LH: *Likelihood* (verossimilhança); NP: *Number of parameters* (Número de parâmetros); FDR: *False Discovery Rate* (Taxa de falsa descoberta); P: *p-value of LTR* (valor de p do LTR); Genes em negrito apresentaram tanto FDR como P-valor inferiores a 0.05.

CONCLUSÕES GERAIS

- Os genomas cloroplastidiais da família da Sapindaceae apresentam uma estrutura conservada quanto ao tamanho de seu genoma, tamanhos da região longa de cópia única (LSC), tamanho da região curta de cópia única (SSC), tamanho das regiões invertidas repetidas (IR) e número e ordem dos genes.
- As variações nos números de genes estão associadas a eventos de expansão e retração das IR, pseudogenização do gene *rps2* e a quantidade de genes codificadores de tRNAs.
- *Serjania erecta* apresentou um genoma cloroplastidial com estrutura similar ao de outras Sapindaceae. Entre os genomas cloroplastidiais avaliados *S. erecta*, apresentou a menor quantidade de repetições complexas e microssatélites (SSR).
- Os genes dos cloroplastos são extremamente conservados na família Sapindaceae e apenas o gene *psaI* não está sobre seleção negativa.
- Os genes *ndhF* e *ycf1* apresentam elevada diversidade nucleotídica e apresentam locus sob seleção positiva, mostrando potencial utilização como marcadores moleculares para a subfamília Sapindoideae.

REFERÊNCIAS

- ACEVEDO-RODRIGUEZ, P. Distributional patterns in Brazilian *Serjania* (Sapindaceae). **Acta Botanica Brasilica**, v. 4, n. 1, p. 69–82, 1990.
- ACEVEDO-RODRÍGUEZ, P. **Systematics of *Serjania* (Sapindaceae) Part I: a revision of *Serjania* sect. *Platycoccus*** *Memoirs of the New York Botanical Garden* 1, 1993.
- ACEVEDO-RODRÍGUEZ, P. et al. Sapindaceae. **Flowering Plants. Eudicots: Sapindales, Cucurbitales, Myrtaceae**, p. 357–407, 2011.
- ACEVEDO-RODRÍGUEZ, P. et al. Generic Relationships and Classification of Tribe Paullinieae (Sapindaceae) with a New Concept of Supertribe Paullinioidae. **Systematic Botany**, v. 42, n. 1, p. 96–114, 2017.
- ARRUDA, A. P. C. B. N. **AVALIAÇÃO DA ATIVIDADE ANTIULCEROGÊNICA E TÓXICA DOS EXTRATOS METANÓLICO E CLOROFÓRMICO DAS FOLHAS DE *Serjania erecta* Radlk (SAPINDACEAE)**. [s.l.] Universidade Estadual Paulista, 2011.
- BROGGINI, L. S. C. et al. Behavioral and enzymatic bioassays with *Serjania erecta* Radlk., Sapindaceae, correlated with cognitive dysfunctions. **Brazilian Journal of Pharmacognosy**, v. 20, n. May 2009, p. 519–528, 2010.
- BUERKI, S. et al. Plastid and nuclear DNA markers reveal intricate relationships at subfamilial and tribal levels in the soapberry family (Sapindaceae). **Molecular Phylogenetics and Evolution**, v. 51, n. 2, p. 238–258, 2009.
- BUERKI, S. et al. The abrupt climate change at the Eocene-Oligocene boundary and the emergence of South-East Asia triggered the spread of sapindaceous lineages. **Annals of Botany**, v. 112, n. 1, p. 151–160, 2013.
- BUERKI, S. et al. An updated infra-familial classification of Sapindaceae based on targeted enrichment data. **American Journal of Botany**, v. 108, n. 7, p. 1234–1251, 2021.
- CARDOSO, C. A. L. et al. Phenolic compounds and antioxidant, antimicrobial and antimycobacterial activities of *Serjania erecta* Radlk. (Sapindaceae). **Brazilian Journal of Pharmaceutical Sciences**, v. 49, n. 4, p. 775–782, 2013.
- DONG, W. et al. *ycf1*, the most promising plastid DNA barcode of land plants. **Scientific Reports**, v. 5, p. 8348, 2015.
- DUGAS, D. V. et al. Mimosoid legume plastome evolution: IR expansion, tandem repeat expansions, and accelerated rate of evolution in *clpP*. **Scientific Reports**, v. 5, n. November, p. 1–13, 2015.
- FERNANDES, R. S. et al. Neutralization of pharmacological and toxic activities of Bothrops jararacussu snake venom and isolated myotoxins by *Serjania erecta* methanolic extract and its fractions. **Journal of Venomous Animals and Toxins Including Tropical Diseases**, v. 17, n. 1, p. 85–93, 2011.

FERRUCCI, M. S.; ACEVEDO-RODRÍGUEZ, P. Three new species of *Serjania* (Sapindaceae) from south America. **Systematic Botany**, v. 30, n. 1, p. 153–162, 2005.

FREITAS, S. T. F. DE. **POTENCIAL DE EXTRATOS DE PLANTAS NO CONTROLE DE INSETOS-PRAGA: UM LEVANTAMENTO CIENCIOMÉTRICO E ANÁLISES DE EFEITOS BIOLÓGICOS DE EXTRATOS FOLIARES DE *Serjania erecta* Radlk EM *Chrysodeixis includens***. [s.l.] INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA GOIANO, 2018.

GOMIG, F. et al. Topical anti-inflammatory activity of *Serjania erecta* Radlk (Sapindaceae) extracts. **Journal of Ethnopharmacology**, v. 118, n. 2, p. 220–224, 2008.

GONÇALVES, D. J. P. et al. Incongruence between gene trees and species trees and phylogenetic signal variation in plastid genes. **Molecular Phylogenetics and Evolution**, v. 138, n. May, p. 219–232, 2019.

GOYAL, S. Medicinal Plants of the Genus *Sapindus* (Sapindaceae) - a Review of Their Botany, Phytochemistry, Biological Activity and Traditional Uses. **Journal of Drug Delivery and Therapeutics**, v. 4, n. 5, p. 7–20, 2014.

GUARIM NETO, G. UTILIZAÇÃO DE VEGETAIS NA MEDICINA TRADICIONAL. I. *Serjania erecta* RADLK. (CINCO-FOLHAS). **FLOVET**, n. 1984, p. 14–20, 2009.

GUARIM NETO, G.; SANTANA, S. R.; DA SILVA, J. V. B. Notas etnobotânicas de espécies de sapindaceae jussieu. **Acta Botanica Brasilica**, v. 14, n. 3, p. 327–334, 2000.

GUIMARÃES, C. C. et al. The glycosylated flavonoids vitexin, isovitexin, and quercetrin isolated from *Serjania erecta* Radlk (Sapindaceae) leaves protect PC12 cells against amyloid- β 25-35 peptide-induced toxicity. **Food and Chemical Toxicology**, v. 86, p. 88–94, 2015.

GUIMARÃES, R.; FORNI-MARTINS, E. R. Chromosome numbers and their evolutionary meaning in the Sapindales order: an overview. **Brazilian Journal of Botany**, n. 0123456789, 2021.

HARRINGTON, M. G. et al. Phylogenetic inference in sapindaceae sensu lato using plastid *matK* and *rbcL* DNA sequences. **Systematic Botany**, v. 30, n. 2, p. 366–382, 2005.

LENZA, E.; FERREIRA, J. N. Biologia Floral de *Serjania erecta* Radlk.(Sapindaceae): Um caso de diocia Temporal. **Boletim do Herbário Ezechias Paulo Heringer**, v. 6, p. 23–37, 2000.

LIANG, Q. et al. The genome assembly and annotation of yellowhorn (*Xanthoceras sorbifolium* Bunge). **GigaScience**, v. 8, n. 6, p. 1–15, 2019.

LIDILHONE, H.; GENISE, V. S.; NEUSA, T. *Paullinia cupana* Kunth (Sapindaceae): A review of its ethnopharmacology, phytochemistry and pharmacology. **Journal of**

- Medicinal Plants Research**, v. 7, n. 30, p. 2221–2229, 2013.
- MOHANTA, T. K. et al. Gene loss and evolution of the plastome. **Genes**, v. 11, n. 10, p. 1–23, 2020.
- MOORE, M. J. et al. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. **Proceedings of the National Academy of Sciences of the United States of America**, v. 107, n. 10, p. 4623–4628, 2010.
- MUELLNER-RIEHL, A. N. et al. Molecular phylogenetics and molecular clock dating of Sapindales based on plastid *rbcL*, *atpB* and *trnL-trnF* DNA sequences. **Taxon**, v. 65, n. 5, p. 1019–1036, 2016.
- OHYAMA, K. et al. Chloroplast gene organization deduced from complete sequence of liverwort *marchantia polymorpha* chloroplast DNA. **Nature**, v. 322, n. 6079, p. 572–574, 1986.
- PEREIRA, J. R.; FAMADAS, K. M. Avaliação " in Vitro " Da Eficiência Do Extrato Da Raiz Do Timbó (*Dahlstedtia Pentaphylla*) (Leguminosae , Papilionoidae , Millettiedae) Sobre Boophilus Vale Do Paraíba , São Paulo , Brasil. **Arq. Inst. Biol**, v. 71, n. 4, p. 443–450, 2004.
- PERES, M. K. **ESTRATÉGIAS DE DISPERSÃO DE SEMENTES NO BIOMA CERRADO: CONSIDERAÇÕES ECOLÓGICAS E FILOGENÉTICAS**. [s.l.] UNIVERSIDADE DE BRASÍLIA, 2016.
- RAVIN, N. V. et al. The loss of photosynthetic pathways in the plastid and nuclear genomes of the non-photosynthetic mycoheterotrophic eudicot *Monotropa hypopitys*. **BMC Plant Biology**, v. 16, n. Suppl 3, 2016.
- REFLORA. **Sapindaceae**. Disponível em: <<http://reflora.jbrj.gov.br/reflora/listaBrasil/FichaPublicaTaxonUC/FichaPublicaTaxonUC.do?id=FB216>>. Acesso em: 22 nov. 2021a.
- REFLORA. **Sapindaceae**. Disponível em: <<http://reflora.jbrj.gov.br/reflora/floradobrasil/FB20949>>. Acesso em: 8 set. 2021b.
- SHINOZAKI, K. et al. The complete nucleotide sequence of the tobacco chloroplast genome: Its gene organization and expression. **EMBO Journal**, v. 5, n. 9, p. 2043–2049, 1986.
- TWYFORD, A. D.; NESS, R. W. Strategies for complete plastid genome sequencing. **Molecular Ecology Resources**, v. 17, n. 5, p. 858–868, 2017.
- URDAMPILLETA, J. D.; FERRUCCI, M. S.; VANZELA, A. L. L. Cytogenetic studies in South American species of *Serjania* (Sapindaceae: Paullinieae). **Plant Biosystems**, v. 146, n. 4, p. 835–846, 2012.
- VAN DER HAM, R. W. J. M.; TOMLIK, A. *Serjania* pollen and the origin of the tribe Paullinieae (Sapindaceae). **Review of Palaeobotany and Palynology**, v. 83, n. 1–3, p. 43–53, 1994.

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

XIAO-MING, Z. et al. Inferring the evolutionary mechanism of the chloroplast genome size by comparing whole-chloroplast genome sequences in seed plants. **Scientific Reports**, v. 7, n. 1, p. 1–10, 2017.

YANG, J. et al. De novo genome assembly of the endangered *Acer yangbiense*, a plant species with extremely small populations endemic to Yunnan Province, China. **GigaScience**, v. 8, n. 7, p. 1–10, 2019.