



Universidade Federal de Goiás
Instituto de Ciências Biológicas



Programa de Pós-graduação em Genética e Biologia Molecular

MILLENA SILVA MENDES

**Caracterização do genoma cloroplastidial de *Caryocar cuneatum*
(Caryocaraceae)**

Orientadora: Dra. Mariana Pires de Campos Telles
Coorientador: Dr. Rhewter Nunes

**GOIÂNIA - GO
2023**



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 Outro*: _____

*No caso de mestrado/doutorado profissional, indique o formato do Trabalho de Conclusão de Curso, permitido no documento de área, correspondente ao programa de pós-graduação, orientado pela legislação vigente da CAPES.

Exemplos: Estudo de caso ou Revisão sistemática ou outros formatos.

2. Nome completo do autor

Millena Silva Mendes

3. Título do trabalho

Caracterização do genoma cloroplastidial de *Caryocar cuneatum* (Caryocaraceae)

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(a) autor(a) e ao(a) 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 **Mariana Pires De Campos Telles, Professor do Magistério Superior**, em 07/07/2023, às 13:39, 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 **Millena Silva Mendes, Discente**, em 17/07/2023, às 21:33, 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 **3877199** e o código CRC **18F6F8F4**.

Referência: Processo nº 23070.021355/2023-17

SEI nº 3877199



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 Outro*: _____

*No caso de mestrado/doutorado profissional, indique o formato do Trabalho de Conclusão de Curso, permitido no documento de área, correspondente ao programa de pós-graduação, orientado pela legislação vigente da CAPES.

Exemplos: Estudo de caso ou Revisão sistemática ou outros formatos.

2. Nome completo do autor

Millena Silva Mendes

3. Título do trabalho

Caracterização do genoma cloroplastidial de *Caryocar cuneatum* (Caryocaraceae)

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 **Millena Silva Mendes, Discente**, em 19/12/2023, às 06:37, 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 **4274623** e o código CRC **3FE79DDC**.

Referência: Processo nº 23070.021355/2023-17

SEI nº 4274623

Millena Silva Mendes

**Caracterização do genoma cloroplastidial de *Caryocar cuneatum*.
(Caryocaraceae)**

Dissertação apresentada ao Programa de Pós-Graduação em Genética e Biologia Molecular do Instituto de Ciências Biológicas, 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

Orientadora: Dra. Mariana Pires de Campos Telles

Coorientador: Dr. Rhewter Nunes

GOIÂNIA – GO 2023

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

Mendes, Millena Silva

Caracterização do genoma cloroplastidial de *Caryocar cuneatum* (Caryocaraceae) [manuscrito] / Millena Silva Mendes. - 2023.
45 f.: il.

Orientador: Prof. Mariana Pires de Campos Telles; co-orientador Rhewter Nunes.

Dissertação (Mestrado) - Universidade Federal de Goiás, Instituto de Ciências Biológicas (ICB), Programa de Pós-Graduação em Genética e Biologia Molecular, Goiânia, 2023.

Bibliografia. Anexos.

1. Genoma organelar. 2. Plastoma. 3. Comparação genômica. I. Telles, Mariana Pires de Campos, orient. II. Título.

CDU 57



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

Ata nº 113 da sessão de Defesa de Dissertação de **Millena Silva Mendes**, que confere o título de Mestre(a) em **Genética e Biologia Molecular**, na área de concentração em **Genética e Biologia Molecular**.

Ao/s **vinte e oito dias do mês de abril de dois mil e vinte e três**, a partir da(s) **14h30**, no(a) sala do **INCT no ICB I**, realizou-se a sessão pública de Defesa de Dissertação intitulada **“Caracterização do genoma cloroplastidial de *Caryocar cuneatum* (Caryocaraceae)”**. Os trabalhos foram instalados pelo(a) Orientador(a), Professor(a) Doutor(a) **Mariana Pires de Campos Telles (ICB/UFG)** com a participação dos demais membros da Banca Examinadora: Professor(a) Doutor(a) **Rafael Barbosa Pinto**, membro titular externo; Professor(a) Doutor(a) **Cintia Pelegrineti Targueta de Azevedo Brito**, membro titular externo. Durante a arguição os membros da banca fizeram várias correções, mas **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) **aprovada** pelos seus membros. Proclamados os resultados pelo(a) Professor(a) Doutor(a) **Mariana Pires de Campos Telles (ICB/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) **vinte e oito dias do mês de abril de dois mil e vinte e três**.
TÍTULO SUGERIDO PELA BANCA



Documento assinado eletronicamente por **Mariana Pires De Campos Telles, Professor do Magistério Superior**, em 29/04/2023, às 10:02, 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 **Rafael Barbosa Pinto, Usuário Externo**, em 02/05/2023, às 08:02, 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 **Cintia Pelegrineti Targueta de Azevedo Brito, Usuário Externo**, em 02/05/2023, às 10:07, 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 **3709692** e o código CRC **5D85FF9E**.

Dedico, a todos que não tiveram a oportunidade de chegar até onde eu cheguei, aos que acharam que não fossem capaz e principalmente aos que não tiveram oportunidade pela sua classe social.

“A única luta que se perde é a que se abandona”

AGRADECIMENTOS

Primeiramente, agradeço a Universidade Federal de Goiás e a Fundação de Amparo à Pesquisa do Estado de Goiás (FAPEG) (Nº processo no 202110267000674) por ter me fornecido a oportunidade e me concedido a bolsa nível de mestrado proporcionando-me dedicar e realizar meus estudos exclusivamente. Agradeço ao Programa de Pós-Graduação em Genética e Biologia Molecular ao Instituto de Ciências Biológicas da Universidade Federal de Goiás.

Agradeço ao Instituto Nacional de Ciência e Tecnologia em Ecologia, Evolução e Conservação da Biodiversidade (INCT – EECBio), dentro do grupo de trabalho de Genética e Genômica Evolutiva, fomentado 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), por ter fornecido a estrutura e suporte para o desenvolvimento da pesquisa.

Agradeço também à toda equipe e ao Laboratório de Genética & Biodiversidade (LGBio) do ICB/UFG por ter me acolhido de maneira extremamente gentil e receptível assim que entrei ao laboratório. Agradeço em especial minha orientadora Dra. Mariana Pires de Campos Telles e ao meu Coorientador Dr. Rhewter Nunes, por ter aceitado meu pedido de orientação, me acolhendo e instruindo cientificamente. Agradeço pela paciência, fizeram diferença me ajudando a seguir um novo passo na minha pós-graduação.

Agradeço todos os meus colegas de laboratório, e imensamente ao Leonado C. J. Corvalán e Marla Arianne Almeida por ajudar e compartilhar seus conhecimentos comigo sempre que necessário, a paciência de vocês foi crucial na minha aprendizagem e espero algum dia poder retribuir de alguma forma.

Também agradeço aos meus pais, minha mãe Cleuza Francisco da Silva, meu pai Nardeli Abadio Mendes, meu irmão Markley da Silva Mendes por me apoiar e fornecer, da forma que conseguem, a estrutura mesmo que simples para conseguir alcançar meus objetivos, a toda minha família e amigos.

“Por muito tempo na história, o anônimo foi uma
mulher”

(Virgínia wolff)

SUMÁRIO

1. INTRODUÇÃO	9
2. OBJETIVOS	14
2.3. Objetivo geral	14
2.3. Objetivos específicos	14
3. MATERIAIS E MÉTODOS.....	15
3.1. Extração e sequenciamento cloroplastidial de <i>Caryocar cuneatum</i>	15
3.2. Anotação do genoma cloroplastidial de <i>Caryocar cuneatum</i>	15
3.3. Análise comparativa estrutural e regiões repetitivas.....	16
3.4. Reconstrução filogenética	17
4. RESULTADOS.....	20
4.1. Anotação do genoma cloroplastidial de <i>Caryocar cuneatum</i>	20
4.2. Análise comparativa estrutural e regiões repetitivas de espécies de Malpighiales	22
5. DISCUSSÃO	33
6. CONCLUSÕES	36
7. REFERÊNCIA	37
8. ANEXOS	41

RESUMO

A família Caryocaraceae (1845) é pertencente à ordem Malpighiales, que apresenta 36 famílias, 716 gêneros e 16.065 espécies. Caryocaraceae é constituída por dois gêneros *Anthodiscus* G. Mey. com dez espécies e *Caryocar* L. com 16 espécies. O presente trabalho tem como objetivo realizar a montagem de novo do genoma cloroplastidial de *Caryocar cuneatum* Wittm. e realizar análises comparativas quanto à estrutura e a composição do genoma com outras espécies de Malpighiales. Para isso, foi sequenciado o DNA total de um indivíduo de *C. cuneatum* utilizando a plataforma *MiSeq* (Illumina), *paired-end* (2x300) com o kit *MiSeq* V3 600 ciclos. O plastoma foi montado utilizando o programa *NOVOPlasty* v3.2 e anotado utilizando o *CHLOROBX*. Para as análises comparativas, foram utilizados genomas cloroplastidiais de 8 espécies pertencentes a ordem de Malpighiales (*Caryocar brasiliense*, *Caryocar glabrum*, *Lophopyxis maingayi*, *Drypetes indica*, *Aspidopterys concava*, *Byrsonima crassifolia*, *Balanops balansae*, *Couepia ovalifolia*). O genoma cloroplastidial de *C. cuneatum* apresentou um tamanho de 165.767 pb, composto de uma região maior de cópia única de 83.968 pb, uma região menor de cópia única de 11.854 pb, separadas por duas regiões repetidas e invertidas de 34.973 pb. Apresenta 131 genes, 90 codificadores de proteínas, 33 de RNAs transportadores e 8 RNAs ribossomais. O gene *rpl32* não está presente nas espécies em *C. cuneatum* e nas outras duas espécies de *Caryocar*. Para as outras espécies, verificou-se variação no tamanho, estrutura e composição genômica. O gene *InfA* está pseudogenizado em todas as espécies analisadas e não foi encontrado em *A. concava*. Na análise de mutações sinônimas e não sinônimas o gene *rpl20* está sobre neutralidade e os outros genes estão em seleção negativa. Para a diversidade nucleotídica nos genomas de *Caryocar* foi possível identificar duas regiões de hotspots: *trnS-GCU - trnG-UCC* e *rbcL - atpB*, o qual podem ser utilizados como possíveis marcadores moleculares de DNA *barcoding*. A árvore filogenética obteve bons valores de suporte para os nós em Caryocaraceae validando a posição sistemática de *C. cuneatum* dentro da família e evidência sua relação com seus grupos irmãos *C. brasiliense* e *C. glabrum*. Em conclusão esse estudo fornece o primeiro genoma de *C. cuneatum* sequenciado, informação sobre a caracterização genômica e os conhecimentos aqui fornecidos podem ser utilizadas para novos estudos e desenvolvimentos de tecnológicas.

Palavras chaves: Genoma organelar. Plastoma. Comparação genômica.

ABSTRACT

The Caryocaraceae family (1845) belongs to the order Malpighiales, which has 36 families, 716 genera and 16,065 species. Caryocaraceae consists of two genera *Anthodiscus* G. Mey. with ten species and *Caryocar* L. with 16 species. The present work aims to carry out the de novo assembly of the chloroplastial genome of *Caryocar cuneatum* Wittm. and to carry out comparative analyzes regarding the structure and composition of the genome with other species of Malpighiales. For this, the total DNA of an individual of *C. cuneatum* was sequenced using the MiSeq platform (Illumina), paired-end (2x300) with the MiSeq V3 600 cycles kit. The plastome was assembled using the NOVOPlasty v3.2 program and annotated using the CHLOROBX. For comparative analyses, chloroplast genomes of 8 species belonging to the Malpighiales order (*Caryocar brasiliense*, *Caryocar glabrum*, *Lophopyxis maingayi*, *Drypetes indica*, *Aspidopterys concava*, *Byrsonima crassifolia*, *Balanops balansae*, *Couepia ovalifolia*) were used. The chloroplast genome of *C. cuneatum* had a size of 165,767 bp, composed of a single copy major region of 83,968 bp, a single copy minor region of 11,854 bp, separated by two inverted repeat regions of 34,973 bp. It has 131 genes, 90 protein coding, 33 transfer RNAs and 8 ribosomal RNAs. The *rpl32* gene is not present in species in *C. cuneatum* and in the other two species of *Caryocar*. The size of the plastid genome of the genus *Caryocar* presents itself a lot. For the other species, there was variation in size, structure and genomic composition. The *InfA* gene is pseudogenized in all analyzed species and was not found in *A. concava*. In the analysis of synonymous and non-synonymous mutations, the *rpl20* gene is under neutrality and the other genes are under negative selection. For the nucleotide diversity in *Caryocar* genomes, it was possible to identify two hotspot regions: *trnS-GCU - trnG-UCC* and *rbcL - atpB*, which can be used as possible molecular markers for DNA barcoding. The phylogenetic tree obtained good support values for nodes in Caryocaraceae, validating the systematic position of *C. cuneatum* within the family and evidence of its relationship with its sister groups *C. brasiliense* and *C. glabrum*. In conclusion, this study provides the first genome of *C. cuneatum* sequenced, information on genomic characterization and the knowledge provided here can be used for further studies and technological.

Keywords: Organellar genome. Plastome. Genomic comparison.

1. INTRODUÇÃO

O cloroplasto é a organela responsável pela fotossíntese nas plantas e como função secundária desempenha a síntese de aminoácidos, nucleotídeos, ácidos graxos, fitormônios, vitaminas e assimilação de enxofre e nitrogênio (BOBIK e BURCH-SMITH, 2015). O cloroplasto contém seu próprio DNA, denominado cpDNA ou genoma cloroplastidial (RODRÍGUEZ-EZPETETA et al., 2005; SATO, 2019). Geralmente, o material genético dessa organela em espermatófitas se dispõe em formato circular, com estrutura quadripartida, com duas cópias de repetições invertidas (IRs - *inverted repeats*), separadas pela região longa de cópia única (LSC- *large single copy*) e região curta de cópia única (SSC- *small single copy*) (JASEN; RUHLMAN, 2012; XIAO-MING et al., 2017). O número de genes do genoma do cloroplasto de angiospermas que realizam fotossíntese varia de 100 a 130 genes distintos, e o conteúdo de seus genomas são altamente conservados (PROVAN et al., 2001).

Algumas modificações estruturais podem ocorrer em determinadas linhagens de plantas, como rearranjos estruturais significativos, perdas de regiões repetidas invertidas ou famílias inteiras de genes. As estruturas mais aberrantes do genoma plastidial pertencem a plantas parasitárias que perderam genes conforme faziam transição independente para um organismo não fotossintetizantes (WICKE e NAUMANN, 2018).

Os genomas cloroplastidiais têm potencial utilização em estudos evolutivos e ajuda na resolução de relações filogenéticas de táxons, isso se deve ao fato desse genoma ser altamente conservado pela herança extranuclear materna. Além disso, possuem milhões de cópias em uma única célula vegetal. São importantes para obtenção de marcadores *barcoding*. Essa é uma técnica usada para identificar regiões do genoma de espécies conhecidas, de acordo com Dong et al. (2015) *ycfla* ou *ycflb* é a região do genoma plastidial mais variável e pode servir como DNA *barcoding* padrão para identificar plantas terrestres. O conhecimento entre as relações de espécies é necessário para entender padrões evolutivos de diversidade e suas características similares herdadas de um ancestral em comum (EMBRAPA, 2005; LOPEZ et al., 2016).

A família Caryocaraceae Voigt é pertencente à ordem Malpighiales. Esta ordem apresenta atualmente 36 famílias, 716 gêneros e 16.065 espécies. Caryocaraceae é constituída por dois gêneros *Anthodiscus* G. Mey. com dez espécies: *A. amazoniens* Gleason & ACSm., *A. chocoensis* Prance, *A. fragrans* Sleumer, *A. klugii* Standl, ex Prance, *A. mazarunensis* Gilly, *A. montanus* Gleason, *A. obovatus* Benth. ex Wittm., *A. peruanus* Baill., *A. pilosus* Duckc e *A.*

trifoliatus G. Mey; atualmente aceitas e *Caryocar* L. com 16 espécies atualmente aceitas: *C. amygdaliferum* Mutis ex Cav., *C. amygdaliforme* G. Don, *C. brasiliense* A.St.-Hil., *C. coriaceum* Wittm., *C. costaricense* Donn.Sm., *C. cuneatum* Wittm., *C. dentatum* Gleason, *C. edule* Casar. *C. glabrum* (Aubl.) Pers., *C. gracile* Wittm., *C. harlingii* Prance & Encam., *C. microcarpum* Ducke, *C. montanum* Prance, *C. nuciferum* L., *C. pallidum* ACSm. e *C. villosum* (Aubl.) Pers. (APG IV, 2016; PRANCE; SILVA, 1973).

A distribuição da família ocorre na região neotropical entre as Américas, desde a região da Costa Rica até a região do Paraguai (PRANCE e SILVA, 1973). O gênero *Caryocar* ocorre de maneira mais ampla do que *Anthodiscus*, nos países: Costa Rica, Paraguai, nas Guianas até o Sul do Brasil e no bioma Amazônico, onde é sua maior diversidade de ocorrências de espécies (PRANCE, 1990, 2014). Já o gênero *Anthodiscus* ocorre na região noroeste da América do Sul e apenas na região Norte e Oeste da Floresta Amazônica e uma pequena porção da Floresta Atlântica na Bahia (PRANCE e SILVA, 1973). No território brasileiro ocorrem os dois gêneros e 17 espécies, *Anthodiscus* G. Mey com cinco espécies e *Caryocar* L. com 12 espécies (MEDEIROS E AMORIM, 2015; ZAPPI, 2015). Dentre as espécies de Caryocaraceae, *C. brasiliense* ocorre em o maior número de diferentes ecorregiões principalmente em Savanas (PRANCE, 1973; NUNES et al., 2020).

A morfologia de Caryocaraceae é constituída por arvores e pode ocorrer como arbustos ou subarbustos (PRANCE E SILVA, 1973). As folhas são alternas ou opostas, em formato trifolioladas, pode aparecer com ou sem estípulas, na maioria das ocorrências com estípelas na base dos folíolos, o formato da nervação é braquidódroma e camptódroma, são os padrões usuais, embora algumas espécies de *Anthodiscus* sejam hipódromas (PRANCE e SILVA 1973, PRANCE e PIRANI, 2020).

Suas flores são projetadas em formato de inflorescência terminal em racemo possuindo flor terminal. Suas flores são grandes com simetria actinomórficas e hermafroditas. A corola da flor pode apresentar de quatro a cinco pétalas e pouco frequente acima dessas quantidades, e o cálice também segue o mesmo padrão. As pétalas são imbricadas se apresentando total ou parcialmente sobrepostas no botão floral. As flores têm de 55 a 750 estames e a anteras são biloculares. O ovário é rugoso e tem de quatro a seis carpelos em *Caryocar*, e de oito a 20 em *Anthodiscus*. Sua fruta é do tipo drupa com uma a quatro sementes em *Caryocar*, e de 8 a 20 em *Anthodiscus*. O mesocarpo é indecente e o endocarpo é duro e espinhoso. As sementes são reniformes com dois cotilédones e endosperma fino ou ausente (PRANCE E SILVA, 1973; PRANCE e PIRANI, 202).

No que diz respeito as informações filogenéticas, essa família não apresenta uma resolução bem clara de seus aspectos taxonômicos com outros clados de Malpighiales. Não há informações moleculares suficientes que mostram as relações entre alguns táxons dessa ordem, dessa maneira, os valores de suportes para nós não são consistentes para as filogenias moleculares dos clados de Crisobalanoideis, Malpighioides, Putranjivoids e Caryocaraceae formando uma politomia entre eles (XI et al., 2012).

Nome *Cayocar* possui uma etimologia do grego *caryon* e possui significado de núcleo ou noz fazendo referência ao centro do fruto e *kara* significa cabeça. *Caryon* e *kara* fazem referência ao fruto que possui um formato globoso. O nome vernacular pequi vem da língua tupi, o qual o prefixo *py* significa pele ou casca e *qui* significa espinho, fazendo referência aos espinhos do endocarpo ou caroço do fruto. (CARVALHO, 2009).

A morfologia de suas folhas é oposta, com duas folhas se inserindo no mesmo nó do caule em lados opostos. A inflorescência é terminal nascendo no final da haste quando ocorre sem flores é do tipo racemo ou do tipo botrioides quando ocorre com flores, seu pedúnculo é conspícuo e a raque pode ser alongada ou curta. As flores frequentemente são amarelas, podem ser brancas, rosadas ou vermelhas, o pedicelo é inserido no ápice da inflorescência, o cálice das flores é do tipo cupuliforme. As sépalas podem ocorrer a quantidade de cinco a seis, as pétalas também como nas sépalas, podem ocorrer de cinco a seis sendo unidas pela base. Os estames ocorrem entre 55 a 750 unidades, longos projetam-se para fora da corola ficando maiores que as pétalas, anteras podem ser do tipo basifixas ou dorsifixas. A fruta em forma de drupa, possui pirenos delimitados por endocarpo muricado a tuberculado ou espinhoso podendo ocorrer de quatro a seis unidades (PRANCE E SILVA, 1973; PRANCE e PIRANI, 202; REFLORA, 2023).

A distribuição desse gênero é neotropical, e no Brasil são reconhecidas 12 espécies: *C. brasiliense* Cambess., *C. coriaceum* Wittm., *C. cuneatum* Wittm., *C. dentatum* Gleason., *C. edule* Casar., *C. glabrum* (Aubl.) Pers., *C. gracile* Wittm., *C. microcarpum* Ducke., *C. montanum* Prance., *C. nuciferum* L., *C. pallidum* A.C.Sm., *C. villosum* (Aubl.) Pers. (MEDERIOS E AMORIM, 2015).

Esse gênero é considerado de grande potencial econômico, o qual praticamente todas as partes da planta tem algum uso comercial, a castanha, casca, folhas, madeira principalmente o seu fruto que possui altos teores de óleo, vitaminas e minerais neles encontrados (ALMEIDA e SILVA, 1994; POZO, 1997; OLIVEIRA et al., 2008). O fruto contém uma polpa que fica aderida ao caroço, normalmente contém espinhos, ele é bastante utilizado na culinária local. É consumido fresco ou em conserva, geleia e em pratos típicos, como frango com pequi. Também

é utilizado na produção de óleos e fabricação de licores. Sua semente é extraída da castanha usada como alimento ou também para a produção de óleo medicinal (NUNES et al., 2020).

Além de ser utilizado na alimentação também é usado pelas populações tradicionais locais onde ele ocorre, o qual os saberes populares que são passados de geração á geração e culturais na medicina natural, o óleo é utilizado como anti-inflamatório (BEZERRA; BARROS; COELHO, 2015) para tratamento de bronquites, gripes, resfriados e controle de tumores, tratamentos oftalmológicos relacionados à deficiência de vitamina A, por possuir em sua composição vitamina A e C, tiamina e sais minerais (CARVALHO, PEREIRA, ARAUJO, 2015). Além disso, estudos demonstram sua viabilidade, de acordo com Pereira et al. (2020) a polpa e o óleo de *C. coriaceum* tem abundância em ácidos graxos oleico e palmítico. Sendo utilizados para o controle de bactérias multirresistentes. Em *C. brasiliense* também apresenta eficiência quando usado junto com o fluido magnético com dextrano (DexMF) no tratamento do tumor sólido de Ehrlich clínico, agindo contra as células tumorais (MIRANDA-VILELA et al., 2014). E também apresenta atividade anti-inflamatória contra os efeitos da atividade física intensa e atua como modulador do colesterol total, principalmente em homens (MIRANDA-VILELA et al., 2009).

Outro fator importante é sua utilização como potencial econômico na compra e venda desse fruto, uma importante participação de renda para as populações rurais do cerrado, esses são vendidos em feiras livres, comercializados em festas típicas e por vendedores ambulantes. No estado de Goiás também é vendido na Central de Abastecimento de Goiás – CEASA, que distribui o alimento para outros vendedores dentro do estado. Esse fruto é obtido de plantio e extrativismo (BRAIT, 2008; SILVA et al., 2021).

A espécie *Caryocar cuneatum* Wittm. é endêmica do Brasil, e seu nome vernacular é também chamado de pequizeiro ou pequi assim como outras espécies de *Caryocar*, como *C. Brasiliense*, e muitas vezes confundido entre comunidades que fazem uso delas. A morfologia de suas folhas contem o ápice dos folíolos apiculados, margem serradas ou crenadas. Dispõe suas flores em inflorescências em forma de racemo com pedúnculos alongados, a cor de suas pétalas podem ser róseas e brancas, com formato oblongas. Produz drupa elipsóide como fruto, contendo dois locus e duas sementes, mas pode ocorrer três ou mais, a textura do seu fruto é liso. As sementes estão em forma de pineros com endocarpo espinhosos. Sua forma de vida se desenvolve como árvore, pode crescer até 15 metros de altura com 35 metros de diâmetro e seu substrato é terrícola. Sua distribuição geográfica ocorre pelo Norte, Nordeste e Centro-oeste, nos estados do Pará, Tocantins, Bahia, Maranhão, Piauí e Goiás (PRANCE E SILVA, 1973; PRANCE e PIRANI, 202; REFLORA, 2013). É relatado ocorrendo em bioma

da Caatinga do Nordeste brasileiro e nas ecorregiões de Desertos e Arbustos Xericos, mas a maioria das ocorrências são principalmente no Cerrado e próximo ao bioma Caatinga (NUNES, 2020; PRANCE, 2014).

De acordo com Nunes et al. (2020), dentro da família Caryocaraceae as principais espécies representadas na literatura são do gênero *Caryocar*, o qual *C. brasiliense* lidera a quantidade de estudos, desde anatomia, morfologia, caracterização físico-química até genética. Para a área da genética, os estudos se restringem á diversidade genética, como informações da caracterização cariotípica obtidas para *C. brasiliense*, *C. microcarpum* e *C. villosum* (EHRENDORF et al., 1984). Estudos para desenvolvimento de marcadores microssatélites (SSR) transferidos para *C. coriaceum*, *C. edule*, *C. glabrum*, *C. pallidum* e *C. villosum* (COLLEVATTI et al., 1999).

Para estudos genômicos, existe o sequenciamento genômico cloroplastidial apenas para duas espécies: *C. brasiliense* (NUNES et al., 2020) e *C. glabrum*, de acordo com o banco de dados genômicos NCBI (*National Center for Biotechnology Information*). Atualmente, *C. brasiliense* se destaca apresentando grande parte de análises dessa área, quando comparados com outras espécies desse gênero, não existindo mais pesquisas genômicas para as demais espécies de *Caryocar*. Dessa maneira, este estudo se propõe a disponibilizar *de novo* o genoma cloroplastidial de *C. cuneatum* e realizar análises estruturais e comparativas com as outras espécies do gênero, com intuito de fornecer recursos genômicos e contribuir com o conhecimento filogenético do clado.

2. OBJETIVOS

2.3. Objetivo geral

Realizar uma montagem *de novo* do genoma cloroplastidial de *Caryocar cuneatum* Wittm. (Caryocaraceae) e realizar análises comparativas com genomas cloroplastidiais de espécies da ordem Malpighiales.

2.3. Objetivos específicos

1. Sequenciar e montar o primeiro genoma cloroplastial de *Caryocar cuneatum*;
2. Realizar análises estruturais e comparativas entre os genomas de cloroplasto de *Caryocar cuneatum* e as espécies pertencentes aos clados dos Crisobalanoides, Malpighioids, Putranjivoids e Caryocaraceae;
3. Realizar análises das mutações não sinônimas e sinônimas entre genes homólogos do genoma cloroplastidial das espécies pertencentes a Malpighiales;
4. Identificar regiões de hotspots através da análise de diversidade nucleotídicas nos genomas cloroplastidiais;
5. Validar a posição taxonomica de *Caryocar cuneatum* através da construção de uma filogenia molecular utilizando espécies de Malpighiales.

3. MATERIAIS E MÉTODOS

3.1. Extração e sequenciamento cloroplastidial de *Caryocar cuneatum*

Para a obtenção do genoma de *Caryocar cuneatum* foram coletadas folhas de uma árvore adulta na região da Chapada dos Veadeiros, na cidade de Teresina de Goiás, Goiás (SISGEN - AAD5076). Inicialmente, foi realizada a extração de DNA utilizando o método CTAB 2% (DOYLE e DOYLE, 1987). Em seguida, o DNA foi quantificado utilizando eletroforese em gel de agarose de 1% e o *Qubit dsDNA High Sensitivity Assay Kit* (*ThermoFischer Scientific*). O preparo da biblioteca genômica foi construído com o protocolo *Nextera DNA Flex* (Illumina) seguindo os seguintes passos, tagmentação do DNA, limpeza pós-tagmentação, amplificação do DNA tagmentado e limpeza final. A avaliação da qualidade da biblioteca foi realizada no *Bioanalyzer 2100* (*Agilent*). Em seguida, biblioteca foi quantificada com o *Qubit dsDNA High Sensitivity Assay Kit* e em seguida diluída para a concentração de 15 pM. O sequenciamento foi realizado na plataforma MiSeq (Illumina) no modo *paired-end* (2x300), sendo 300 pb para *Forward* e 300 pb para *Reverse*, utilizando o kit MiSeq V3 600 ciclos (Illumina). Todas as etapas foram conduzidas no Laboratório de Genética & Biodiversidade – LGBio, da Universidade Federal de Goiás, em Goiânia - Goiás – Brasil.

3.2. Anotação do genoma cloroplastidial de *Caryocar cuneatum*

A montagem do genoma cloroplastidial de *C. cuneatum* foi realizado no programa *NOVOplasty v2.2* (DIERCKXSENS et al., 2017). A anotação do genoma cloroplastidial de *C. cuneatum* foi feita utilizando o programa *webserver software GeSeq* (TILLICH et al., 2017). A predição de genes codificadores de proteínas e RNAs ribossômicos (rRNAs) foi realizada nos programas *BLAT search* (KENT, 2002) e a predição de RNAs transportador (tRNAs) utilizando *tRNAscan-SE v2.0.7*. Os genes foram verificados e os que estavam sobrepostos foram removidos e anotados manualmente, no programa *Ugene* (OKONECHNIKOV et al., 2012). O mapa gráfico do genoma cloroplastidial foi construído no *Draw Organelle Genome Maps* (OGDRAW) (LOHSE et al., 2013).

3.3. Análise comparativa estrutural e regiões repetitivas

Para análise comparativa, o genoma cloroplastidial de *C. cuneatum* foi comparado ao de 8 espécies pertencentes à ordem Malpighiales (Tabela 1). Os genomas cloroplastidiais das 8 espécies foram obtidos no banco de dados *Genome (RefSeq)* do NCBI para serem anotados. O programa *REPuter* foi utilizado para a identificação das estruturas de repetições complexas. Foram feitas análises para quatro tipos de repetições: *Forward*, *Reverse*, *Palindromic* e *Complement* (KURTZ et al., 2001). Para tanto, foi estabelecido o tamanho mínimo da repetição de 30 pb e uma distância de *Hamming* de 3 pb. Para a identificação e caracterização das regiões microssatélites foi utilizado o programa *Web MISA* (BIER et al., 2017) e os números de repetições mínimas foram definidas para 10, 5, 3, 3, 3 para as repetições mononucleotídicas, dinucleotídicas, trinucleotídicas, tetranucleotídicas, pentanucleotídicas e hexanucleotídicas.

Para identificação do tamanho das quatro regiões dos genomas cloroplastidiais, as IRs, LSC, SSC e os genes localizados em suas extremidades foi utilizado o *Geneious prime* (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) de alinhamento múltiplo.

Tabela1: Espécies utilizadas para análises comparativas e seus respectivos números de referência.

Espécie	ID	Clado	Família
<i>Caryocar cuneatum</i>	-	Caryocaraceae	Caryocaraceae
<i>Caryocar brasiliense</i>	MK726375 NC_061179.1/M	Caryocaraceae	Caryocaraceae
<i>Caryocar glabrum</i>	N116504.1 NC_050684.1/M	Caryocaraceae	Caryocaraceae
<i>Lophopyxis maingayi</i>	N504795.0 NC_052002.1/M	Putranjivoid	Lophopyxidaceae
<i>Drypetes indica</i>	N504791.1 NC_060542.1/OL	Putranjivoid	Putranjivaceae
<i>Aspidopterys concava</i>	471043.1 NC_037192.1/MF	Malpighioid	Malpighiaceae
<i>Byrsonima crassifolia</i>	359248.1	Malpighioid	Malpighiaceae

	NC_052006.1/M		
<i>Balanops balansae</i>	N504796.1	Chrysobalanoid	Balanopaceae
	NC_030549.1/KX		
<i>Couepia ovalifolia</i>	180055.1	Chrysobalanoid	Chrysobalanaceae

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

Para as análises de diversidade nucleotídica foram utilizadas três espécies do gênero *Caryocar*, sendo elas: *C. brasiliense*, *C. glabrum* e *C. cuneatum*. Inicialmente, os genomas foram alinhados utilizando o programa *MAFFT* (KATOH e STANDLEY, 2013). A estimativa de diversidade nucleotídica (π) foi realizada utilizando o programa *DnaSP v6* definindo janelas de 600 pb e passos de 200 pb (ROZAS et al., 2017).

Para as relações entre mutações não sinônimas e sinônimas (ka/ks). O cálculo foi realizado utilizando 79 regiões codificadoras de proteínas (CDS - *protein-coding DNA sequence*) dos genomas cloroplastidiais das 9 espécies utilizadas para as análises comparativas, juntamente com o alinhamento no programa *MAFFT*. Foi considerado o valor superior a 1 indicando processo de seleção positiva, inferior a 1 indicando seleção negativa e igual a 1 indicando seleção neutra, ou seja, ausência de seleção de genes (NEI e KUMAR, 2000).

3.4. Reconstrução filogenética

Para a reconstrução filogenética dos genomas cloroplastidiais da ordem de Malpighiales foram selecionadas 34 espécies (Tabela 2), dessas, três pertencentes à família Caryocaraceae, uma pertencente à família Lophopyxidaceae, duas à família Putranjivaceae, quatro pertencente à família Malpighiaceae, uma à família Balanopaceae e 18 a Chrysobalanaceae. Além disso, a espécie *Vitis vinífera* foi utilizada como *outgroup* (grupo externo) para a construção da árvore filogenética.

As CDS dessas espécies foram obtidas no NCBI e alinhadas utilizando o *MAFFT version 7* (KATOH e STANDLEY, 2013), posteriormente concatenadas utilizando um *Script* no *Castfasta2phym1*. As regiões filogeneticamente não informativas foram retiradas utilizando o *Webserver Gblocks* (TALAVERA e 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 47.622 pares de bases na construção da árvore filogenética e o melhor modelo foi selecionado utilizando o *ModelFinder* (KALYAANAMOORTHY et al., 2017). Com isso, foi selecionado o modelo TVMF+R2 e os valores de suporte de nó na árvore foram avaliados utilizando o método de *bootstrapping* com 1000 réplicas.

Tabela 2: Tabela das 34 espécies utilizada para análise filogenética e um grupo externo.

Espécie	ID	Clado	Família
<i>Caryocar cuneatum</i>	-	Caryocaraceae	Caryocaraceae
<i>Caryocar glabrum</i>	NC_061179.1/MN116504.1	Caryocaraceae	Caryocaraceae
<i>Caryocar brasiliense</i>	MK726375	Caryocaraceae	Caryocaraceae
<i>Lophopyxis maingayi</i>	NC_050684.1/MN504795.0	Putranjivoid	lophopyxidaceae
<i>Drypetes indica</i>	NC_052002.1/MN504791.1	Putranjivoid	putranjivaceae
<i>Sibangea similis</i>	NC_052005.1/MN504794.1	Putranjivoid	putranjivaceae
<i>Aspidopterys concava</i>	NC_060542.1/OL471043.1	Malpighioid	Malpighiaceae
<i>Banisteriopsis caapi</i>	NC_037945.1/MH041648.1	Malpighioid	Malpighiaceae
<i>Bunchosia argentea</i>	NC_041491.1/MK390344.1	Malpighioid	Malpighiaceae
<i>Byrsonima crassifolia</i>	NC_037192.1/MF359248.1	Malpighioid	Malpighiaceae
<i>Balanops balansae</i>	NC_052006.1/MN504796.1	Chrysobalanoid	Balanopaceae
<i>Acioa guianensis</i>	NC_030534.1/KX158870.1	Chrysobalanoid	Chrysobalanaceae
<i>Afrolicania elaeosperma</i>	NC_030544.1/KX180050.1	Chrysobalanoid	Chrysobalanaceae
<i>Angelesia splendens</i>	NC_030545.1/KX180051.1	Chrysobalanoid	Chrysobalanaceae
<i>Atuna racemosa</i>	NC_030546.1/KX180052.1	Chrysobalanoid	Chrysobalanaceae
<i>Chrysobalanus icaco</i>	NC_024061.1/KJ414480.1	Chrysobalanoid	Chrysobalanaceae
<i>Couepia ovalifolia</i>	NC_030549.1/KX180055.1	Chrysobalanoid	Chrysobalanaceae
<i>Dactyladenia bellayana</i>	NC_030555.1/KX180062.1	Chrysobalanoid	Chrysobalanaceae
<i>Exellodendron barbatum</i>	NC_030558.1/KX180065.1	Chrysobalanoid	Chrysobalanaceae
<i>Gaulettia elata</i>	NC_030559.1/KX180066.1	Chrysobalanoid	Chrysobalanaceae
<i>Geobalanus oblongifolius</i>	NC_030571.1/KX180079.1	Chrysobalanoid	Chrysobalanaceae
<i>Grangeria borbonica</i>	NC_030560.1/KX180067.1	Chrysobalanoid	Chrysobalanaceae
<i>Hirtella macrosepala</i>	NC_030561.1/KX180068.1	Chrysobalanoid	Chrysobalanaceae
<i>Hunga gerontogea</i>	NC_030564.1/KX180072.1	Chrysobalanoid	Chrysobalanaceae
<i>Hymenopus glabriflorus</i>	NC_030567.1/KX180075.1	Chrysobalanoid	Chrysobalanaceae
<i>Kostermanthus robustus</i>	NC_030565.1/KX180073.1	Chrysobalanoid	Chrysobalanaceae

<i>Leptobalanus sprucei</i>	NC_024065.1/KJ414484.1	Chrysobalanoid	Chrysobalanaceae
<i>Licania alba</i>	NC_024064.1/KJ414483.1	Chrysobalanoid	Chrysobalanaceae
<i>Magnistipula butayei</i>	NC_030576.1/KX180084.1	Chrysobalanoid	Chrysobalanaceae
<i>Maranthes gabunensis</i>	NC_030577.1/KX180085.1	Chrysobalanoid	Chrysobalanaceae
<i>Moquilea minutiflora</i>	NC_030573.1/KX180081.1	Chrysobalanoid	Chrysobalanaceae
<i>Neocarya macrophylla</i>	NC_030580.1/KX180088.1	Chrysobalanoid	Chrysobalanaceae
<i>Parastemon urophyllus</i>	NC_030517.1/KX180089.1	Chrysobalanoid	Chrysobalanaceae
<i>Parinari campestris</i>	NC_024067.1/KJ414486.1	Chrysobalanoid	Chrysobalanaceae
<i>Vitis vinifera</i>	NC_007957.1		Vitaceae

4. RESULTADOS

4.1. Anotação do genoma cloroplastidial de *Caryocar cuneatum*

O genoma cloroplastidial completo de *C. cuneatum* apresenta o tamanho de 165.767 pb, apresentando as quatro regiões esperadas para plantas terrestres. Uma região maior de cópia única (LSC) de 83.968 pb, uma região menor de cópia única (SSC) de 11.854 pb, separadas por duas regiões repetidas e invertidas (IRa e IRb - *inverted repeat*) de tamanho 34.973 pb. Dessa forma, apresenta uma estrutura quadripartida em formato circular (Figura 1).

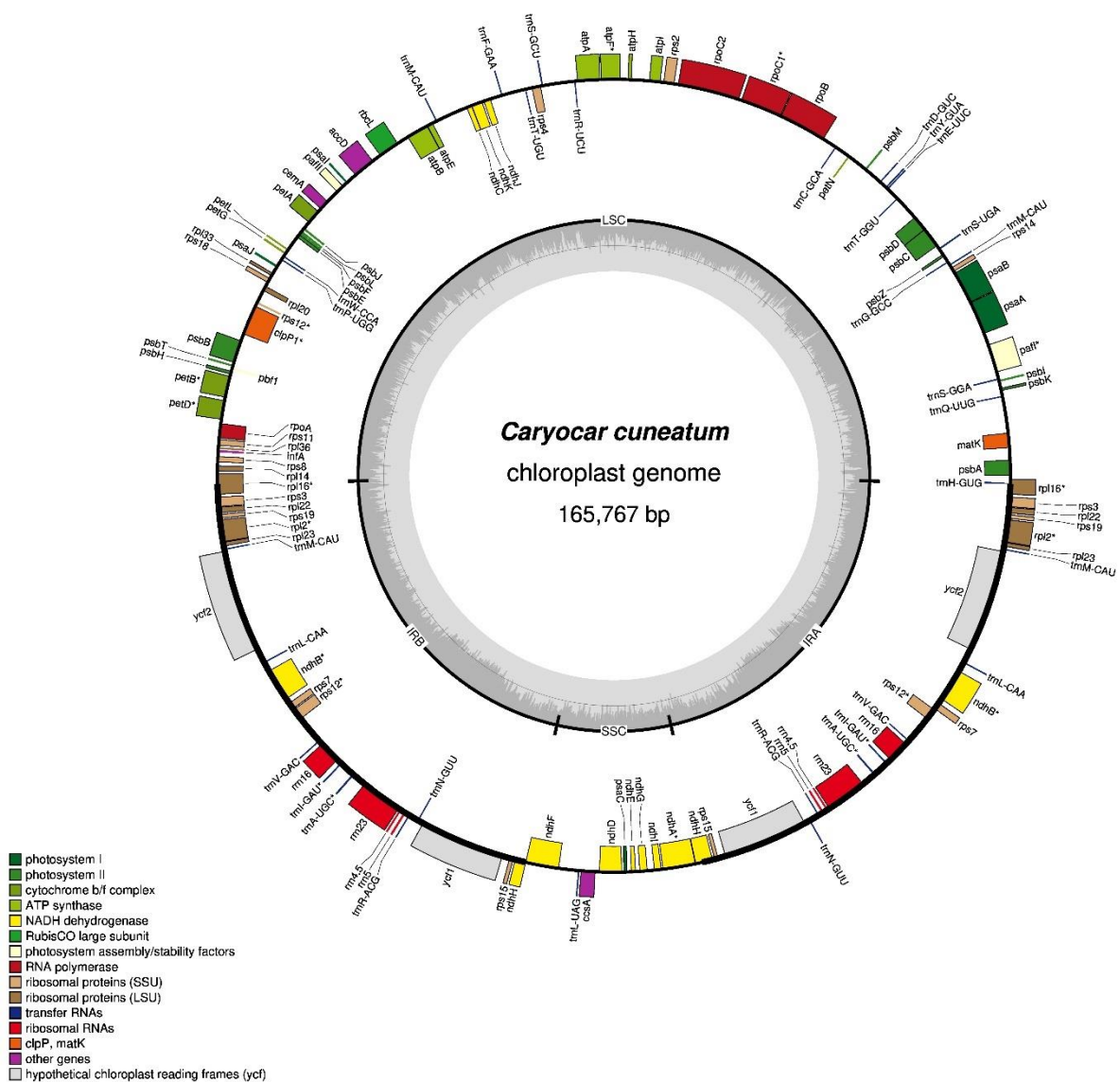


Figura 1. Mapa genômico circular do cloroplasto de *Caryocar cuneatum*. O círculo externo representa os genes e estão coloridos de acordo com seu grupo funcional. O círculo interno demarca as regiões invertidas repetidas (IRA e IRb - *inverted repeat regions*), região maior de cópia única (LSC - *large single copy*), região menor de cópia única (SSC - *small single copy*), e também mostram o conteúdo GC em barras. * genes que possuem íntrons.

Foram anotados 131 genes, sendo esses, 90 codificadores de proteínas, 33 de RNAs transportadores (tRNAs) e 8 RNAs ribossômicos (rRNAs). Considerando os genes que codificam proteínas, 13 apresentaram íntrons. Desses, apenas *ycf3* e *clpP1* apresentam dois íntrons e os demais apresentam um único íntron. Dos genes codificadores de RNAs, 33 são codificadores de tRNAs. Os genes *trnM-CAU*, *trnL-CAA*, *trnI-GAU*, *trnA-UGC*, *trnR-ACG*, *trnN-GUU*, *trnV-GAC* apresentam cópias. Os genes que apresentam apenas um íntron são *trnL-GAU* e *trnA-UGC*. Foram identificados 8 genes codificadores de rRNAs (Tabela 4). O conteúdo de Guanina + Citosina (GC%) é de 36,70% em todo o genoma.

Tabela 3: Categorização dos genes em *C. cuneatum*.

Categoria	Grupos de genes	Genes
Auto-replicação	Subunidade ribossômica maior	<i>rpl2</i> **, <i>rpl14</i> , <i>rpl16</i> ¹ , <i>rpl33</i> , <i>rpl20</i> , <i>rpl23</i> **, <i>rpl22</i> **, <i>rpl36</i> .
	Subunidade ribossômica menor	<i>rps2</i> , <i>rps3</i> , <i>rps4</i> , <i>rps7</i> , <i>rps8</i> , <i>rps11</i> , <i>rps15</i> **, <i>rps12</i> *** ¹ , <i>rps14</i> , <i>rps18</i> , <i>rps19</i> ***.
	RNA polimerase DNA dependente	<i>rpoA</i> , <i>rpoB</i> , <i>rpoC1</i> ¹ , <i>rpoC2</i> .
	Genes de RNAs ribossômicos	<i>rrn16</i> **, <i>rrn23</i> **, <i>rrn4.5</i> **, <i>rrn5</i> **.
	Genes de RNAs transportadores	<i>trnA-UGC</i> ** ¹ , <i>trnH-GUG</i> , <i>trnQ-UGG</i> , <i>trnS-GGA</i> , <i>trnM-CAU</i> ****, <i>trnG-GCC</i> , <i>trnS-UGA</i> , <i>trnT-GGU</i> , <i>trnE-UUC</i> , <i>trny-GUA</i> , <i>trnD-GUC</i> , <i>trnC-GCA</i> , <i>trnR-UCU</i> , <i>trnS-GCU</i> , <i>trnT-UGU</i> , <i>trnF-GAA</i> , <i>trnW-CCA</i> , <i>trnP-UGG</i> , <i>trnL-CAA</i> ** ¹ , <i>trnV-GAC</i> ** ¹ , <i>trnI-GAU</i> ** ¹ , <i>trnR-ACG</i> ** ¹ , <i>trnN-GUU</i> ** ¹ , <i>trnL-UAG</i> .
Fotossíntese	Fotossistema 1	<i>psaA</i> , <i>psaB</i> , <i>psaC</i> , <i>psaI</i> , <i>psaJ</i> .
	Fotossistema 2	<i>psbA</i> , <i>psbB</i> , <i>psbC</i> , <i>psbD</i> , <i>psbE</i> , <i>psbF</i> , <i>psbH</i> , <i>psbI</i> , <i>psbJ</i> , <i>psbK</i> , <i>psbL</i> , <i>psbM</i> , <i>psbN</i> , <i>psbT</i> , <i>psbZ</i> .
NADH desidrogenase	NADH desidrogenase	<i>ndhA</i> ** ¹ , <i>ndhB</i> ** ¹ , <i>ndhC</i> , <i>ndhD</i> , <i>ndhF</i> , <i>ndhE</i> , <i>ndhG</i> , <i>ndhH</i> ** ¹ , <i>ndhJ</i> , <i>ndhI</i> , <i>ndhK</i> .

	Complexo citocromo b/f	<i>petA, petB¹, petD¹, petG, petL, petN.</i>
	ATP sintase	<i>atpA, atpB, atpE, atpF¹, atpH atpI.</i>
Outros genes	Subunidade grande da rubisco	<i>rbcL</i>
	Maturase K	<i>matK</i>
	Proteína de membrana	<i>cemA</i>
	Subunidade de acetil-CoAcarboxilase	<i>accB</i>
	Gene de síntese citocromo tipo C	<i>ccsA</i>
	Protease	<i>clpP1²</i>
	Quadros de leitura abertos para hipotética conservação de cloroplasto	<i>ycf1^{**}, ycf2^{**}, ycf3^{**}, ycf4^{**2}</i>
	Pseudogenes	<i>infA</i>

Nota: **: genes duplicados, ***: genes triplicados, ****: genes quadruplicados, ¹ genes com um íntron, ² genes com dois íntrons.

4.2. Análise comparativa estrutural e regiões repetitivas de espécies de Malpighiales

As análises comparativas entre as nove espécies da ordem Malpighiales demonstram o menor tamanho genômico é de *Lophopyxis maingayi* (119.741 pb) e *Drypetes indica* (120.596 pb). Esse encurtamento se deve ao fato dessas espécies não possuírem IRs (JIN et al., 2020) diminuindo seu tamanho e também a quantidade de genes (Tabela 4). Para o tamanho da LSC de *Balanops balansae* apresenta o maior tamanho (88.825 pb) aproximado de *Couepia ovalifolia* (88.799 pb) e *Balanops balansae* (88.825 pb). O maior SSC foi de *Aspidopterys concava com* 31.539 pb e os menores são de *C. brasiliense* com 11.582 pb. No total, *C. brasiliense* apresenta o maior tamanho genômico.

Caryocar cuneatum apresenta o genoma semelhante ao *C. brasiliense* e próximo de *C. glabrum*. O tamanho das LSC varia de *C. cuneatum* (83.968 pb), *C. brasiliense* (84.135 pb) e para *C. glabrum* (83.900 pb). O tamanho das IRs contém uma pequena variação, entre *C. cuneatum*, *C. brasiliense* e *C. glabrum*. Nas SSC *C. brasiliense* e *C. glabrum* apresentam tamanho de 11.854 pb e para *C. cuneatum* existe uma pequena diferença entre os tamanhos dessa região.

Tabela 4: Descrição das nove espécies de genomas cloroplastidiais da ordem Malpighiales utilizados na análise comparativa.

Espécies	ID	GS	LSC	IR	SSC	GC	GC	GC	GC	Prote in	tRN A	rRN A	Gene
						(%) total	(%) IR	(%) SSC	(%) LSC				
<i>Caryocar cuneatum</i>	-	165.76 7	83.968	34.973	11.854	36,70	39,60	31,50	35,50	90	33	8	131
<i>Caryocar brasiliense</i>	MK726 375	165.79 3	84.135	34.903	11.582	36,70	39,60	31,50	35,50	90	33	8	131
<i>Caryocar glabrum</i>	NC_06 1179.1	165.69 0	83.900	34.969	11.852	36,70	39,60	31,50	35,50	90	33	8	131
<i>Lophopyxis maingayi</i>	NC_05 0684.1	119.74 1	-	-	-	36,0	-	-	-	78	26	4	107
<i>Drypetes indica</i>	NC_05 2002.1	120.59 6	-	-	-	36,0	-	-	-	78	27	4	109
<i>Aspidopterys concava</i>	NC_06 0542.1	160.44 1	88.737	19.962	31.539	36,50	40,90	37,40	34,30	86	31	8	125
<i>Byrsonima crassifolia</i>	NC_03 7192.1	160.21 2	88.448	26.975	17.814	36,80	42,40	30,80	34,50	87	33	8	128
<i>Balanops balansae</i>	NC_05 2006.1	160.93 0	88.825	26.748	18.609	37,30	42,70	31,40	35,20	90	33	8	129
<i>Couepia ovalifolia</i>	NC_03 0549.1	161.92 6	88.799	26.875	19.377	36,30	42,50	30	33,90	87	33	8	126

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

O número total de genes para as três espécies de *Caryocar* foram iguais e apresenta 131 genes. Além disso, a quantidade total de rRNAs para todas as espécies também foram iguais, com 4 genes duplicados (*rrn16*, *rrn23*, *rrn 4.5 + rrn5* e *rrn23*). Os genes codificadores de tRNAs também compõem a mesma quantidade com 33 genes e a mesma quantidade de íntrons com 19.

Considerando o conteúdo GC total para as espécies foi observado uma pequena variação, as espécies de *Caryocar* possuem o mesmo valor de 36,7% e das demais regiões do genoma o conteúdo GC também são iguais. Para as outras espécies o conteúdo total varia para *Byrsonima crassifolia* (36,80%) com a maior porcentagem *L. maingayi*. (30,0%). Para as demais regiões, as IRs possuem variação entre a espécie *B. balansae* (42,70%) e a menor *A. concava* (40,90%), SSC há variação considerável, a maior porcentagem para *A. concava* (37,40%) e menor para *C. ovalifolia* (30%), LSC com maior para *B. crassifolia* (34,30%) e menor para *C. ovalifolia* (33,90%), assim como mostra a tabela 4.

Para análise dos genes que flanqueiam as regiões de transição entre as IRs dos genomas cloroplastidiais foram identificados padrões de identificação nos genes que flanqueiam as regiões de transição. Tais padrões foram chamados de *Standard* evidenciam grupos de genes em determinada ordem e região e contém a finalidade de identificar semelhança entre eles (Figura 2). Na região LSC-IRb, foram identificados três padrões, *Standard 1*, flanqueado pelos genes *rpl14*, *rpl16*, e *rps3* em *C. brasiliense* e *C. cuneatum*, *Standard 2* flanqueado pelos genes *psbA*, *trnH-GUG*, *rpl16* e *rps3* em *C. glabrum*, *Standard 3* flanqueado pelos genes *rps3*, *rpl22*, *rps19* e *rpl2* *A. concava* e *Standard 4* flanqueados pelos genes *rps3*, *rps22*, *rps19* e *rpl12* pelas demais espécies.

Na região IRa e LSC foi identificado cinco padrões, *Standard 1* flanqueado pelos genes *rps3*, *rpl16* e *psbA* em *C. brasiliense* e *C. cuneatum*. *Standard 2* flanqueado pelos genes *rps3* e *rpl16* em *C. glabrum*. *Standard 3* flanqueado pelos genes *rpl2*, *rps19* - fragmento e *trH-GUG* em *A. concava* e *B. balansae*. *Standard 4* flanqueado pelos genes *rps19* e *trH-GUG* em *B. Crassifolia*. *Standard 5* flanqueado pelos genes *rpl2*, *rps19* em *C. ovalifolia*. Na região entre IRb e SSC foram identificados três padrões, *Standard 1* flanqueado pelos genes *ndhH* e *ndhF* para as três espécies de *Caryocar* e *Standard 2* flanqueado pelo gene *rrn23* em *A. concava*. *Standard 3* flanqueado pelos genes *ycf1* e *ndhF* em *B. Crassifolia*, *B. balansae*, *C. ovalifolia*. E na região entre SSC e IRa foram identificados três padrões, *Standard 1* flanqueado pelos genes *ndhA*, *ndhH* e *rps15* nas três espécies de *Caryocar*. *Standard 2* flanqueado pelo gene *rrn23* - fragmento em *A. concava* e *Standard 3*, flanqueado pelo gene *ycf1* em *B. crassifolia*, *B. balansae*, *C. ovalifolia* (Figura 2).

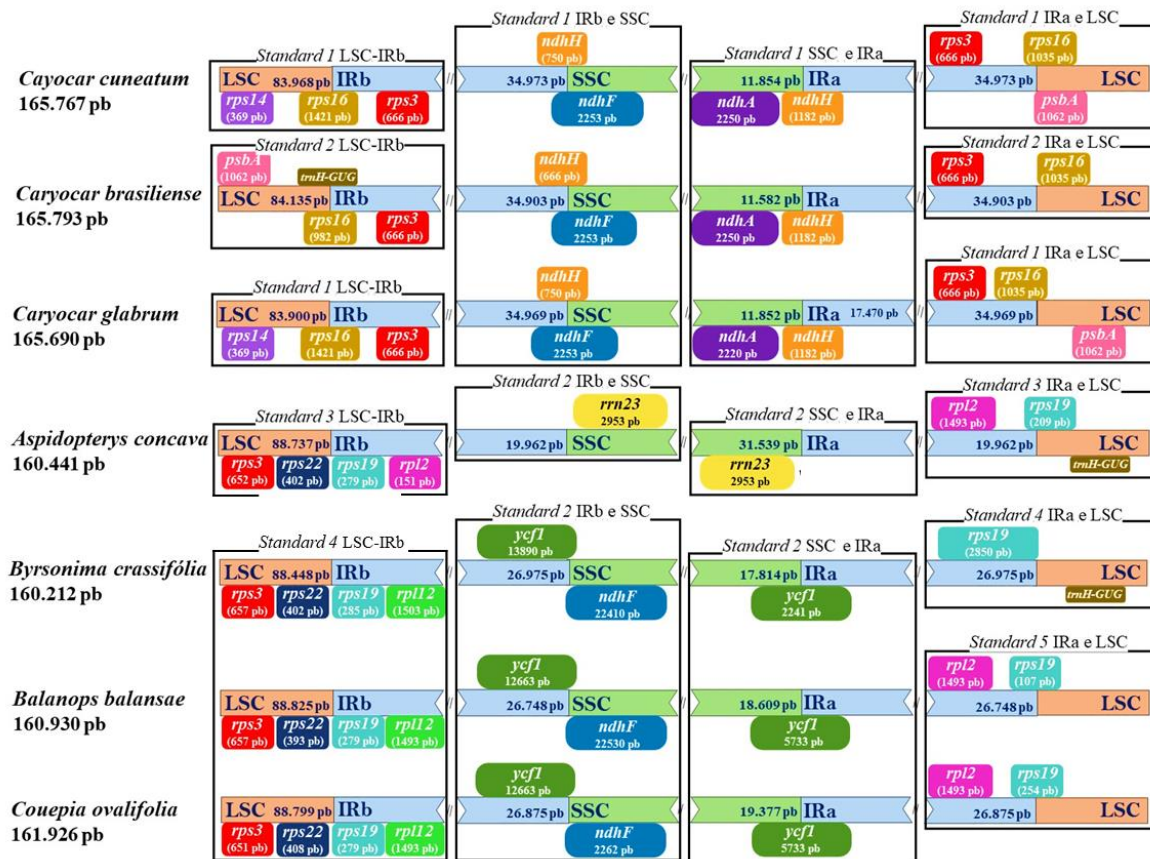


Figura 2. Comparações entre os genes que flanqueiam as quatro regiões dos genomas cloroplastidiais de 7 espécies utilizadas em análise comparativa

Nota: Regiões invertida repetida (IRa e IRb - inverted repeat regions) em azul, região longa de cópia única (LSC - large single copy) em salmão e região curta de cópia única (SSC - small single copy) em verde.

Com relação as repetições microssatélites na espécie *C. cuneatum* foram encontrados 85 microssatélites em seu genoma. Dentro de *Caryocar*, na espécie *C. brasiliense* foram encontrados 86 microssatélites para *C. glabrum* foram encontrados 88 microssatélites. Para as outras espécies, houve grande variação no valor total quando comparadas com *Caryocar*, 35 para *L. maingayi*, 64 para *D. indica*, 81 para *A. concava*, 49 para *B. crassifolia*, 69 para *B. balansae*, 66 para *C. ovalifolia*. As espécies *L. maingayi*, *A. concava*, *B. crassifolia*, *B. balansae* não possuem motivo de microssatélites Hexanucleotídico (Figura 3 e Tabela 5).

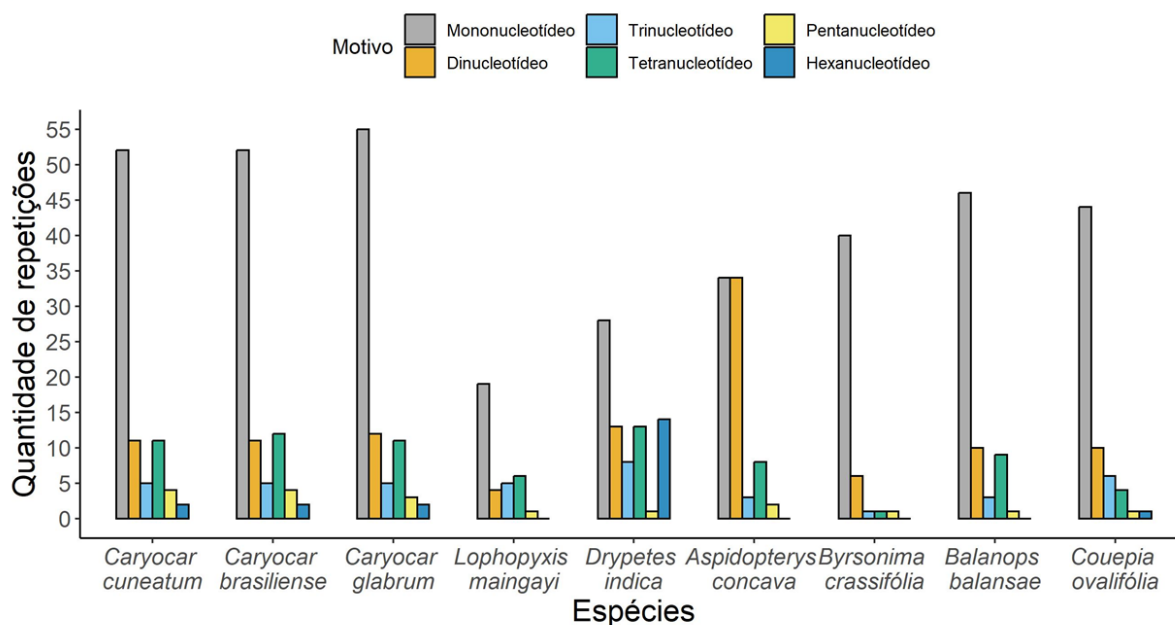


Figura 3. Comparações das regiões repetitivas das nove espécies utilizadas para análises comparativas de Malpighiales

Nota: Disposição dos microssatélites por motivos de repetições (Motif).

Tabela 5: Caracterização de microssatélites dos genomas cloroplastidiais utilizada para análise comparativa das 9 espécies de Malpighiales.

Espécies	Motivo	Quantidade
<i>Caryocar cuneatum</i>	Mononucleotídeo	52
	Dinucleotídeo	11
	Trinucleotídeo	5
	Tetranucleotídeo	11
	Pentanucleotídeo	4
	Hexanucleotídeo	2
<i>Caryocar brasiliense</i>	Mononucleotídeo	52
	Dinucleotídeo	11
	Trinucleotídeo	5
	Tetranucleotídeo	12
	Pentanucleotídeo	4
	Hexanucleotídeo	2
<i>Caryocar glabrum</i>	Mononucleotídeo	55
	Dinucleotídeo	12
	Trinucleotídeo	5
	Tetranucleotídeo	11
	Pentanucleotídeo	3

	Hexanucleotídeo	2
<i>Lophopyxis maingayi</i>	Mononucleotídeo	19
	Dinucleotídeo	4
	Trinucleotídeo	5
	Tetranucleotídeo	6
	Pentanucleotídeo	1
	Hexanucleotídeo	0
<i>Drypetes indica</i>	Mononucleotídeo	28
	Dinucleotídeo	13
	Trinucleotídeo	8
	Tetranucleotídeo	13
	Pentanucleotídeo	1
	Hexanucleotídeo	14
<i>Aspidopterys concava</i>	Mononucleotídeo	34
	Dinucleotídeo	34
	Trinucleotídeo	3
	Tetranucleotídeo	8
	Pentanucleotídeo	2
	Hexanucleotídeo	0
<i>Byrsonima crassifolia</i>	Mononucleotídeo	40
	Dinucleotídeo	6
	Trinucleotídeo	1
	Tetranucleotídeo	1
	Pentanucleotídeo	1
	Hexanucleotídeo	0
<i>Balanops balansae</i>	Mononucleotídeo	46
	Dinucleotídeo	10
	Trinucleotídeo	3
	Tetranucleotídeo	9
	Pentanucleotídeo	1
	Hexanucleotídeo	0
<i>Couepia ovalifolia</i>	Mononucleotídeo	44
	Dinucleotídeo	10
	Trinucleotídeo	6
	Tetranucleotídeo	4

Pentanucleotídeo	1
Hexanucleotídeo	1

O número de repetições complexas entre *Caryocar* variou em um total de 49 repetições para *C. brasiliense*, 49 para *C. glabrum* e 48 para *C. cuneatum*. Entre as outras espécies houve variação de 51 para *L. maingayi*, 300 para *D. indica*, 190 para *A. concava*, 66 para *B. crassifolia*, 61 em *B. balansae* e 70 para *C. ovalifolia*. Dentro do gênero *Caryocar* apenas *C. glabrum* possui repetições *complement* e *C. brasiliense* e *C. cuneatum* junto de *L. maingayi* e *D. indica* não contém esse tipo de repetição. *Polindromic* e *forward* foram as repetições com maior abundância em todas as espécies (Figura 4 e Tabela 6).

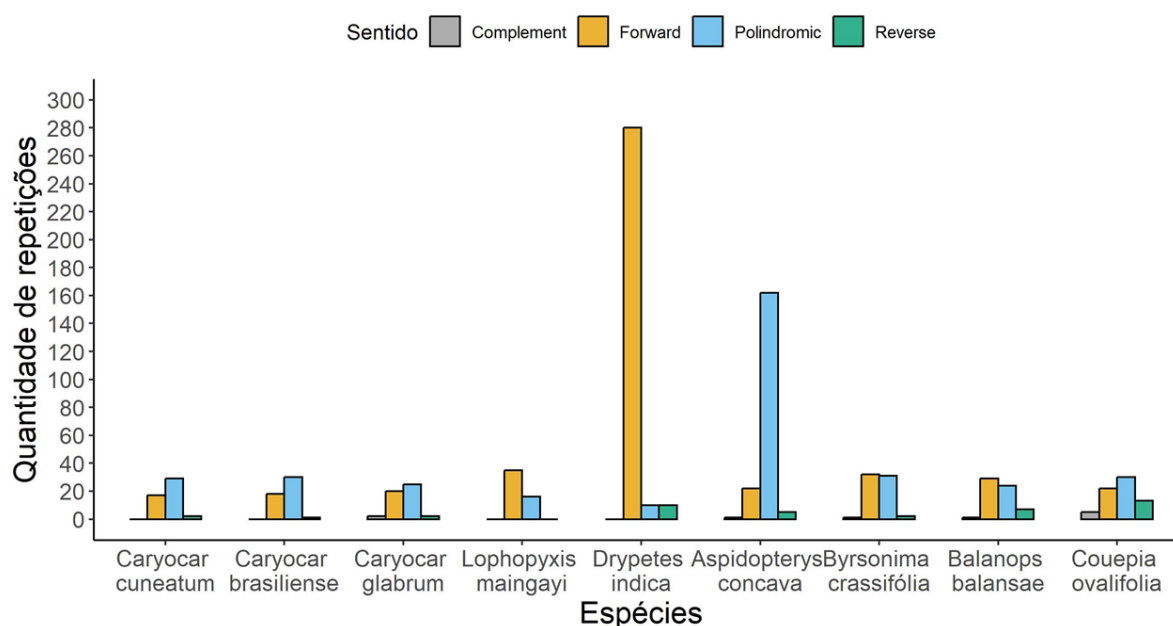


Figura 4. Disposição das repetições complexas por tipos; *Complement*, *forward*, *polindromic* e *reverse*.

Tabela 6: Caracterização do número de repetições complexas para cada espécie utilizando o programa *Reputer*.

Espécies	Sentido	Quantidade
<i>Caryocar cuneatum</i>	<i>Forward</i>	17
	<i>Reverse</i>	2
	<i>Polindromic</i>	29
	<i>Complement</i>	0
<i>Caryocar brasiliense</i>	<i>Forward</i>	18
	<i>Reverse</i>	1

	<i>Polindromic</i>	30
<i>Caryocar glabrum</i>	<i>Complement</i>	0
	<i>Forward</i>	20
	<i>Reverse</i>	2
	<i>Polindromic</i>	25
<i>Lophopyxis maingayi</i>	<i>Complement</i>	2
	<i>Forward</i>	35
	<i>Reverse</i>	0
	<i>Polindromic</i>	16
<i>Drypetes indica</i>	<i>Complement</i>	0
	<i>Forward</i>	280
	<i>Reverse</i>	10
	<i>Polindromic</i>	10
<i>Aspidopterys concava</i>	<i>Complement</i>	0
	<i>Forward</i>	22
	<i>Reverse</i>	5
	<i>Polindromic</i>	162
<i>Byrsonima crassifolia</i>	<i>Complement</i>	1
	<i>Forward</i>	32
	<i>Reverse</i>	2
	<i>Polindromic</i>	31
<i>Balanops balansae</i>	<i>Complement</i>	1
	<i>Forward</i>	29
	<i>Reverse</i>	7
	<i>Polindromic</i>	24
<i>Couepia ovalifolia</i>	<i>Complement</i>	1
	<i>Forward</i>	22
	<i>Reverse</i>	13
	<i>Polindromic</i>	30
	<i>Complement</i>	5

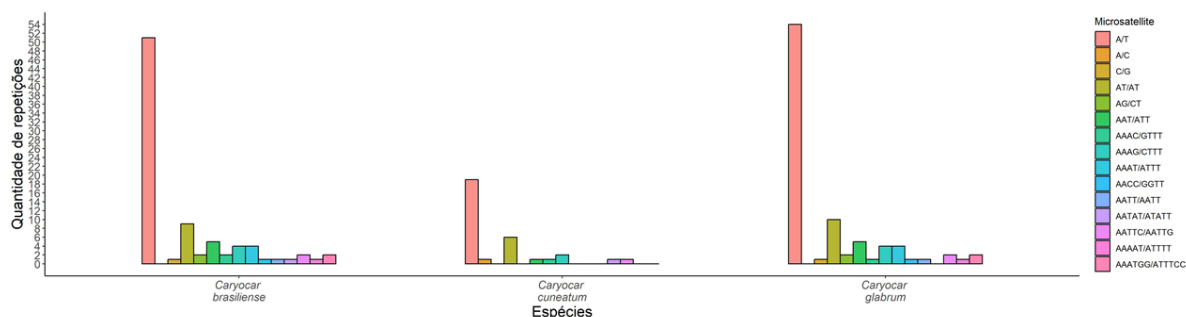


Figura 5: Classificação nucleotídica dos microssatélites de acordo com os seus tipos de motivos espécies de *Caryocar*

O alinhamento feito no Mauve para as três espécies do gênero *Caryocar* revelou alta colinearidade. Não é verificado recombinação entre os genes dessas espécies (Figura 7).

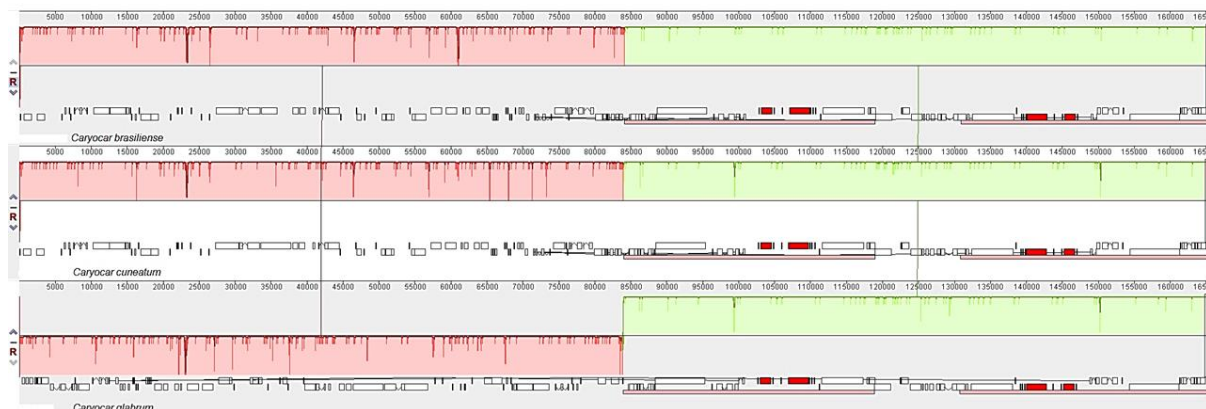


Figura 6. Alinhamento dos genomas cloroplastidiais das espécies *C. brasiliense*, *C. cuneatum* e *C. glabrum*

A diversidade nucleotídica (π) do gênero *Caryocar* variou entre o valor mínimo de 0 até 0,01778 com a mediana de 0,00596 (Figura 7). Foram identificados 2 hotspots de diversidade nucleotídica, esses definidos com valores acima do dobro da mediana de 0,0118. Foram encontrados na posição intergênica *trnS-GCU* - *trnG-UCC*, o valor de π mais alto 0,01778, e *rbcL* - *atpB*, π entre 0.01444 e 0,0133.

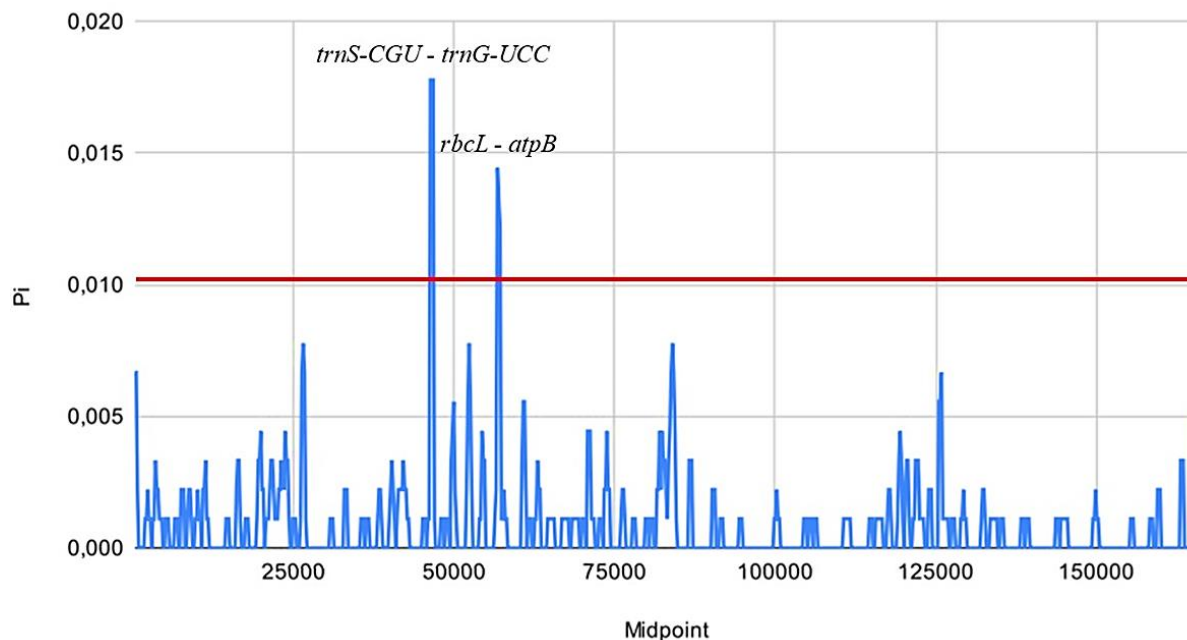


Figura 7: Gráfico gerado pelo cálculo de diversidade nucleotídicas evidenciando duas regiões de hotspots em *Caryocarp*

Nota: Pi. Valor de diversidade nucleotídica. Midpoint. posição nucleotídica.

No cálculo feito para as mutações sinônimas e não sinônimas (K_a/K_s) os valores médios foram de 0,344420857 com variação entre 0 para o gene *psbL* e próximo de 1 para os demais 78 genes e indicando seleção negativa. O gene *rpL20* obteve o valor de 1,018, indicando neutralidade (Figura 8).

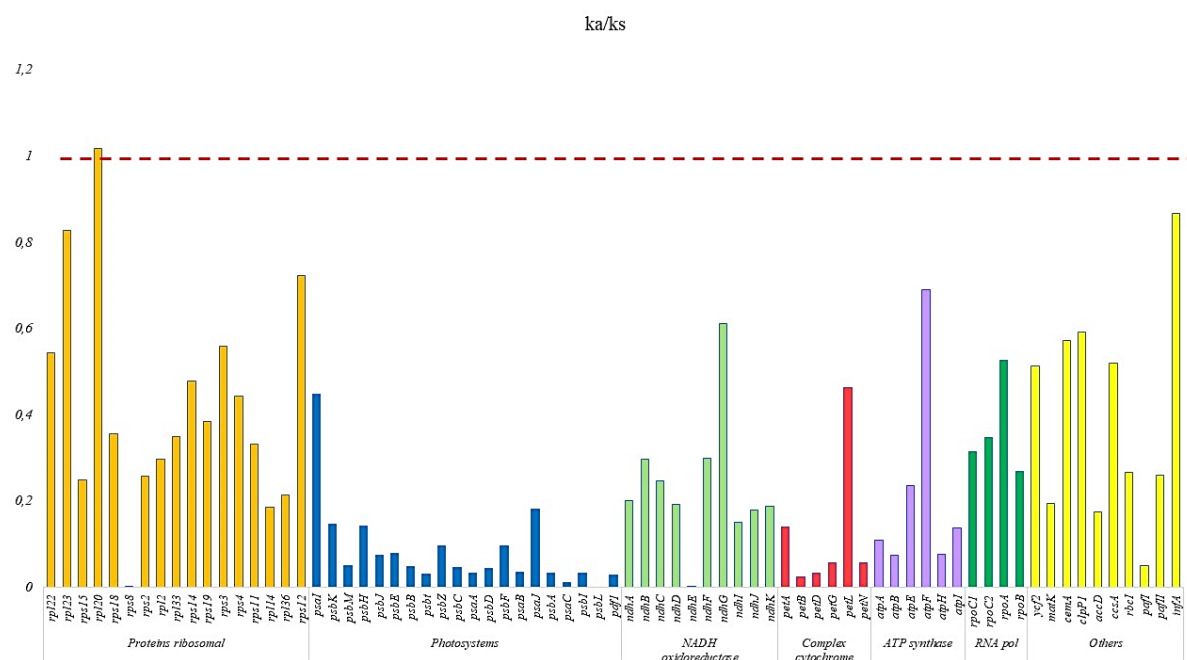


Figura 8: Mutações sinônimas e não sinônimas (K_a/K_s) para as 9 espécies da análise comparativa

A árvore filogenética gerada com 50 CDS resultou em sua maioria nós com altos valores de suportes e alguns abaixo de 75 na família Malpighiaceae, e as espécies *Afrolicania elaeoperma*, *Hymenopus glabriflorus*, *Gaulettia elata*, *Licania alba*, *Leptobalamus sprucei*, *Parastemon urophyllus* e outros dois ramos dentro da família Chrysobalanaceae (Figura 9).

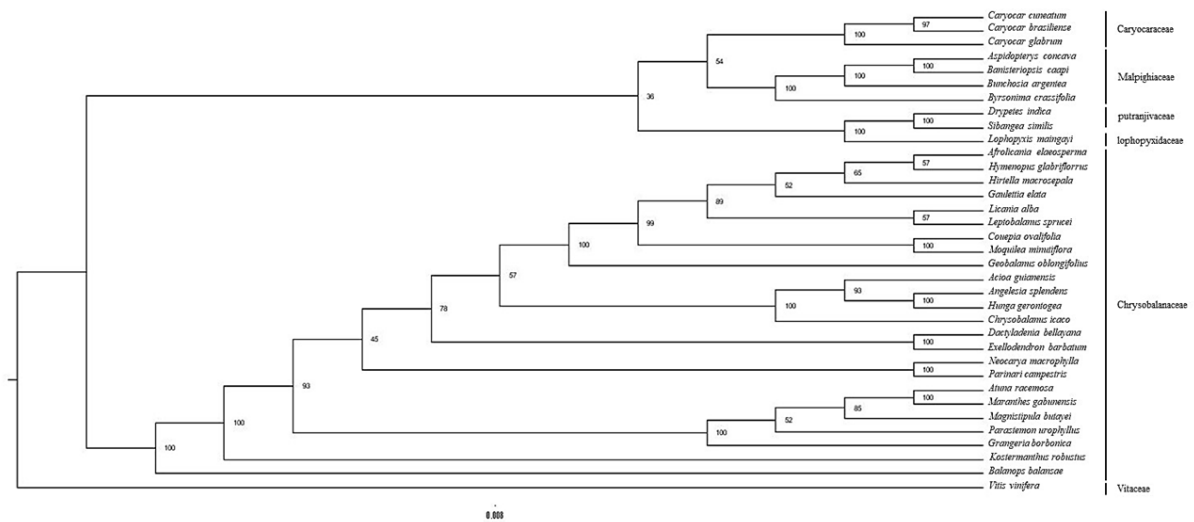


Figura 9. Árvore filogenética obtida pelo método de máximo verossimilhança para a família Malpighiaceae

5. DISCUSSÃO

O tamanho do genoma cloroplastidial de *C. cuneatum* apresenta-se dentro do esperado quando comparado com espécies sequenciadas desse gênero. Em análise comparativa das três espécies do gênero *Caryocar*, os três genomas estão conservados com nenhuma variação no conteúdo GC, na quantidade de genes (131), tanto para os CDS (90) como tRNA (33) e rRNA (8). As pequenas variações ocorrem na parte estrutural no tamanho do genoma, nas LSC, IRs, SSC e na composição de pares de bases, modificando minimamente as repetições microssatélites e complexas, a composição genica é semelhante apresentando os mesmos genes.

O gene *rpl32*, não foi encontrado em *Caryocar*. Ele faz parte da categoria da subunidade ribossômica maior. A perda de genes em genomas cloroplastidiais pode ser explicado por duas formas, a perda total do gene ou sua transferência para o núcleo. Há relatado casos de perda de *rpl32* em outras espécies como: *Populus trichocarpa*, (STEANE,2005) e *Euphorbia schimperi*, o qual sofreram transferência para o núcleo (ALQAHTAHI; JASEN, 2021). Em *Thalictrum coreanum* também é documentado a transferência nuclear através da realização de transcriptoma nuclear (SEONGJUN; JASEN; SEONJOO, 2015). A falta desse gene nos genomas das três espécies de *Caryocar* pode indicar uma característica compartilhada entre todo o grupo, porém é necessário a realização de outros estudos para detectar o seu destino.

Além disso, não está relatado como a falta dos genes *rpl32* pode afetar o desenvolvimento de suas funções biológicas celulares, espera-se que sua atividade não seja comprometida por existir outros genes participantes da mesma categoria de subunidade ribossômica maior, que em sua falta auxiliem na atividade de composição ribossômica.

O gene *InfA* é um pseudogene e se encontra nas oito espécies analisadas menos em *A. concava*, ele já foi descrito como um gene móvel e que sofreu pseudogenização por provavelmente ter sido transferido várias vezes do cloroplasto para o núcleo pode ter perdido partes do seu tamanho e adquirido um códon de parada. Além disso, não se sabe ao certo se ele pode ter sofrido perda total no genoma cloroplastidial ou transferência nuclear (MILLEN et al., 2001).

Nas regiões de transição também há conservação entre os genomas de *Caryocar*, os genes *rps16*, *rps3* estão nos limites de LSC com IRs das três espécies e nas regiões de IR com SSC permanecem o mesmo padrão com os mesmos genes. Nas outras espécies comparadas fora desse gênero, o gene que predomina na região LSC com IRs é o *rps19* que em *A. concava* e *B. balansae* está pseudogenizado.

As análises comparativas de regiões de repetições microssatélites revelou a quantidade total conservada entre os três genomas de *Caryocar*. Existem variações de repetições para as espécies analisadas, o padrão de repetições mononucleotídicas e dinucleotídicas é mais alto dentro dos genomas como esperado. Para as análises de repetições complexas também é evidenciado um padrão conservado, em que as repetições *Polidromic* é mais facilitada em ocorrer nas regiões de IRs, sendo assim, espécies que possuem maior tamanho dessa região terá maior quantidade dessa repetição, como ocorre em *Caryocar*, porém em *A. concava* isso não ocorre. Em *L. maingay* e *D. indica* tem maior número de repetições do tipo *Foward*.

A análise de mutações sinônimas e não sinônimas é importante para entender como está atuando a evolução de genes e a taxa de mutações nucleotídicas, é usada para apontar seleção positiva com valores da razão Ka/Ks acima de 1, seleção negativa com valores abaixo de 1 e sobre neutralidade com valores igual a 1. A maioria dos genes avaliados se encontram em seleção negativa. O gene *rpl20* está sobre neutralidade, indicando que as mutações nele ocorridas não favorecem nenhuma das duas seleções. O *rpl20* pertence a subunidade maior ribossômica e sua função é atuar na autorreplicação, a neutralidade aponta que a seleção natural não está agindo sobre ele e não tem impacto nas realizações de suas funções biológicas. Para os demais genes que estão em seleção negativa indica o caráter de conservação, não favorecendo mutações não sinônimas sobre eles. Nenhum gene se encontra em seleção positiva o que não favorece a diversificação e sim sua conservação.

Investigando a diversidade nucleotídica nos genomas de *Caryocar* aqui apresentados, foi possível identificar duas regiões de *hotspots*: *trnS-GCU - trnG-UCC* e *rbcL - atpB*, é importante ressaltar que essas regiões de diversidade se deram em espaços intergênicos, que em geral apresentam maior diversidade nucleotídica do que regiões gênicas que tendem ser mais conservadas. A região *trnS-GCU - trnG-UCC* também já foi identificada em outras espécies como *hotspots* e podem ser utilizados como marcadores moleculares de *DNA barcoding* em *Phyllanthus urinaria* (FANG, et al., 2022) *Echinacanthus* (GAO; DENG; WANG, 2019). Essas duas regiões de *hotspots* podem ter utilizadas como possíveis marcadores moleculares de *DNA barcoding* discriminando *C. cuneatum* das demais espécies do gênero, para sua validação e utilização medicinal, gastronômica e outros fins comerciais.

A árvore filogenética obteve bons valores de suporte para os nós em *Caryocaraceae*, validando a posição sistemática de *C. cuneatum* dentro da família e evidência sua relação com seus grupos irmãos *C. brasiliense* e *C. glabrum*. Como esperado, as espécies se enquadraram dentro das suas respectivas famílias e clados, porém para análise de parentesco entre essas famílias os valores de suportes para nós ficaram fracos abaixo de 75. Dessa forma, não é

possível ter precisão para verificar suas relações sistemáticas. Faz-se necessário mais estudos genômicos e morfológicos e incluir novos táxons pertencentes as famílias aqui analisadas para melhorar a resolução da relação filogenética pertencentes a esses clados citados.

6. CONCLUSÕES

Neste trabalho, foi possível disponibilizar pela primeira vez o genoma cloroplastidial da espécie *C. cuneatum*. A partir das análises comparativas, conclui-se que existe baixa variação entre os genomas das espécies *C. Cuneatum*, *C. brasilense* e *C. glabrum*. Para as outras espécies, ocorrem variação no tamanho e estrutura genômica, principalmente *L. maingayi* e *D. indica* que apresentam o menor tamanho genômico e não contém a IRa. Foi possível identificar duas regiões de *hotspots*: *trnS-GCU - trnG-UCC* e *rbcL - atpB* e podem ser usadas como marcador molecular para *DNA barcoding* como descrito em outros estudos. A árvore filogenética evidencia as relações entre as espécies de *Caryocar* e valida a posição taxonômica de *C. cuneatum*. Em conclusão, esse estudo fornece o primeiro genoma de *C. cuneatum* sequenciado, informação sobre a caracterização genômica e os conhecimentos aqui fornecidos podem ser utilizadas para novos estudos e desenvolvimentos de tecnológicas como *DNA barcoding* e estudos filogenéticos.

7. REFERÊNCIA

- ALMEIDA, S. P. DA; SILVA, J. A. DA. Piqui e buriti: importância alimentar para a população dos cerrados. **Embrapa-Cpac**, Planaltina, v. 54. p. 38, 1994. Disponível em: <<https://www.infoteca.cnptia.embrapa.br/bitstream/doc/548665/1/doc54.pdf>> Acesso em 2. Mar. 2023.
- ALGAHTANI, A. A.; JANSEN, K. R. The evolutionary fate of *rpl32* and *rps16* losses in the *Euphorbia schimperi* (Euphorbiaceae) plastome. *Scientific Reports*. v.11 n.1 p. 7466 DOI: 10.1038/s41598-021-86820-z.
- ALVES, H. S. Sequenciamento e análise do genoma cloroplastidial de eucalipto (*Eucalyptus grandis*). **Esalq**, p. 64, 2005.
- APG IV. The Catalogue of Life Partnership. APG IV: Angiosperm Phylogeny Group classification for the orders and families of flowering plants. 2016, disponível em: <<https://doi.org/10.15468/fzuaam>> Acesso em: 18 mar. 2023.
- BEIER, S. et al. MISA-web: A web server for microsatellite prediction. **Bioinformatics**, v. 33, n. 16, p. 2583–2585, 2017.
- BEZERRA, N. K. M. S.; BARROS, T. L.; COELHO, N. P. M. F. ação do óleo de pequi (*Caryocar brasiliense*) no processo cicatricial de lesões cutâneas em ratos. **Rev. brasileira de plantas medicinais**, Campinas, v. 17, n. 4, p.875-880, 2015. DOI: https://doi.org/10.1590/1983-084X/14_061, Disponível em <<https://www.scielo.br/j/rbpm/a/qvVGZBDmZnXT7Q5twpcs74s/?format=pdf&lang=pt> > Acesso em 1 fev. 2023.
- BOBIK, K.; BURCH-SMITH, T. M. Chloroplast signaling within, between and beyond cells. **Frontiers in Plant Science**, v. 6, n. 781, p. 1–26, 2015. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4593955/> > Acesso em: 10 dez. 2022.
- BRAIT, D, A. J. **características físicas e químicas de pequizeiro (*Caryocar brasiliense* CAMB) de ocorrência natural em rio verde- GO**. Dissertação (Agricultura) Universidade Federal de Goiás, Goiás, 2008. Disponível em: [https://files.cercomp.ufg.br/weby/up/217/o/Joys_Dias_de_Assis_Brait_%E2%80%9393_Caracter%C3%ADsticas_f%C3%ADsticas_e_qu%C3%ADmicas_de_pequizeiro_\(Caryocar_brasiliense_CAMB\)_de_ocorr%C3%Aancia_natural_em_Rio.PDF?1489243867](https://files.cercomp.ufg.br/weby/up/217/o/Joys_Dias_de_Assis_Brait_%E2%80%9393_Caracter%C3%ADsticas_f%C3%ADsticas_e_qu%C3%ADmicas_de_pequizeiro_(Caryocar_brasiliense_CAMB)_de_ocorr%C3%Aancia_natural_em_Rio.PDF?1489243867) > Acesso em: 02 mar. 2023.
- BUSO, G. S. C. Marcadores moleculares e análise filogenética. **Embrapa**, p. 22, 2005.
- CARVALHO, L. S.; PEREIRA, K. F.; ARAÚJO, E. G. Características botânicas, efeitos terapêuticos e princípios ativos presentes no pequi (*caryocar brasiliense*). *Arquivo Ciências da Saúde*, v. 19, n. 2, p. 147-157, 2015. Disponível em: <https://ojs.revistasunipar.com.br/index.php/saude/article/view/5435/3122> > Acesso em: 12 jan. 2023.
- CARVALHO, P. E. R. Espécies arbóreas brasileiras. **Embrapa Informação Tecnológica**. Embrapa Florestas, Brasília, v. 3, 2008. Disponível em <<https://ainfo.cnptia.embrapa.br/digital/bitstream/CNPF-2010/46379/1/CT230.pdf>> Acesso em 10. mar. 2023.
- COLLEVATTI, R. G.; VIANELLO, R.; GRATTAPAGLIA, D. Development and characterization of microsatellite markers for genetic analysis of a Brazilian endangered tree species *Caryocar brasiliense*. **Heredity**, v. 83, n. 1, p. 748–756, 1999.
- DANIELL, H. et al. Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. **Genome Biology**, v. 17, n. 1, p. 1–29, 2016.
- DAVIS, C. C. et al. Explosive radiation of Malpighiales supports a mid-cretaceous origin of modern tropical rain forests. **The American naturalist**, v. 165, n. 3, 2005.

- DARLING, A. C. E. et al. Implicitfunction.Pdf. **Genome Research**, v. 14, n. 7, p. 1394–1403, 2004.
- DIERCKXENS, N.; MARDULYN, P.; SMITS, G. NOVOPlasty: De novo assembly of organelle genomes from whole genome data. **Nucleic Acids Research**, v. 45 n. 4, p. 2-9, 2017.
- DOMINGUES, H. M. B. O jardim botânico do Rio de Janeiro. **Espaços da Ciência no Brasil: 1800-1930**, v. XIX, p. 15–56, 2001.
- DONG, W. et al. ycf1, the most promising plastid DNA barcode of land plants. **Scientific Reports**, v. 5, p. 8348, 2015. DOI: 10.1038/srep08348.
- DOYLE, J. J.; DOYLE, J. L. **Doyle_plantDNAextractCTAB_1987.pdf**. **Phytochemical Bulletin**New York, 1987. Disponível em: <https://webpages.uncc.edu/~jeweller2/pages/BINF8350f2011/BINF8350_Readings/Doyle_plantDNAextractCTAB_1987.pdf> Acesso em 20 já. 2023.
- EHRENDORFER, F.; MORAWETZ, W.; DAWE, J. The neotropical angiosperm families Brunelliaceae and Caryocaraceae: First karyosystematical data and affinities. **Plant Systematics and Evolution**, v. 145, n. 3–4, p. 183–191, 1984.
- FANG, H. et al. Application of chloroplast genome in the identification of *Phyllanthus urinaria* and its common adulterants. **Front Plant**. China, v. 13, 2022. DOI: 10.3389/fpls.2022.1099856.
- GAO, C.; DENG, Y.; WANG, J.; The Complete Chloroplast Genomes of *Echinacanthus* Species (Acanthaceae): Phylogenetic Relationships, Adaptive Evolution, and Screening of Molecular Markers. **Front Plant Sci**. v. 9, n. 1989, 2019. doi: 10.3389/fpls.2018.01989.
- JANSEN, R.K, RUHLMAN T.A. **Plastid Genomes of Seed Plants**. **Adv Photosynth Respir**. Dordrecht. Springer, v. 35 p. 103–126, 2012. <https://doi.org/10.1007/978-94-007-2920-9>.
- JIN, D. M. et al. The Loss of the Inverted Repeat in the Putranjivoid Clade of Malpighiales. **Frontiers in Plant Science**, v. 11, n. June, p. 1–9, 2020.
- KALYAANAMOORTHY, S. et al. ModelFinder: Fast model selection for accurate phylogenetic estimates. **Nature Methods**, v. 14, n. 6, p. 587–589, 2017.
- KATOH, K.; STANDLEY, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. **Molecular Biology and Evolution**, v. 30, n. 4, p. 772–780, 2013.
- KEARSE, M. et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. **Bioinformatics**, v. 28, n. 12, p. 1647–1649, 2012.
- KENT, W. J. BLAT —The BLAST -Like Alignment Tool. **Genome Research**, v. 12, n. 4, p. 656–664, 2002.
- KUBITZKI, K. The Families and Genera of Bats. **The American Naturalist**, v. 41, n. 490, p. 671–672, 1907.
- KURTZ, S. et al. REPuter: The manifold applications of repeat analysis on a genomic scale. **Nucleic Acids Research**, v. 29, n. 22, p. 4633–4642, 2001.
- LOHSE, M. et al. OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. **Nucleic acids research**, v. 41, n.1, p. 575–581, 2013.
- MARGULIS, L.; BERMUDES, D. Symbiosis as a mechanism of evolution: status of cell symbiosis theory. **Symbiosis**, v. 1, p. 101–124, 1985.
- MEDEIROS, H.; AMORIM, A. M. A. *Caryocaraceae*. **Lista de Espécies da Flora do Brasil**. Jardim Botânico do Rio de Janeiro. 2015. Disponível em: <<http://floradobrasil2015.jbrj.gov.br/jabot/floradobrasil/FB6687>> Acesso em 10 mar. 2023.

- MILLEN, R. S. et al. Many parallel losses of infa from chloroplast dna during angiosperm evolution with multiple independent transfers to the nucleus. **Plant Cell**, v. 13, n. 3, p. 645–658, 2001.
- JIN, M. D. et al. The Loss of the Inverted Repeat in the Putranjivoid Clade of Malpighiales. v.11 n. 942, 2020. OI: 10.3389/fpls.2020.00942.
- MIRANDA-VILELA, A. L. et al. Pequi fruit (*Caryocar brasiliense* Camb.) pulp oil reduces exercise-induced inflammatory markers and blood pressure of male and female runners. **Nutrition Research**, v. 29, n. 12, p. 850–858, 2009. DOI: 10.1016/j.nutres.2009.10.022. acesso em: <<https://www.sciencedirect.com/science/article/abs/pii/S0271531709002073?via%3Dihub>> Acesso em: 12 jan. 2023.
- MIRANDA-VILELA, A. L. et al. Dextran-functionalized magnetic fluid mediating magnetohyperthermia for treatment of Ehrlich-solid-tumor-bearing mice: Toxicological and histopathological evaluations. **Tumor Biology**, v. 35, n. 4, p. 3391–3403, 2014. Disponível em <<https://link.springer.com/article/10.1007/s13277-013-1447-y>> Acesso em: 12 jan. 2023.
- NEI, M. KUMAR, S. Molecular evolution and phylogenetics. **Oxford University Press**, v. 86, p.385, 2000. DOI<<https://doi.org/10.1046/j.1365-2540.2001.0923a.x>>.
- NEWMASTER, S. G.; FAZEKAS, A. J.; RAGUPATHY, S. DNA barcoding in land plants: Evaluation of rbcL in a multigene tiered approach. **Canadian Journal of Botany**, v. 84, n. 3, p. 335–341, 2006.
- NGUYEN, L. T. et al. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. **Molecular Biology and Evolution**, v. 32, n. 1, p. 268–274, 2015.
- NUNES, R. et al. Caryocaraceae Voigt (Malpighiales): a Synthesis Based on Science Mapping and Systematic Review. **Botanical Review**, New York, v. 86, n. 3–4, p. 338–358, 2020.
- NUNES, R. et al. Complete chloroplast genome sequence of caryocar brasiliense camb. (caryocaraceae) and comparative analysis brings new insights into the plastome evolution of Malpighiales. **Genetics and Molecular Biology**, v. 43, n. 2, p. 1–7, 2020.
- OKONECHNIKOV, K. et al. Unipro UGENE: A unified bioinformatics toolkit. **Bioinformatics**, v. 28, n. 8, p. 1166–1167, 2012.
- OLDENBURG, D. J.; BENDICH, A. J. The linear plastid chromosomes of maize: terminal sequences, structures, and implications for DNA replication. **Current Genetics**, v. 62, n. 2, p. 431–442, 2016.
- OLDENBURG, D. J.; BENDICH, A. J. DNA maintenance in plastids and mitochondria of plants. **Frontiers in Plant Science**, v. 6, n. OCTOBER, p. 1–15, 2015.
- OLIVEIRA, M. E. B. et al. Aspetos Agronômicos e de Qualidade do Pequi. **Embrapa**, Fortaleza v. 113, p. 33, 2008. Disponível em <<https://www.infoteca.cnptia.embrapa.br/infoteca/bitstream/doc/426706/1/Dc113.pdf>> Acesso em 1 mar. 2023.
- PEREIRA, F. F. G. et al. Characterization, antibacterial activity and antibiotic modifying action of the *Caryocar coriaceum* Wittm. pulp and almond fixed oil. **Natural Product Research**, v. 34, n. 22, p. 3239–3243, 2020.
- POZO, O. V. C. *O pequi (Caryocar brasiliense): uma alternativa para o desenvolvimento sustentável do Cerrado no norte de Minas Gerais*. Dissertação (Mestrado em Administração Rural) - Universidade Federal de Lavras. Lavras, 1997.
- PRANCE, G. T.; SILVA, M. F. A monograph of Caryocaraceae. **Flora neotropica**. The New York Botanical Garden, New York, v. 12, n. p. 1-17. 1973.
- PRANCE, G. T. No Title Caryocaraceae. Flora of the Venezuelan. **Missouri Botanical Garden Press**, v. 4, p. 164–170, 1998.

PRANCE, G. T. The Genus *Caryocar* L. (Caryocaraceae): An Underexploited Tropical Resource. **Advances in Economic Botany**, v. 8, p. 177–188, 1990.

PRANCE, G. T. **The Families and Genera of Vascular Plants**. Flowering Plants. Eudicots: Malpighiales. Kubitzki. Caryocaraceae. Kubitzki. p. 1-332. v. 11. Berlin. 2014.

PRANCE, G.T.; PIRAPI, J.R. *Caryocaraceae in Flora e Funga do Brasil*. Jardim Botânico do Rio de Janeiro. 2020. Disponível em: <<https://floradobrasil.jbrj.gov.br/FB6688>>. Acesso em: 16 mar. 2023.

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

QU, X. J. et al. Plastome Reduction in the only Parasitic Gymnosperm *Parasitaxus* Is Due to Losses of Photosynthesis but Not Housekeeping Genes and Apparently Involves the Secondary Gain of a Large Inverted Repeat. **Genome Biology and Evolution**, v. 11, n. 10, p. 2789–2796, 2019.

RAMOS, B. H. et al. Anatomy and micromorphometry of *Caryocar brasiliense* leaves. **Rodriguesia**, v. 66, n. 1, p. 87–94, 2015.

RIBEIRO, A. E. O espaço, o homem e o seu destino no norte de Minas. In: UNIVERSIDADE FEDERAL DE LAVRAS. Departamento de Administração e Economia. Manejo sustentado do cerrado para uso múltiplo: subprojeto agroecologia e desenvolvimento. Lavras, 1996. p. 11-18

RODRÍGUEZ-EZPELETA, N. et al. Monophyly of primary photosynthetic eukaryotes: Green plants, red algae, and glaucophytes. **Current Biology**, v. 15, n. 14, p. 1325–1330, 2005. DOI: <https://doi.org/10.1016/j.cub.2005.06.040>.

ROZAS, J. et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets. **Molecular Biology and Evolution**, v. 34, n. 12, p. 3299–3302, 2017.

SATO, N. Complex origins of chloroplast membranes with photosynthetic machineries: multiple transfers of genes from divergent organisms at different times or a single endosymbiotic event? **Journal of Plant Research**, China, v. 133, n. 1, p. 15–33, 2020. Doi: 10.1007/s10265-019-01157-z.

SEONGJUN, P.; JANSEN, K. R.; SEONJOO, P.; Complete plastome sequence of *Thalictrum coreanum* (Ranunculaceae) and transfer of the *rpl32* gene to the nucleus in the ancestor of the subfamily Thalictroideae. **BMC Plant Biol.** V. 15, n. 1, p. 40, 2015. DOI: 10.1186/s12870-015-0432-6.

SILVA, P. J. et al. Aspectos da comercialização do pequi (*Caryocar brasiliense* Camb.) no estado de Goiás, Brasil, **Encicoplédia Biósfera**, Jandaia-GO, v.18 n.37; p. 61. 2021. DOI: 10.18677/EnciBio_2021C6

STEANE, A. D. Complete Nucleotide Sequence of the Chloroplast Genome from the Tasmanian Blue Gum, *Eucalyptus globulus* (Myrtaceae), DNA Research, Japão, v. 12 n.5 p. 215-220. 2005. DOI: <https://doi.org/10.1093/dnares/dsi006>.

SHEPHERD, G. J.; GIULIETTI, A. M. **Flora Fanerogâmica do Estado de São Paulo**. Instituto de botânica São Paulo v 3, 2003.

TALAVERA, G.; CASTRESANA, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. **Systematic Biology**, v. 56, n. 4, p. 564–577, 2007.

TILLICH, M. et al. GeSeq - Versatile and accurate annotation of organelle genomes. **Nucleic Acids Research**, v. 45, n. 1, p. 6–11, 2017.

VITTA, F. **Flora Fanerogâmica do Estado de São Paulo**. Instituto de botânica. São Paulo v. 5, 2007

VIEIRA, L. D. N. DESENVOLVIMENTO DE UM PROTOCOLO DE CONÍFERAS E SEQUENCIAMENTO DO GENOMA PLASTIDIAL DE *Podocarpus lambertii* Klotzch ex Endl . Florianópolis. **Dissertação**

submetida ao Programa de Pós-Graduação em Recursos Genéticos Vegetais da Universidade Federal de Santa Catarina, p. 79, 2014.

WICKE, S., NAUMANN, J. Molecular evolution of plastid genomes in parasitic flowering plants. *botanical research*, Reino Unido, v. 85, p. 315–347, 2018.

XI, Z. et al. Phylogenomics and a posteriori data partitioning resolve the Cretaceous angiosperm radiation Malpighiales. *Proceedings of the National Academy of Sciences of the United States of America*, Washington, v. 109, n. 43, p. 17519–17524, 2012. DOI: <https://doi.org/10.1073/pnas.120581810>.

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.

ZAPPI, D. C. et al. Growing knowledge: an overview of seed plant diversity in Brazil. *Rodriguésia*, Rio de Janeiro, v. 66, n. 4, p. 1085-1113, 2015. Disponível em: <http://dx.doi.org/10.1590/2175-7860201566411>. Acesso em: 04 mar. 2023.

8. ANEXOS

Tabela S1: Genomas retirados no banco de dados no NCBI pertencentes ao clados Caryocaraceae, Putranjivoid, Malpighioid e Chrysobalanoid e seus respectivos números de referência.

Espécies	Replicons	Clado	Famílias
<i>Caryocar cuneatum</i>		Caryocaraceae	Caryocaraceae
<i>Caryocar glabrum</i>	NC_061179.1/MN116504.1	Caryocaraceae	Caryocaraceae
<i>Caryocar brasiliense</i>	MK726375	Caryocaraceae	Caryocaraceae
<i>Lophopyxis maingayi</i>	NC_050684.1/MN504795.0	Putranjivoid	lophopyxidaceae
<i>Drypetes chevalieri</i>	NC_051999.1/MN504788.1	Putranjivoid	putranjivaceae
<i>Drypetes diopa</i>	NC_052000.1/MN504789.1	Putranjivoid	putranjivaceae
<i>Drypetes hainanensis</i>	NC_052001.1/MN504790.1	Putranjivoid	putranjivaceae
<i>Drypetes indica</i>	NC_052002.1/MN504791.1	Putranjivoid	putranjivaceae
<i>Drypetes lateriflora</i>	NC_052003.1/MN504792.1	Putranjivoid	putranjivaceae
<i>Drypetes longifolia</i>	NC_052004.1/MN504793.1	Putranjivoid	putranjivaceae
<i>Sibangea similis</i>	NC_052005.1/MN504794.1	Putranjivoid	putranjivaceae
<i>Aspidopterys concava</i>	NC_060542.1/OL471043.1	Malpighioid	Malpighiaceae
<i>Aspidopterys obcordata</i>	NC_049898.1/MT590775.1	Malpighioid	Malpighiaceae
<i>Banisteriopsis caapi</i>	NC_037945.1/MH041648.1	Malpighioid	Malpighiaceae
<i>Bunchosia argentea</i>	NC_041491.1/MK390344.1	Malpighioid	Malpighiaceae
<i>Byrsonima coccolobifolia</i>	NC_037191.1/MF359247.1	Malpighioid	Malpighiaceae
<i>Byrsonima crassifolia</i>	NC_037192.1/MF359248.1	Malpighioid	Malpighiaceae

<i>Balanops balansae</i>	NC_052006.1/MN504796.1	Chrysobalanoid	Balanopaceae
<i>Balanops pedicellata</i>	NC_052007.1/MN504797.1	Chrysobalanoid	Balanopaceae
<i>Acioa guianensis</i>	NC_030534.1/KX158870.1	Chrysobalanoid	Chrysobalanaceae
<i>Afrolicania elaeosperma</i>	NC_030544.1/KX180050.1	Chrysobalanoid	Chrysobalanaceae
<i>Angelesia splendens</i>	NC_030545.1/KX180051.1	Chrysobalanoid	Chrysobalanaceae
<i>Atuna racemosa</i>	NC_030546.1/KX180052.1	Chrysobalanoid	Chrysobalanaceae
<i>Chrysobalanus icaco</i>	NC_024061.1/KJ414480.1	Chrysobalanoid	Chrysobalanaceae
<i>Couepia caryophylloides</i>	NC_030547.1/KX180053.1	Chrysobalanoid	Chrysobalanaceae
<i>Couepia grandiflora</i>	NC_030548.1/KX180054.1	Chrysobalanoid	Chrysobalanaceae
<i>Couepia guianensis</i>	NC_024063.1/KJ414482.1	Chrysobalanoid	Chrysobalanaceae
<i>Couepia ovalifolia</i>	NC_030549.1/KX180055.1	Chrysobalanoid	Chrysobalanaceae
<i>Couepia paraensis</i>	NC_030550.1/KX180056.1	Chrysobalanoid	Chrysobalanaceae
<i>Couepia polyandra</i>	NC_030551.1/KX180058.1	Chrysobalanoid	Chrysobalanaceae
<i>Couepia rankiniae</i>	NC_030552.1/KX180059.1	Chrysobalanoid	Chrysobalanaceae
<i>Couepia sandwithii</i>	NC_030553.1/KX180060.1	Chrysobalanoid	Chrysobalanaceae
<i>Couepia subcordata</i>	NC_030554.1/KX180061.1	Chrysobalanoid	Chrysobalanaceae
<i>Dactyladenia bellayana</i>	NC_030555.1/KX180062.1	Chrysobalanoid	Chrysobalanaceae
<i>Dactyladenia buchneri</i>	NC_030556.1/KX180063.1	Chrysobalanoid	Chrysobalanaceae
<i>Dactyladenia floretii</i>	NC_030557.1/KX180064.1	Chrysobalanoid	Chrysobalanaceae
<i>Exellodendron barbatum</i>	NC_030558.1/KX180065.1	Chrysobalanoid	Chrysobalanaceae
<i>Gaulettia elata</i>	NC_030559.1/KX180066.1	Chrysobalanoid	Chrysobalanaceae
<i>Geobalanus oblongifolius</i>	NC_030571.1/KX180079.1	Chrysobalanoid	Chrysobalanaceae
<i>Grangeria borbonica</i>	NC_030560.1/KX180067.1	Chrysobalanoid	Chrysobalanaceae
<i>Hirtella macrosepala</i>	NC_030561.1/KX180068.1	Chrysobalanoid	Chrysobalanaceae

<i>Hirtella physophora</i>	NC_024066.1/KJ414485.1	Chrysobalanoid	Chrysobalanaceae
<i>Hirtella racemosa</i>	NC_024060.1/KJ414479.1	Chrysobalanoid	Chrysobalanaceae
<i>Hirtella suffulta</i>	NC_030562.1/KX180070.1	Chrysobalanoid	Chrysobalanaceae
<i>Hirtella zanzibarica</i>	NC_030563.1/KX180071.1	Chrysobalanoid	Chrysobalanaceae
<i>Hunga gerontogea</i>	NC_030564.1/KX180072.1	Chrysobalanoid	Chrysobalanaceae
<i>Hymenopus glabriflorus</i>	NC_030567.1/KX180075.1	Chrysobalanoid	Chrysobalanaceae
<i>Hymenopus heteromorphus</i>	NC_024062.1/KJ414481.1	Chrysobalanoid	Chrysobalanaceae
<i>Hymenopus macrophyllus</i>	NC_030568.1/KX180076.1	Chrysobalanoid	Chrysobalanaceae
<i>Kostermanthus robustus</i>	NC_030565.1/KX180073.1	Chrysobalanoid	Chrysobalanaceae
<i>Leptobalanus sprucei</i>	NC_024065.1/KJ414484.1	Chrysobalanoid	Chrysobalanaceae
<i>Licania alba</i>	NC_024064.1/KJ414483.1	Chrysobalanoid	Chrysobalanaceae
<i>Licania canescens</i>	NC_030566.1/KX180074.1	Chrysobalanoid	Chrysobalanaceae
<i>Licania majuscula</i>	NC_030569.1/KX180077.1	Chrysobalanoid	Chrysobalanaceae
<i>Licania membranacea</i>	NC_030570.1/KX180078.1	Chrysobalanoid	Chrysobalanaceae
<i>Licania micrantha</i>	NC_030572.1/KX180080.1	Chrysobalanoid	Chrysobalanaceae
<i>Licania ovalifolia</i>	NC_030574.1/KX180082.1	Chrysobalanoid	Chrysobalanaceae
<i>Magnistipula butayei</i>	NC_030576.1/KX180084.1	Chrysobalanoid	Chrysobalanaceae
<i>Maranthes gabunensis</i>	NC_030577.1/KX180085.1	Chrysobalanoid	Chrysobalanaceae
<i>Maranthes glabra</i>	NC_030578.1/KX180086.1	Chrysobalanoid	Chrysobalanaceae
<i>Maranthes kerstingii</i>	NC_030579.1/KX180087.1	Chrysobalanoid	Chrysobalanaceae
<i>Moquilea minutiflora</i>	NC_030573.1/KX180081.1	Chrysobalanoid	Chrysobalanaceae
<i>Moquilea tomentosa</i>	NC_030575.1/KX180083.1	Chrysobalanoid	Chrysobalanaceae
<i>Neocarya macrophylla</i>	NC_030580.1/KX180088.1	Chrysobalanoid	Chrysobalanaceae
<i>Parastemon urophyllus</i>	NC_030517.1/KX180089.1	Chrysobalanoid	Chrysobalanaceae

<i>Parinari campestris</i>	NC_024067.1/KJ414486.1	Chrysobalanoid	Chrysobalanaceae
<i>Parinari capensis</i>	NC_030581.1/KX180090.1	Chrysobalanoid	Chrysobalanaceae
<i>Parinari curatellifolia</i>	NC_030582.1/KX180091.1	Chrysobalanoid	Chrysobalanaceae
<i>Parinari oblongifolia</i>	NC_030583.1/KX180092.1	Chrysobalanoid	Chrysobalanaceae
Total de gêneros	33 gêneros		
Total de espécies	68 espécies		

Figura S2: Tabela das 34 espécies utilizada para análise filogenética e um grupo externo

Espécies	Replicons	Clado	Famílias
<i>Caryocar cuneatum</i>		Caryocaraceae	Caryocaraceae
<i>Caryocar glabrum</i>	NC_061179.1/MN116504.1	Caryocaraceae	Caryocaraceae
<i>Caryocar brasiliense</i>	MK726375	Caryocaraceae	Caryocaraceae
<i>Lophopyxis maingayi</i>	NC_050684.1/MN504795.0	Putranjivoid	lophopyxidaceae
<i>Drypetes indica</i>	NC_052002.1/MN504791.1	Putranjivoid	putranjivaceae
<i>Sibangea similis</i>	NC_052005.1/MN504794.1	Putranjivoid	putranjivaceae
<i>Aspidopterys concava</i>	NC_060542.1/OL471043.1	Malpighioid	Malpighiaceae
<i>Banisteriopsis caapi</i>	NC_037945.1/MH041648.1	Malpighioid	Malpighiaceae
<i>Bunchosia argentea</i>	NC_041491.1/MK390344.1	Malpighioid	Malpighiaceae
<i>Byrsonima crassifolia</i>	NC_037192.1/MF359248.1	Malpighioid	Malpighiaceae
<i>Balanops balansae</i>	NC_052006.1/MN504796.1	Chrysobalanoid	Balanopaceae
<i>Acioa guianensis</i>	NC_030534.1/KX158870.1	Chrysobalanoid	Chrysobalanaceae

<i>Afrolicania elaeosperma</i>	NC_030544.1/KX180050.1	Chrysobalanoid	Chrysobalanaceae
<i>Angelesia splendens</i>	NC_030545.1/KX180051.1	Chrysobalanoid	Chrysobalanaceae
<i>Atuna racemosa</i>	NC_030546.1/KX180052.1	Chrysobalanoid	Chrysobalanaceae
<i>Chrysobalanus icaco</i>	NC_024061.1/KJ414480.1	Chrysobalanoid	Chrysobalanaceae
<i>Couepia ovalifolia</i>	NC_030549.1/KX180055.1	Chrysobalanoid	Chrysobalanaceae
<i>Dactyladenia bellayana</i>	NC_030555.1/KX180062.1	Chrysobalanoid	Chrysobalanaceae
<i>Exellodendron barbatum</i>	NC_030558.1/KX180065.1	Chrysobalanoid	Chrysobalanaceae
<i>Gaulettia elata</i>	NC_030559.1/KX180066.1	Chrysobalanoid	Chrysobalanaceae
<i>Geobalanus oblongifolius</i>	NC_030571.1/KX180079.1	Chrysobalanoid	Chrysobalanaceae
<i>Grangeria borbonica</i>	NC_030560.1/KX180067.1	Chrysobalanoid	Chrysobalanaceae
<i>Hirtella macrosepala</i>	NC_030561.1/KX180068.1	Chrysobalanoid	Chrysobalanaceae
<i>Hunga gerontogea</i>	NC_030564.1/KX180072.1	Chrysobalanoid	Chrysobalanaceae
<i>Hymenopus glabriflorus</i>	NC_030567.1/KX180075.1	Chrysobalanoid	Chrysobalanaceae
<i>Kostermanthus robustus</i>	NC_030565.1/KX180073.1	Chrysobalanoid	Chrysobalanaceae
<i>Leptobalanus sprucei</i>	NC_024065.1/KJ414484.1	Chrysobalanoid	Chrysobalanaceae
<i>Licania alba</i>	NC_024064.1/KJ414483.1	Chrysobalanoid	Chrysobalanaceae
<i>Magnistipula butayei</i>	NC_030576.1/KX180084.1	Chrysobalanoid	Chrysobalanaceae
<i>Maranthes gabunensis</i>	NC_030577.1/KX180085.1	Chrysobalanoid	Chrysobalanaceae
<i>Moquilea minutiflora</i>	NC_030573.1/KX180081.1	Chrysobalanoid	Chrysobalanaceae
<i>Neocarya macrophylla</i>	NC_030580.1/KX180088.1	Chrysobalanoid	Chrysobalanaceae
<i>Parastemon urophyllus</i>	NC_030517.1/KX180089.1	Chrysobalanoid	Chrysobalanaceae
<i>Parinari campestris</i>	NC_024067.1/KJ414486.1	Chrysobalanoid	Chrysobalanaceae
<i>Vitis vinifera</i>	NC_007957.1		Vitaceae
