



UFG

**UNIVERSIDADE FEDERAL DE GOIÁS
ESCOLA DE AGRONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM
GENÉTICA E MELHORAMENTO DE PLANTAS**

**ANÁLISE COMPARATIVA DE SEQUÊNCIAS
COMPLETAS DE GENOMAS DE
CLOROPLASTO DE *PTERODON*
EMARGINATUS E *PTERODON PUBESCENS*
(LEGUMINOSAE)**

JULIANA BORGES PEREIRA BRITO FREITAS

Orientadora:
Prof.^a Dr.^a Thannya Nascimento Soares

Fevereiro – 2021



UNIVERSIDADE FEDERAL DE GOIÁS
ESCOLA DE AGRONOMIA

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

Juliana Borges Pereira Brito Freitas

3. Título do trabalho

ANÁLISE COMPARATIVA DE SEQUÊNCIAS COMPLETAS DE GENOMAS DE CLOROPLASTO DE PTERODON EMARGINATUS E PTERODON PUBESCENS (LEGUMINOSAE)

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.



Documento assinado eletronicamente por **Thannya Nascimento Soares, Professora do Magistério Superior**, em 12/03/2021, às 09:48, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **JULIANA BORGES PEREIRA BRITO FREITAS, Discente**, em 17/05/2021, às 12:20, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



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 **1937771** e o código CRC **05A645BB**.

JULIANA BORGES PEREIRA BRITO FREITAS

Análise comparativa de sequências completas de genomas de cloroplasto de *Pterodon emarginatus* e *Pterodon pubescens* (Leguminosae)

Dissertação apresentada ao Programa de Pós-Graduação em Genética e Melhoramento de Plantas, da Universidade Federal de Goiás, como requisito parcial para obtenção do título de Mestre em Genética e Melhoramento de Plantas.

Orientadora:

Prof.^a Dr.^a Thannya Nascimento Soares

Coorientadora:

Prof.^a Dr.^a Adriana Maria Antunes Taquary

Goiânia GO - Brasil

2021

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

Freitas, Juliana Borges Pereira Brito
ANÁLISE COMPARATIVA DE SEQUÊNCIAS COMPLETAS DE
GENOMAS DE CLOROPLASTO DE PTERODON EMARGINATUS E
PTERODON PUBESCENS (LEGUMINOSAE) [manuscrito] / Juliana
Borges Pereira Brito Freitas. - 2021.
48 f.: il.

Orientador: Profa. Dra. Thannya Nascimento Soares; co
orientadora Dra. Adriana Maria Antunes Taquary.
Dissertação (Mestrado) - Universidade Federal de Goiás, Escola
de Agronomia (EA), Programa de Pós-graduação em Genética e
Melhoramento de Plantas, Goiânia, 2021.
Bibliografia. Apêndice.

1. Anotação gênica. 2. Genoma organelar. 3. Leguminosae. 4.
Pterodon. I. Soares, Thannya Nascimento, orient. II. Título.

CDU 57



UNIVERSIDADE FEDERAL DE GOIÁS

ESCOLA DE AGRONOMIA

ATA DE DEFESA DE DISSERTAÇÃO

Ata nº **62/2021** da sessão de Defesa de Dissertação de **Juliana Borges Pereira Brito Freitas**, que confere o título de Mestra em Genética e Melhoramento de Plantas, na área de concentração em Genética e Melhoramento de Plantas.

Aos 11/03/2021, onze dias do mês de março do ano de dois mil e vinte e um, a partir das 14:00, por videoconferência, realizou-se a sessão pública de Defesa de Dissertação intitulada "ANÁLISE COMPARATIVA DE SEQUÊNCIAS COMPLETAS DE GENOMAS DE CLOROPLASTO DE PTERODON EMARGINATUS E PTERODON PUBESCENS (LEGUMINOSAE)". Os trabalhos foram instalados pela Orientadora, **Profa. Thannya Nascimento Soares - ICB/UFG** - com a participação dos demais membros da Banca Examinadora: **Prof. Alexandre Siqueira Guedes Coelho - EA/UFG**, membro titular interno; **Profa. Renata de Oliveira Dias - ICB/UFG**, membro titular externo. 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 candidata **APROVADA** pelos seus membros. Proclamados os resultados pela **Profa. Thannya Nascimento Soares**, Presidente da Banca Examinadora, foram encerrados os trabalhos e, para constar, lavrou-se a presente ata que é assinada pelos Membros da Banca Examinadora, aos 11/03/2021, onze dias do mês de março do ano de dois mil e vinte e um.

TÍTULO SUGERIDO PELA BANCA



Documento assinado eletronicamente por **Thannya Nascimento Soares, Professora do Magistério Superior**, em 12/03/2021, às 09:44, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Renata De Oliveira Dias, Professora do Magistério Superior**, em 12/03/2021, às 12:50, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Alexandre Siqueira Guedes Coelho, Professor do Magistério Superior**, em 12/03/2021, às 13:43, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



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

AGRADECIMENTO

Um sonho antigo que por muitas vezes eu mesma descreditei está se tornando real. Voltar a estudar depois de tanto tempo foi um desafio principalmente tendo que conciliar a função de estudante, mãe e profissional. Por isso, agradeço a todos que estiveram comigo durante este

período, que acreditaram e me apoiaram em todos os momentos. Agradeço primeiramente a Deus por me fortalecer e iluminar, me dando forças para seguir nesta trajetória. Agradeço à minha família que sempre acreditou e se empenhou para que eu pudesse seguir este caminho, em especial às minhas maiores inspirações, minhas filhas Isadora e Helena, que com muito carinho e compreensão sempre estiveram ao meu lado. Meus pais, irmão e avós por acreditarem em mim e ao meu esposo Luiz Carlos por seu companheirismo. Agradeço a toda a equipe do Programa de Genética e Melhoramento de Plantas e do Laboratório de Genética & Biodiversidade (Instituto de Ciências Biológicas UFG) que me auxiliaram durante estes anos de estudo e pesquisa. Ao PRONEM (Programa de Apoio a Núcleos Emergentes PRONEM/FAPEG/CNPq e o grupo de trabalho em Genética e Genômica Evolutiva, do Instituto Nacional de Ciência e Tecnologia - Ecologia, Evolução e Conservação da Biodiversidade (INCT-EECBio CNPq/FAPEG). Às minhas orientadoras professoras Thannya Nascimento Soares e Adriana Maria Antunes Taquary pelo apoio, ensinamento, orientação e paciência. Quero agradecer também aos meus colegas de trabalho do CEPMG Manoel VilaVerde, e aos meus amigos, Dhiovanna, Kleibe, Daniel Lucas, Thaís, Jefte, Ailton, Luis Gabriel, Victor e Victória, que fiz neste percurso e que foram imprescindíveis para a realização deste trabalho. Agradeço a Andrezza, Rhewter, Ramilla e Cíntia por seu importantíssimo apoio durante as análises. Agradeço também a todos os professores do programa, em especial aos professores Alexandre Siqueira, e Renata por suas excelentes aulas e colaboração.

SUMÁRIO

RESUMO	8
ABSTRACT	9
1. INTRODUÇÃO.....	10
2. REFERENCIAL TEÓRICO.....	13
2.1 Classificação taxonômica e distribuição	13
2.2 Descrição botânica	14
2.3 Importância socioeconômica e ambiental	14
2.4 Genoma de Cloroplastos	15
3. MATERIAL E MÉTODOS.....	20
3.1 Obtenção das sequências dos genomas cloroplastidiais de <i>P. pubescens</i> e <i>P. emarginatus</i>	20
3.2 Montagem dos Genomas dos Cloroplastos	21
3.3 Anotação das sequências gênicas, identificação de regiões de microssatélites e desenho de <i>primers</i>	22
3.4 Análises comparativas dos genomas dos cloroplastos	23
4. RESULTADOS	25
4.1 Os genomas cloroplastidiais de <i>P. pubescens</i> e <i>P. emarginatus</i>	26
4.2 Anotação gênica e regiões microssatélites identificadas nos genomas cloroplastidiais de <i>P. pubescens</i> e <i>P. emarginatus</i>	28
4.3 Comparação dos genomas cloroplastidiais de <i>Pterodon emarginatus</i> e <i>P. pubescens</i> com outros de Leguminosae	32
5. DISCUSSÃO.....	37
6. CONCLUSÕES.....	40
7. REFERÊNCIAS BIBLIOGRÁFICAS	41
APÊNDICE	47

RESUMO

FREITAS, J. B. P. B. **Análise comparativa de sequências completas de genomas de cloroplasto de *Pterodon emarginatus* e *Pterodon pubescens* (Leguminosae).** 2021. Dissertação (Mestrado em Genética e Melhoramento de Plantas) – Escola de Agronomia, Universidade Federal de Goiás, Goiânia, 2020.¹

O cloroplasto é uma organela celular que possui genoma próprio. Diversos projetos de sequenciamento dos genomas cloroplastídias têm sido desenvolvidos com espécies do grupo das leguminosas, família à qual pertencem as espécies *Pterodon emarginatus* e *P. pubescens*. Essas duas espécies do gênero *Pterodon* são nativas do bioma Cerrado, são popularmente conhecidas como sucupira branca ou faveira, e apresentam potencial para uso medicinal e madeireiro. O presente trabalho visou conhecer e comparar a estrutura e a organização dos genomas cloroplastídias de *P. emarginatus* e *P. pubescens*, com o intuito de colaborar com o conhecimento acerca das relações filogenéticas entre estas espécies. O sequenciamento foi realizado com a plataforma Illumina MiSeq. Os genomas foram montados por meio de estratégia *de novo* associado à montagem guiada por genoma de referência. Foi realizada a anotação de genes e das regiões repetitivas dos genomas. Além disso, foram estimadas a diversidade nucleotídica e analisadas as relações filogenéticas destas espécies com outras da família Leguminosae, cujos genomas cloroplastídias encontram-se disponíveis em bancos de dados. Estas análises de diversidade e filogenia foram realizadas com base nas sequências gênicas dos genomas cloroplastídias. A sequência completa do genoma cloroplastídial de *P. emarginatus* tem 159.823 pb, enquanto de *P. pubescens* tem 159.818 pb. Ambos os genomas possuem estrutura quadripartida, composta por duas regiões invertidas (IR) com 25.584 pb em ambas as espécies, separadas por uma região de cópia única curta (SSC), com 19.359 pb e 19.367 pb, respectivamente, para *P. pubescens* e *P. emarginatus* e uma região de cópia única longa (LSC), com 89.291 pb em *P. pubescens* e 89.288 pb em *P. emarginatus*. Foram preditos no genoma de cloroplasto, 127 genes nas duas espécies, dentre os quais, 109 genes de cópia simples e 18 duplicados nas regiões invertidas. Foram identificadas 143 regiões microssatélites em *P. emarginatus* e 141 em *P. pubescens*. O conteúdo gênico foi semelhante ao de outros genomas cloroplastídias de espécies de leguminosas, com destaque para a semelhança com *Dipteryx alata* e *Styphnolobium japonicum*, com poucas exceções de perdas e ganhos de genes. As estimativas de diversidade nucleotídica das regiões gênicas foram de 0,062 em LSC, 0,076 em SSC e 0,036 em IR. A análise filogenética demonstrou a proximidade entre os gêneros *Pterodon* e *Dipteryx*, ambos do clado Dipterygeae bem como a relação próxima entre os clados ADA e Cladrastis. Desta forma, as informações genéticas obtidas sobre os genomas cloroplastídias de *P. emarginatus* e *P. pubescens* presentes neste estudo reforçam a semelhança e proximidade evolutiva entre estas espécies. Estes resultados podem contribuir para outros estudos, tais como, sistemática molecular e genética e filogeografia de populações.

Palavras-chaves: Anotação gênica, Genoma organelar, Leguminosae, *Pterodon*.

¹ Orientadora Prof^a. Dr^a. Thannya Nascimento Soares

ABSTRACT

FREITAS, J. B. P. B. **Comparative analysis of complete chloroplast genome sequences of *Pterodon emarginatus* and *P. pubescens* (Leguminosae)**. 2020. Dissertation (Master in Genetics and Plant Breeding) – School of Agronomy, Federal University of Goiás, Goiânia, 2020.¹

The chloroplast is a cellular organelle that has its own genome. Several projects of chloroplast genome sequencing have been developed with species of the legume group, family to which the species *Pterodon emarginatus* and *P. pubescens* belong. These two species of the genus *Pterodon* are native to the Cerrado biome, are popularly known as sucupira branca or faveira, and have potential for medicinal use and lumber. The present work aimed to know and compare the structure and organization of the chloroplastial genomes of *P. emarginatus* and *P. pubescens*, in order to collaborate with the knowledge about the phylogenetic relationships between these species. Sequencing was performed using the Illumina MiSeq platform. The genomes were assembled using a de novo strategy associated with reference genome guided assembly. Annotation of genes and repetitive regions of the genomes was performed. In addition, nucleotide diversity was estimated and phylogenetic relationships of these species with other species of the Leguminosae family, whose chloroplastid genomes are available in databases, were analyzed. These diversity and phylogeny analyses were performed based on the gene sequences of the chloroplastial genomes. The complete sequence of the chloroplastial genome of *P. emarginatus* is 159,823 bp, while that of *P. pubescens* is 159,818 bp. Both genomes have a quadripartite structure, composed of two inverted regions (IR) with 25,584 bp in both species, separated by a short single copy region (SSC), with 19,359 bp and 19,367 bp, respectively, for *P. pubescens* and *P. emarginatus* and a long single copy region (LSC), with 89,291 bp in *P. pubescens* and 89,288 bp in *P. emarginatus*. A total of 127 genes were predicted in the chloroplast genome in both species, among which, 109 single copy genes and 18 duplicated genes in the inverted regions. A total of 143 microsatellite regions were identified in *P. emarginatus* and 141 in *P. pubescens*. The gene content was similar to that of other chloroplastial genomes of legume species, with an emphasis on similarity with *Dipteryx alata* and *Styphnolobium japonicum*, with few exceptions of gene losses and gains. Nucleotide diversity estimates of gene regions were 0.062 in LSC, 0.076 in SSC, and 0.036 in IR. Phylogenetic analysis demonstrated the proximity between the genera *Pterodon* and *Dipteryx*, both from the clade Dipterygeae as well as the close relationship between the clades ADA and Cladrastis. The genetic information obtained on the chloroplastial genomes of *P. emarginatus* and *P. pubescens* in this study reinforces the similarity and evolutionary proximity between these species. These results may contribute to other studies, such as molecular systematics and population genetics and phylogeography.

Keywords: Gene annotation, Organellar genome, Leguminosae, *Pterodon*.

¹ Advisor Prof^a. Dr^a. Thannya Nascimento Soares

1. INTRODUÇÃO

O conteúdo genômico das células vegetais é composto pelo genoma nuclear e pelos genomas organelares, do cloroplasto e da mitocôndria. Os genomas cloroplastidiais surgiram a partir da fagocitose de cianobactérias em eventos de endossimbiose e a herança destes genomas é, em sua maioria, uniparental (Guerrero; Margulis; Berlanga, 2013; Kolodner; Tewari, 1972). Em plantas superiores o genoma cloroplastidial possui estrutura genômica relativamente conservada, sendo geralmente dividido em quatro partes: duas repetições invertidas (IRs), separadas por uma curta região de cópia única (SSC) e uma região também de cópia única longa (LSC). O genoma cloroplastidial possui maquinaria completa para expressão de genes, cuja função está relacionada a processos essenciais para as plantas, como fotossíntese e biossíntese de ácidos graxos. Assim, desempenham um importante papel na reação das plantas quanto ao estresse biótico e abiótico (Saski et al., 2005; Daniell et al., 2016; Xiao-Ming et al., 2017 e Kersten et al., 2016).

Os processos evolutivos dos genomas de cloroplastos em geral são mais lentos e ocorrem em taxas distintas das regiões nucleares. As mutações são menos frequentes em genomas cloroplastidiais, de modo que as mudanças normalmente ocorrem devido a alterações como deleções e inserções de sequências de DNA e perda de regiões invertidas (Daniell et al., 2016; Xiao-Ming et al., 2017). Por essas razões, os estudos desenvolvidos com genomas cloroplastidiais têm sido amplamente utilizados com o intuito por exemplo, de conhecer a origem, as relações entre as espécies e evolução, como nos trabalhos de sequenciamento do genoma cloroplastidial de arroz (Wambugu et al., 2015) e de orquídeas do gênero *Oncidium* (Pan et al., 2012).

Durante muitos anos o método de Sanger foi utilizado para realização do sequenciamento de DNA (Shinozaki et al., 1986; Ohyama et al., 1986 e Souza et al., 2019). Porém as novas tecnologias de sequenciamento, chamadas de sequenciamento de nova geração (*Next Generation Sequencing* – NGS), promoveram um aumento na produção de sequências de nucleotídeos, que passaram a ser geradas com alto rendimento, baixas taxas de erro, menor tempo e reduzidos custos por base sequenciada (Mardis, 2008; Metzker, 2010; Jeon; Kim, 2019). Diversos projetos de sequenciamento têm sido desenvolvidos com

espécies do grupo das leguminosas, uma das maiores famílias de angiospermas. Até o momento, foram depositadas 190 sequências genômicas plastidiais pertencentes a subfamília dos Papilionoideae no banco de dados *GenBank*, da base NCBI (*National Center for Biotechnology Information*), dentre eles, genomas de várias espécies com interesse econômico, como por exemplo *Glycine max* e *Phaseolus vulgaris* (Taquary, 2017; Schmutz et al., 2014; Tang et al., 2014). Os avanços nas tecnologias de sequenciamento e nas ferramentas de bioinformática facilitaram também a realização de estudos do genoma de cloroplastos de diversas espécies, como as espécies do gênero *Pterodon*.

Pterodon emarginatus e *Pterodon pubescens* (Leguminosa) são espécies irmãs. Estas espécies possuem ampla distribuição geográfica no Brasil, sendo encontrados com maior predominância no bioma Cerrado, como apontam os autores Mendonça (2008) e Dutra et al. (2012), e são classificadas por Carvalho (2020) como arbóreas, xerófitas, semidecíduas e heliófitas.

Embora *P. emarginatus* e *P. pubescens* sejam espécies claramente distintas segundo à classificação botânica (Rocha, 2006), popularmente, elas são conhecidas de forma conjunta como sucupira branca ou faveira (Almeida et al., 1998). A sucupira branca possui madeira resistente, é muito utilizada para arborização, construção civil, recuperação de áreas degradadas e reflorestamento (Hansen; Haraguchi; Alonso, 2010; Santos et al., 2010; Hansen et al., 2012). Na medicina popular, é utilizado o óleo extraído de frutos de *P. emarginatus* e *P. pubescens*, devido às propriedades farmacológicas já comprovadas cientificamente. Neste sentido, diversos estudos relatam funções analgésicas, anti-inflamatórias, antitumorais, antioxidantes, angiogênicas e antimicrobianas (Hansen; Haraguchi; Alonso, 2010; Santos et al., 2010; Hansen et al., 2012; Negri; Mattei; Mendes, 2014; Pascoa et al., 2015; De Araújo et al., 2015; Bavaresco et al., 2016).

A literatura científica reporta dois aspectos interessantes para justificar a importância de se realizar estudos genômicos com as espécies *P. emarginatus* e *P. pubescens*. O primeiro está relacionado com a proximidade evolutiva destas espécies, de modo que elas são interessantes para se estudar processos de diversificação. O segundo se relaciona ao fato de elas apresentarem bom potencial para o uso medicinal e madeireiro.

Diante disso, o objetivo deste trabalho foi conhecer e comparar as estruturas dos genomas cloroplastidiais completos de *P. emarginatus* e *P. pubescens*. Além do mais, buscou-se determinar as relações filogenéticas dos genomas cloroplastidiais destas espécies com genomas cloroplastidiais de outras espécies também pertencentes à família

Leguminosae. Os objetivos específicos foram: conhecer e comparar a composição de regiões microssatélites e do conteúdo gênico dos genomas cloroplastidiais de *P. emarginatus* e *P. pubescens*; desenhar pares de *primers* que flanqueiam regiões microssatélites nesses genomas cloroplastidiais com potencial para uso em trabalhos de desenvolvimento de marcadores; avaliar a diversidade nucleotídica entre os genomas cloroplastidiais das duas espécies de *Pterodon* e outras e 20 espécies da família Leguminosae, além de determinar a relação filogenética, com base nas sequências gênicas dos genomas cloroplastidiais de *P. emarginatus* e *P. pubescens*, as 20 espécies da família Leguminosae juntamente com outras duas espécies como *outgroup*.

2. REFERENCIAL TEÓRICO

2.1 Classificação taxonômica e distribuição

Espécies do gênero *Pterodon*, pertencem à família Leguminosae, terceira maior família do grupo das angiospermas. A família das Leguminosae, são compostas por 770 gêneros e mais de 19.500 espécies, nesta família estão integradas à subfamília das Papilionoideae, clado dos Dipterygeae, subgrupo pertencente ao clado ADA, que inclui também os clados Angylocalyceae e Amburanae, e possuem ampla distribuição geográfica. Segundo o REFLORA (Carvalho, 2020), o gênero compreende quatro espécies nativas e não endêmicas do Brasil: *P. abruptus* (Moric.) Benth., *P. apparicioi* Pedersoli, *P. emarginatus* Vogel (sinônimo de *Acosmium inornatum*) (Mohlenbr.) Yakovlev, *P. polygalaeiflorus* Benth. e *Sweetia inornata* (Mohlenbr.) e *P. pubescens* (Benth.) Benth.

Neste sentido, destaca-se então que *P. pubescens* e *P. emarginatus* são espécies irmãs, popularmente conhecidas como faveira ou sucupira branca. No que diz respeito à polinização das flores, as abelhas são as principais responsáveis pela realização e a dispersão dos seus frutos ocorre por anemocoria (Almeida et al., 1998; Barroso et al., 1999; Cardoso et al., 2015; Almeida et al., 2003; Cardoso et al., 2013; Jin; Choi; Choi, 2019).

P. emarginatus e *P. pubescens* são encontradas principalmente no Cerrado brasileiro, um bioma do tipo savana considerado *hotspot* de biodiversidade. Contudo, estão presentes também na Amazônia, Caatinga e Pantanal (Mendonça, 2008; Dutra et al., 2012; Carvalho, 2020). A localização geográfica segundo o REFLORA é semelhante, considerando que *P. pubescens* é encontrada no Centro-Oeste (Distrito Federal, Goiás, Mato Grosso do Sul, Mato Grosso), Nordeste (Bahia, Maranhão, Piauí), Norte (Tocantins) e Sudeste (Minas Gerais, São Paulo), enquanto *P. emarginatus* é mais comumente encontrada no Centro-Oeste (Distrito Federal, Goiás, Mato Grosso do Sul, Mato Grosso), Nordeste (Bahia, Ceará, Maranhão, Piauí), Norte (Rondônia, Tocantins) e no Sudeste (Minas Gerais, São Paulo) (Almeida et al., 1998; Filardi et al., 2007; Dutra et al., 2012; Lima; Tozzi; Mansano, 2015; Bavaresco et al., 2016; Carvalho, 2020).

2.2 Descrição botânica

P. emarginatus e *P. pubescens* apresentam porte arbóreo com altura de até 20 metros (Rocha, 2006). Essas espécies compartilham diversas semelhanças como a presença de frutos secos, indeiscentes e alados. O ovário das flores de *Pterodon* apresentam apenas um óvulo além de possuírem endocarpo muito duro (Lima, 2019; Rocha, 2006).

Apesar de serem semelhantes, é possível notar algumas diferenças entre as espécies. Em *P. emarginatus* as flores apresentam coloração roxa, e as folhas são glabras com 4 a 10 folíolos, enquanto *P. pubescens* possui folhas pubescentes de 6 a 19 folíolos e flores rosa, como exposto na Figura 1 (Rocha, 2006; Hoscheid; Cardoso, 2015; Melo, 2018). Quanto à floração, Afonso (1997) afirma que em *P. emarginatus* ocorre primeiro, geralmente em meados de agosto. A floração em *P. pubescens* tem picos de ocorrência no mês de setembro. Rocha (2006) afirma que é comum encontrar em campo as espécies floridas simultaneamente, podendo inclusive surgir híbridos naturais.



Figura 1. Imagens das árvores e flores das espécies *Pterodon pubescens* (A e B) e *Pterodon emarginatus* (C e D). Foto: Mariana Pires de Campos Telles, 2016.

2.3 Importância socioeconômica e ambiental

P. pubescens e *P. emarginatus* possuem grande importância econômica, principalmente por seu uso medicinal. Estudos científicos comprovaram as propriedades farmacológicas destas espécies em várias funções, tais como, analgésicas, anti-inflamatórias, antitumorais, antioxidantes, angiogênicas e antimicrobianas (Hansen; Haraguchi; Alonso, 2010; Santos et al., 2010; Hansen et al., 2012; Negri; Mattei; Mendes, 2014; Pascoa et al., 2015; De Araújo et al., 2015; Bavaresco et al., 2016). Testes realizados com as duas espécies demonstraram que são consideradas seguras para o consumo e encontram-se disponíveis

comercialmente no mercado de plantas medicinais (Alves, 2012; Hansen; Haraguchi; Alonso, 2010; Bavaresco et al., 2016).

O óleo extraído dos frutos e sementes (Figura 2) de *P. pubescens* e *P. emarginatus* é utilizado no tratamento de inflamações de garganta, reumatismos, infecções respiratórias, úlceras, diabetes, gastrites, inflamações no útero e ovário, dores musculares, contra vermes intestinais, entre outras doenças (Alves, 2012; Hansen; Haraguchi; Alonso, 2010; Santos et al., 2010; Bavaresco et al., 2016). Os óleos essenciais extraídos das espécies têm sido bastante estudados e as pesquisas buscam desenvolver encapsulamento destes óleos por seu valor medicinal (Alves et al., 2014; Oliveira et al., 2016). Diversos compostos ativos são instáveis e a encapsulação tem a finalidade de prevenir a volatilização e oxidação, além de aumentar expressivamente a vida útil dos óleos essenciais para o uso na fitoterapia (Alves et al., 2014).



Figura 2. Frutos de *Pterodon emarginatus*. Foto: Mariana Pires de Campos Telles, 2016.

Além das propriedades medicinais, as árvores do gênero *Pterodon* podem ser utilizadas na recuperação de áreas degradadas e extração de madeira, pois possuem madeira pesada e com boa durabilidade. Estas espécies têm como características o crescimento lento, tolerância à baixa fertilidade do solo e à luz solar direta. Neste sentido, podem ser utilizadas para o reflorestamento. *P. emarginatus* e *P. pubescens* possuem madeira dura e considerada resistente à umidade, útil também na construção civil (Lorenzi, 2008; Hansen; Haraguchi; Alonso, 2010; Bavaresco et al., 2016).

2.4 Genoma de Cloroplastos

Genomas nucleares de plantas normalmente são grandes e complexos, o que dificulta o processo de sequenciamento e montagem do genoma completo. Muitos estudos, como o de *Stryphnodendron adstringens* por Souza et al. (2019) e *Herbaceous bamboos* por Wang et al. (2018), dentre outros, vêm sendo realizados nos últimos anos com o objetivo de

montar e caracterizar genomas organelares, principalmente genomas de cloroplastos. Os genomas organelares podem ser montados e anotados separadamente do genoma nuclear.

Genomas nucleares e mitocondriais são maiores e mais variáveis em sequências que os genomas dos cloroplastos, pois apresentam diversos elementos repetitivos. Além disso, os genomas das mitocôndrias são constantemente rearranjados, enquanto os genomas dos cloroplastos são relativamente mais conservados. As principais alterações nos genomas cloroplastidiais ocorrem, por exemplo, devido à perda de repetição invertida (IR), perda de genes e *indels*, além de ocorrerem diminuições no comprimento das regiões intergênicas e dos íntrons (Khan et al., 2019; Li; Su; Wang, 2018; Delannoy et al., 2011; Xue et al., 2019).

As células vegetais contêm em torno de 1000 a 1700 cópias dos genomas de cloroplastos. Um cloroplasto presente na folha de *Arabidopsis thaliana* pode conter de 20 a 35 cópias do DNA cloroplastidial (Zoschke; Liere; Börner, 2007). A quantidade de DNA cloroplastidial é altamente variável durante o desenvolvimento da planta. Bendich, (1987) propõe que esta grande quantidade de genomas cloroplastidiais se deve à maior necessidade de ribossomos organelares. Portanto, a discrepância no número de genomas nucleares e cloroplastidiais, surge pelas diferenças nos mecanismos pelos quais a expressão dos genes é controlada.

Os genomas de cloroplastos em plantas superiores geralmente são circulares e possuem tamanhos variados. O tamanho do DNA cloroplastidial varia de 15.553 pares de bases (bp) em *Asarum minus* (Mohanta et al., 2020) a 521.168 bp em *Floydiella terrestris* (Brouard et al., 2010). São frequentemente divididos em quatro partes: uma região de cópia única curta (SSC), uma região de cópia única longa (LSC) e duas repetições invertidas de cerca de 25 kb (IRa e IRb). Os cloroplastos possuem a maquinaria completa para expressão gênica, com genes de RNA ribossomal (rRNAs), genes de RNA transportador (tRNAs) e genes de RNA mensageiro (mRNAs) (Daniell et al., 2016; Xiao-Ming et al., 2017; Kersten et al., 2016).

Os cloroplastos possuem em média entre 120 e 130 genes essenciais. Os genes do cloroplasto podem ser de cópia única ou duplicados nas regiões invertidas, codificam proteínas, tRNAs e rRNAs, incluindo proteínas com funções que ainda não são bem definidas, tais como os genes *ycf*. A quantidade de genes codificadores de proteínas é variável entre as espécies. A maioria dos genes do cloroplasto codificam proteínas essenciais para a fotossíntese. Como as bactérias, muitos genes do cloroplasto são organizados em

operons cuja transcrição é realizada usando um ou mais promotores (Saski et al., 2005; Daniell et al., 2016; Xiao-Ming et al., 2017; Kersten et al., 2016 e Zhang et al., 2020).

A maioria dos genes cloroplastidiais não apresenta íntrons, regiões não traduzidas identificadas em genomas de eucariotos. Quanto ao número de íntrons, a maior parte dos genes contém apenas um, enquanto raros genes apresentam dois íntrons, como por exemplo, o gene *ycf3*. Os íntrons contribuem para a complexidade dos genomas, uma vez que durante os processos de expressão gênica aumentam a capacidade de codificação do genoma através do *splicing* alternativo (Graveley, 2001). O estudo das regiões intrônicas auxilia na inferência de relações filogenéticas, devido às suas sequências não codificantes apresentarem evolução rápida (Daniell et al., 2016; Kersten et al., 2016 e Zhang et al., 2020).

A estrutura do cloroplasto é relativamente conservada, principalmente nas regiões gênicas. No entanto, durante o processo padrão de evolução das plantas alguns genes foram transferidos do cloroplasto para as mitocôndrias ou para o núcleo. Os cloroplastos são considerados organelas semi-independentes, pois se encontram em estreita e constante relação com o núcleo. A expressão dos genes codificados pelo DNA cloroplastidial é dependente de um extenso grupo de fatores de origem nuclear. A maioria das proteínas do cloroplasto são codificadas pelo genoma nuclear, mas os genes cloroplastidiais também contribuem para síntese proteica (Berger, 2019; Muschner, 2005; Zhang et al., 2014; Xu et al., 2015; Yi et al., 2016). O proteoma dos cloroplastos consiste em milhares de proteínas que, além da fotossíntese, participam da biossíntese de ácidos graxos, aminoácidos, hormônios, pigmentos, vitaminas, nucleotídeos e metabólitos secundários (Saski et al., 2005; Daniell et al., 2016; e Kersten et al., 2016).

Os primeiros genomas de cloroplastos completos sequenciados, montados e anotados foram das espécies *Nicotiana tabacum* (Shinozaki et al., 1986) e *Marchantia polymorpha* (Ohyama et al., 1986), há mais de 30 anos. Estes genomas foram sequenciados usando a tecnologia Sanger. O sequenciamento feito com o método Sanger demanda muito tempo e tem um elevado custo, por isso os estudos mais recentes têm usado as tecnologias de sequenciamento NGS para estudar os genomas cloroplastidiais (Shinozaki et al., 1986; Ohyama et al., 1986 e Souza et al., 2019; Wang et al., 2018).

Com o advento das tecnologias NGS, plataformas de sequenciamento de nova geração, possibilitou-se produzir sequências de nucleotídeos com alto rendimento, com baixas taxas de erro, em um espaço curto de tempo e reduzidos custos por cada base sequenciada (Mardis, 2008; Metzker, 2010; Jeon; Kim, 2019). Embora diversas plataformas

NGS estejam disponíveis para a realização do sequenciamento do genoma dos cloroplastos, atualmente, a plataforma Illumina é a principal ferramenta NGS utilizada para este fim (Daniell et al., 2016; Cronn et al., 2008; Quail et al., 2012).

A plataforma Illumina produz uma grande quantidade de *reads* curtos em cada corrida e permite o sequenciamento com base na estratégia *paired-end* em que ambas as extremidades de cada fragmento são sequenciadas, de modo que os *reads* são formados aos pares. O processo de montagem do genoma envolve a análise e cortes de qualidade, através de ferramentas como os *softwares* FastQC e o Trimmomatic, que detectam e removem bases que apresentam baixa qualidade (Andrews, 2010; Bolger; Lohse; Usadel, 2014). Após a análise de qualidade, os *reads* podem ser mapeados em um genoma de referência e agrupados em sequências consenso chamadas de *contigs* e os *contigs* unidos formam os *scaffolds*, constituídos por *contigs* e *gaps*. Os *gaps* são trechos não sequenciados que são preditos pelos *softwares* de montagem (Miller; Koren; Sutton, 2010). Se não houver um genoma de referência disponível a montagem dos *reads* é feita pela metodologia *de novo*, usando algoritmos como os grafos de Bruijn.

Os grafos de Bruijn realizam a montagem das sequências por meio da quebra dos *reads* em segmentos menores de tamanho K (*K-mers*). Os grafos mostram diferentes caminhos ou possibilidades para a montagem da sequência. Os *k-mers* são organizados para a formação dos grafos de modo que os nós comecem sempre com o número de bases -1. Por exemplo, se o valor de *k-mer* for 7, o nó irá iniciar com as 6 últimas bases referentes ao nó anterior. Os caminhos alternativos, gerados por exemplo por erros de sequenciamento, são formados como bolhas no grafo. Os *softwares* de montagem determinam o melhor caminho para a montagem da sequência observando os que apresentam maiores coberturas de sequenciamento (Schatz; Delcher; Salzberg, 2010). A montagem do genoma tende a produzir menos *gaps* quando o sequenciamento apresenta boa cobertura e alta qualidade das sequências, o que diminui a taxa de erros na montagem (Miller; Koren; Sutton, 2010; Taquary, 2017).

A montagem do genoma do cloroplasto pode ser realizada usando a metodologia *de novo* ou pode ser guiada por referência. A montagem com referência é preferível quando existir genomas de referência com sequências de alta qualidade e de referência para a espécie em estudo. O uso do genoma de referência facilita o processo de montagem, porém apresenta como desvantagem não permitir a resolução de grandes variações estruturais na montagem (Miller; Koren; Sutton, 2010; Taquary, 2017). O uso de genomas de outras

espécies dificulta a montagem das regiões onde os genomas são diferentes. Portanto, em estudos com espécie não modelo, ou seja, que não possuem um genoma de referência, o algoritmo dos grafos *de* Bruijn é uma estratégia bastante utilizada. A montagem *de novo* pode ser feita por vários *softwares*, por exemplo, por meio de montadores como Masurca (Zimin et al., 2013), SOAPdenovo (Luo et al., 2012), SPAdes (Bankevich et al., 2012) e outros.

Os grafos *de* Bruijn se apresentam neste contexto como um algoritmo eficiente para a montagem *de novo* dos genomas (Schatz; Delcher; Salzberg, 2010). A montagem *de novo* encontra desafios quando os genomas são grandes e com extensas regiões repetitivas (Schlebusch; Illing, 2012). Trabalhos como o de Souza et al. (2019) com *S. adstringens* e Antunes et al. (2020) com *Dipteryx alata*, têm utilizado a estratégia de bioinformática de combinar a montagem *de novo* com a montagem com referência, com o intuito de remover possíveis *gaps* que possam ter surgido durante a montagem *de novo* e assim obter a montagem do genoma completo do cloroplasto.

Vários trabalhos obtiveram sucesso a partir da montagem e a caracterização de sequências do genoma cloroplastidial de espécies de leguminosas, tais como *P. vulgaris*, *G. max*, *Cicer arietinum* e *Medicago truncatula*, dentre outros (Taquary, 2017; Schmutz et al., 2010; Varshney et al., 2013; Schmutz et al., 2014; Tang et al., 2014). Neste contexto, analisar o genoma tem se apresentado de forma promissora para identificação de dados genéticos em larga escala, como por exemplo, para espécies do gênero *Pterodon* (Caicedo; Purugganan, 2005; Amar et al., 2014).

Há poucos estudos genéticos disponíveis na literatura para o gênero *Pterodon*, entre eles, está a descrição do número de cromossomos para as duas espécies, *P. pubescens* e *P. emarginatus*, que possuem $2n = 16$ cromossomos (Rocha, 2006; Dutra et al., 2012; Santana, 2014). No banco de dados *GenBank*, da base NCBI (*National Center for Biotechnology Information*) para *Pterodon* estão disponíveis apenas 46 sequências depositadas. Com isto, as estratégias de sequenciamento de DNA e suas anotações, contribuem para um melhor entendimento das características genômicas cloroplastidiais das espécies.

3. MATERIAL E MÉTODOS

3.1 Obtenção das sequências dos genomas cloroplastidiais de *P. pubescens* e *P. emarginatus*

Para a montagem do genoma cloroplastidial da espécie *P. pubescens*, foram utilizadas as sequências brutas de um indivíduo proveniente de Goiânia, GO, Brasil (espécime depositado no Herbário da Universidade Federal de Goiás - n° 61013), que teve seu genoma parcialmente sequenciado por Melo (2018). Neste estudo, o DNA total do indivíduo foi isolado utilizando o protocolo de extração proposto por Doyle; Doyle (1987). A biblioteca de sequenciamento foi preparada com o Kit *Nextera DNA Flex Library Prep* (Illumina) e o sequenciamento foi realizado na plataforma MiSeq, com o kit de 600 ciclos v3 (2x300bp, *paired-end reads*).

Para *P. emarginatus*, as sequências foram obtidas no presente estudo. Folhas jovens foram coletadas de um indivíduo adulto proveniente de Planaltina, DF, Brasil, sendo também depositado no Herbário da Universidade Federal de Goiás (n° 68.411). O DNA cloroplastidial foi extraído utilizando o kit *Chloroplast Isolation Kit* (Sigma-Aldrich). As análises qualitativa e quantitativa do DNA, foram conduzidas por eletroforese horizontal em gel de agarose e ensaio Qubit® DNA Fluorometer (Invitrogen), respectivamente. A biblioteca de DNA foi preparada seguindo o protocolo do Kit para Preparo de Biblioteca da *Agilent Technologies*, SureSelect^{QXT}. A validação das bibliotecas foi realizada com o *Bioanalyzer* (*Agilent Technologies* 2100) e PCR em tempo real (*Applied Biosystems StepOne Plus*) utilizando respectivamente, o kit *Agilent High Sensitivity DNA* e o kit *Kapa Library Quantification* (*Kapa Biosystems*). A plataforma Illumina MiSeq foi utilizada para o sequenciamento, realizado usando o kit de 600 ciclos v3 (2x300bp, *paired-end reads*).

A avaliação da qualidade dos *reads* foi realizada utilizando o *software* FastQC (Andrews, 2010) e as bases com valor de Phred <20, juntamente com os adaptadores de sequenciamento, foram removidos com o uso do *software* Trimmomatic (Bolger; Lohse; Usadel, 2014).

Os *reads* de cloroplastos de *P. pubescens* e *P. emarginatus*, foram mapeados e filtrados do DNA do genoma total através da comparação com sequências de DNA de cloroplastos de outras espécies de plantas, a partir de banco de dados no RefSeq do NCBI

GenBank com as sequências de cloroplastos depositadas de todas as espécies de angiospermas. Os *reads* foram alinhados utilizando o *software bowtie2* contra este banco de dados (Langmead et al., 2009) e todos os *reads* mapeados foram separados para a posterior montagem das sequências de cloroplastos de *P. pubescens* e *P. emarginatus*.

3.2 Montagem dos Genomas dos Cloroplastos

A montagem dos genomas dos cloroplastos foi feita utilizando a associação de montagem *de novo* e montagem guiada por genoma de referência. Para a montagem *de novo*, foi utilizado o *pipeline* Fast-Plast, através de uma abordagem combinando o método baseado em grafos de Bruijn do *software* Spades 3.11.1 para a montagem das sequências dos cloroplastos (Bankevich et al., 2012). Sequências com cobertura abaixo de 10x foram removidos da montagem. Sequências com alta cobertura foram analisadas com *Blast* (*Basic Local Alignment Search Tool*) (Altschul et al., 1990), com o intuito de identificar as espécies que apresentam genoma de cloroplasto mais semelhante para uso na montagem guiada por referência. Com o uso do *software* Yasra, foi realizada a montagem guiada por referência (Ratan, 2009), usando o genoma de *Indigofera tinctoria* (KJ468098.1), disponível no Genbank.

A montagem guiada por referência foi usada para ordenar e orientar os *scaffolds* gerados pela montagem *de novo* e para preencher os *gaps* entre esses *scaffolds*, usando o *software* BWA (Li; Durbin, 2009). As sequências consenso do genoma dos cloroplastos de *P. pubescens* e *P. emarginatus* foram obtidas com o CLC Genomics Workbench 12 (CLCBio, Arhus Denmark). Os *reads* do genoma total foram alinhados às sequências consenso obtidas com a junção da montagem *de novo* e a montagem guiada por referência, com o objetivo de encontrar sequências para preencher quaisquer lacunas entre os *scaffolds*. A qualidade da montagem final foi avaliada com o *script* *Assemblathon*.

Após o processo de montagem, os genomas completos de *P. pubescens* e *P. emarginatus* foram alinhados. O alinhamento apresentou 4 trechos, com 150, 361, 812 e 387 pares de base em *P. emarginatus* que não estavam presentes na montagem de *P. pubescens*. *Primers* foram desenhados para amplificar essas regiões e Reações em Cadeia da Polimerase (PCRs) foram realizadas e confirmadas através de eletroforese horizontal em gel de agarose (Apêndice) para averiguar a existência destes trechos. Essa estratégia confirmou que as regiões existiam e não haviam sido montadas. Uma maneira de identificar estas regiões seria sequenciar usando a técnica de Sanger a partir dos produtos de amplificação. No entanto,

com a paralização dos trabalhos no laboratório devido a doença Covid-19, foi necessária uma nova estratégia para a montagem destas regiões. As sequências das regiões ausentes em *P. pubescens* foram extraídas do genoma cloroplastidial de *P. emarginatus* e alinhadas com a utilização do *software Bowtie2* contra os *reads* produzidos no sequenciamento de *P. pubescens*, obtendo assim as sequências consenso. Deste modo, como os genomas das duas espécies são semelhantes o uso de trechos do genoma de *P. emarginatus* como referência permitiu montar as sequências das regiões de *P. pubescens*.

3.3 Anotação das sequências gênicas, identificação de regiões de microssatélites e desenho de *primers*

A anotação de sequências codantes (CDS), RNAs ribossomais (rRNAs) e RNAs transportadores (tRNAs), presentes nas sequências cloroplastidiais de *P. pubescens* e *P. emarginatus*, foi realizada nos programas *Dual Organellar GenoMe Annotator* (DOGMA) (Wyman; Jansen; Boore, 2004) e *GeSeq* (*Annotation of Organellar Genomes*) (Tillich et al., 2017). A presença de genes de tRNA e rRNA foi confirmada com os programas Aragorn, tRNAscan-SE e RNAmmer (Lowe; Eddy, 1997; Lagesen et al., 2007). O programa OGDRAW (Draw Organelle Genome Maps) (<http://ogdraw.mpimp-golm.mpg.de/>), foi utilizado para construir gráficos representativos dos genomas dos cloroplastos circulares (Lohse et al., 2014).

O programa MISA (*MicroSatellite identification tool*) foi usado para realizar a identificação de regiões microssatélites nas sequências de cloroplastos (Beier et al., 2017). Os parâmetros para a busca foram: mono e dinucleotídeos ≥ 10 bases (mínimo de dez e cinco repetições em tandem respectivamente) e trinucleotídeos e tetranucleotídeos ≥ 12 bases (mínimo de quatro e três repetições respectivamente), pentanucleotídeo ≥ 15 bases (mínimo de três repetições) e hexanucleotídeos ≥ 18 bases (mínimo de 3 repetições). A análise feita com o *repeat masker* teve a função de identificar elementos repetitivos, de sequências de baixa complexidade e repetições intercaladas.

A localização e o tamanho das repetições (*forward*, *reverse*, complementar e palindrômicas) foram feitas com o *software* REPuter. O programa Primer3 foi utilizado para o desenho dos *primers* (Untergasser et al., 2012). Os parâmetros que delinearam o desenho de *primers* para amplificação das regiões microssatélites cloroplastidiais são 150-400pb para o tamanho do produto da amplificação, 30-60% de percentagem GC, 56-62 °C de valor de temperatura de *melting* (T_m), e tamanhos dos *primers* entre 18-27 pb. Evitou-se a escolha

de regiões microssatélites mononucleotídeas e dinucleotídeas com motivo AT, pois apesar de serem frequentes no genoma podem gerar estruturas em forma de grampos, diminuindo a eficácia da amplificação por PCR.

3.4 Análises comparativas dos genomas dos cloroplastos

As análises comparativas dos genomas completos de cloroplastos de *P. pubescens* e *P. emarginatus* foram realizadas utilizando o programa mVISTA. As duas espécies foram comparadas entre si e com três outras espécies: *D. alata*, *Styphnolobium japonicum* e *Lupinus luteus*. O genoma cloroplastidial de *P. emarginatus* foi utilizado como genoma de referência e a análise foi realizada utilizando o modo Shuffle-LAGAN (Frazer et al., 2004).

A análise comparativa de diversidade de nucleotídeos utilizou as sequências gênicas de *P. pubescens*, *P. emarginatus* e de outras vinte espécies de leguminosas, todas da família Fabaceae subfamília Papilionoideae, distribuídas em diferentes clados. As referidas espécies foram utilizadas também para a análise filogenética juntamente com mais duas espécies de outras famílias como *outgroups*. As espécies utilizadas como *outgroups* foram, *Cucumis sativa* pertencente à família Cucurbitaceae e *Fragaria Vesca* da família Rosaceae (Tabela 1). Todas as sequências estão disponíveis no bando de dados do NCBI.

Tabela 1: Espécies de Leguminosae e grupos externos utilizados para a análise filogenética.

Classificação	Tribo	Espécie	Acesso do NCBI
Clado ADA	Dipterygeae	<i>Dipteryx alata</i>	MT119080
	Dipterygeae	<i>Pterodon emarginatus</i>	Presente estudo

	Dipterygeae	<i>Pterodon pubescens</i>	Presente estudo
	Angylocalyceae	<i>Angylocalyx braunii</i>	NC_047392.1
	Amburaneae	<i>Dussia macrophyllata</i>	NC_047341.1
Clado Cladrastis		<i>Styphnolobium japonicum</i>	KY872756.1
Clado Dalbergioides lato senso	Dalbergieae	<i>Arachis hypogaea</i>	NC_037358.1
	Dalbergieae	<i>Dalbergia hainanensis</i>	NC_036961.1
	Dalbergieae	<i>Stylosanthes hamata</i>	NC_039159.1
Clado Genistoides lato senso	Genisteae	<i>Lupinus luteus</i>	NC_023090.1
	Sophoreae	<i>Maackia floribunda</i>	NC_034774.1
	Ormosieae	<i>Ormosia hosiei</i>	NC_039418.1
Clado NPAAA	Phaseoleae	<i>Cajanus cajan</i>	NC_031429.1
	Cicereae	<i>Cicer arietinum</i>	NC_011163.1
	Phaseoleae	<i>Glycine max</i>	NC_007942.1
	Indigofereae	<i>Indigofera tinctoria</i>	NC_026680.1
	Loteae	<i>Lotus japonicus</i>	NC_002694.1
	Trifolieae	<i>Medicago truncatula</i>	NC_003119.8
	Phaseoleae	<i>Mucuna macrocarpa</i>	NC_044116.1
	Phaseoleae	<i>Phaseolus vulgaris</i>	NC_009259.1
	Trifolieae	<i>Trifolium subterraneum</i>	NC_011828.1
	Phaseoleae	<i>Vigna radiata</i>	NC_013843.1
<i>Outgroups</i>	Benincaseae	<i>Cucumis sativus</i>	DQ119058.1
	Potentilleae	<i>Fragaria vesca</i>	JF345175.1

O *software* ClustalW foi utilizado para alinhar as seqüências gênicas (Thompson; Higgins; Gibson, 1994). Os genes foram concatenados com a utilização do *software* MEGA X (Kumar et al., 2018). Para estimar a diversidade de nucleotídeos das seqüências de genes dos cloroplastos foi utilizado o *software* DnaSP6 (Rozas et al., 2017), a análise foi realizada individualmente cada gene.

Seqüências gênicas codificadoras de proteínas do cloroplasto de *P. emarginatus* e *P. pubescens* foram investigadas e comparadas entre si, bem como, com as seqüências das outras três espécies do Clado ADA disponíveis em banco de dados. Foram extraídas para análise entre as espécies de *Pterodon* e entre as cinco espécies do clado ADA, respectivamente 73 e 65 genes codificadores de proteínas, todos alinhados separadamente usando MAFFT. Com o uso do *software* DnaSP6, foi calculada a razão Ka/Ks para estimar as taxas de substituição sinônimas (Ks) e não sinônimas (Ka) dos genes. A razão Ka / Ks pode indicar se um determinado gene codificador de proteína está sofrendo pressão de seleção. A substituição sinônima não altera o aminoácido dentro de uma cadeia peptídica,

enquanto a substituição não sinônima muda a sequência em uma cadeia. Se o valor obtido na análise for igual a 1 é um indicativo de seleção neutra. Um valor de $Ka / Ks > 1$ está relacionado com a seleção positiva do gene, contudo $Ka / Ks < 1$ indica que o gene está sendo afetado pela seleção negativa ou seleção purificadora (Yang, 2003).

A matriz concatenada dos genes de cloroplastos foi usada para análise filogenética, realizada no programa MEGA X (Kumar et al., 2018). A árvore filogenética foi construída pelo método de máxima verossimilhança com base no modelo Tamura-Nei. Para avaliar o suporte dos nós foi utilizado o método *bootstrap* com 1000 replicatas. O *software* Figtree v. 1.4.3 (<http://tree.bio.ed.ac.uk/software/figtree/>) foi utilizado para a edição da árvore filogenética.

4. RESULTADOS

4.1 Os genomas cloroplastidiais de *P. pubescens* e *P. emarginatus*

Para a montagem dos genomas cloroplastidiais foram utilizados 369.732 *reads* para *P. pubescens* e 186.576 *reads* para *P. emarginatus*. *P. pubescens* apresentou cobertura média de 395, enquanto *P. emarginatus* uma média de 178. O tamanho médio dos insertos foram de 240 pb e 303 pb para *P. pubescens* e *P. emarginatus*, respectivamente.

Os tamanhos dos genomas cloroplastidiais das espécies são bem próximos, sendo que o de *P. emarginatus* é ligeiramente maior (159.823 pb) do que o de *P. pubescens* (159.818 pb), com uma diferença de 5 pb apenas. Ambos os genomas cloroplastidiais possuem estrutura quadripartida composta por duas regiões invertidas (IR) com 25.584 pb, semelhantes nas duas espécies, separadas por uma região de cópia única longa (LSC) com 89.291 pb em *P. pubescens* e 89.288 pb em *P. emarginatus* e uma região de cópia única curta (SSC) apresentando 19.359 pb e 19.367 pb para *P. pubescens* e *P. emarginatus*, respectivamente (Tabela 2).

Tabela 2. Estatísticas das montagens dos genomas cloroplastidiais de *Pterodon emarginatus* e *P. pubescens*. Entre parênteses estão os números de genes de cópia única.

Características do genoma	<i>P. pubescens</i>	<i>P. emarginatus</i>
Tamanho do genoma (pb)	159.818	159.823
Tamanho do LSC (pb)	89.291	89.288
Tamanho do SSC (pb)	19.359	19.367
Tamanho do IR (pb)	25.584	25.584
Número de genes (cópia única)	127 (109)	127 (109)
Genes codificador de proteína (cópia única)	82(75)	82(75)
Genes tRNA (cópia única)	37(30)	37(30)
Genes rRNA (cópia única)	8(4)	8(4)
Genes duplicados no IR	18	18
Conteúdo GC (%)	34,86	34,84
Conteúdo GC no LSC (%)	32,28	32,25
Conteúdo GC no SSC (%)	28,95	28,91
Conteúdo GC no IR (%)	41,61	41,61
Total de <i>reads</i>	369.732	186.576
Cobertura média	395	178
Tamanho médio do inserto (pb)	240	303

O conteúdo GC foi de 35% do genoma cloroplastidial total, 42% nas regiões IR e 32% e 29% nas regiões LSC e SSC, respectivamente (Tabela 2). A montagem *de novo* associada à montagem com referência gerou as sequências consenso, em que na imagem (Figura 3A e 3B) ressaltam a posição e a direção dos genes distribuídos nestas quatro regiões.

A

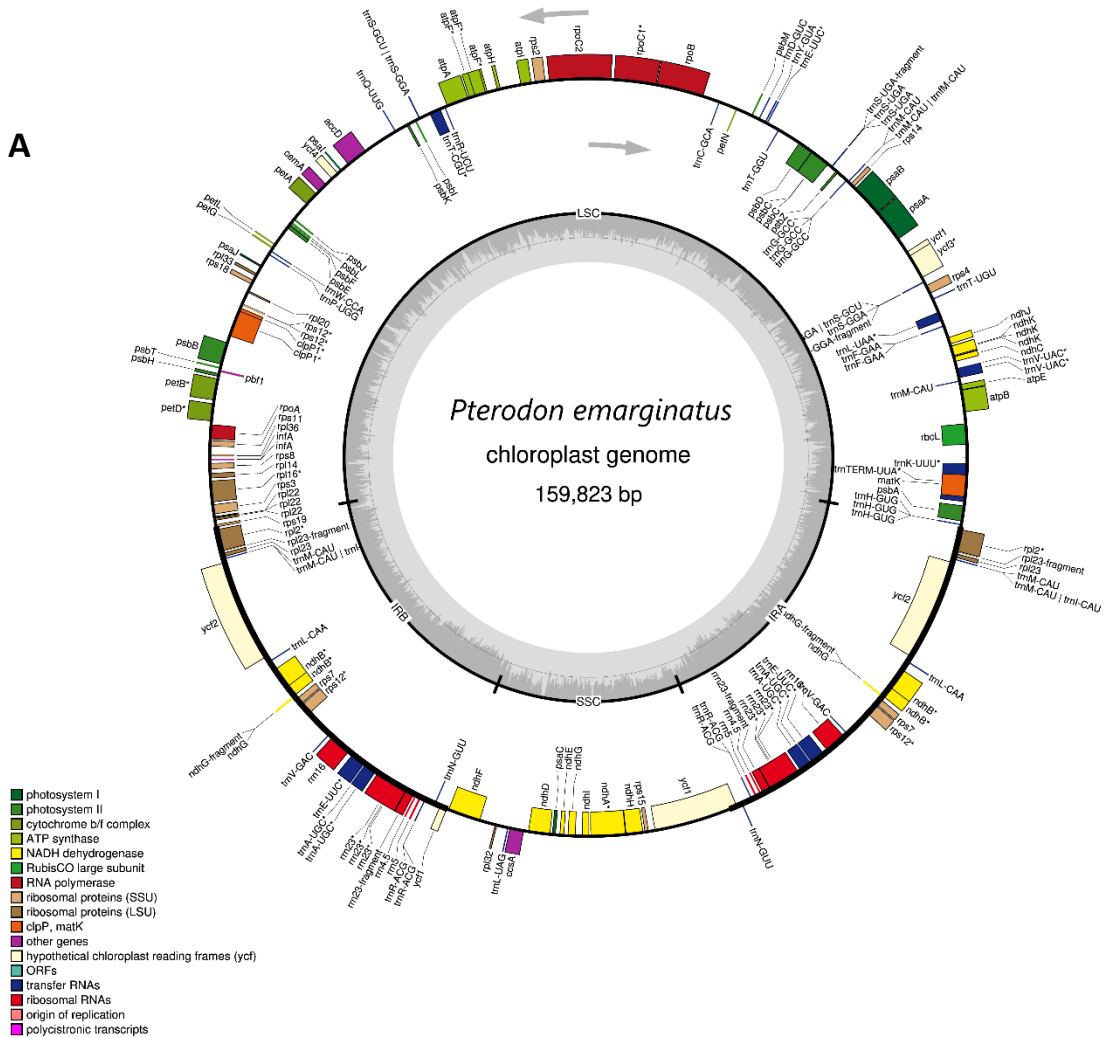


Figura 3A. Mapas circulares dos genomas cloroplastidiais de *Pterodon emarginatus*. Os genes que aparecem na parte externa da figura são transcritos no sentido antihorário, enquanto os genes na parte interna são transcritos no sentido horário. Apresenta a cópia única longa (LSC), as regiões invertidas (IRs) e cópia única curta (SSC). O conteúdo do GC é representado em cinza no interior da figura. A cor da legenda indica a função genética do gene.

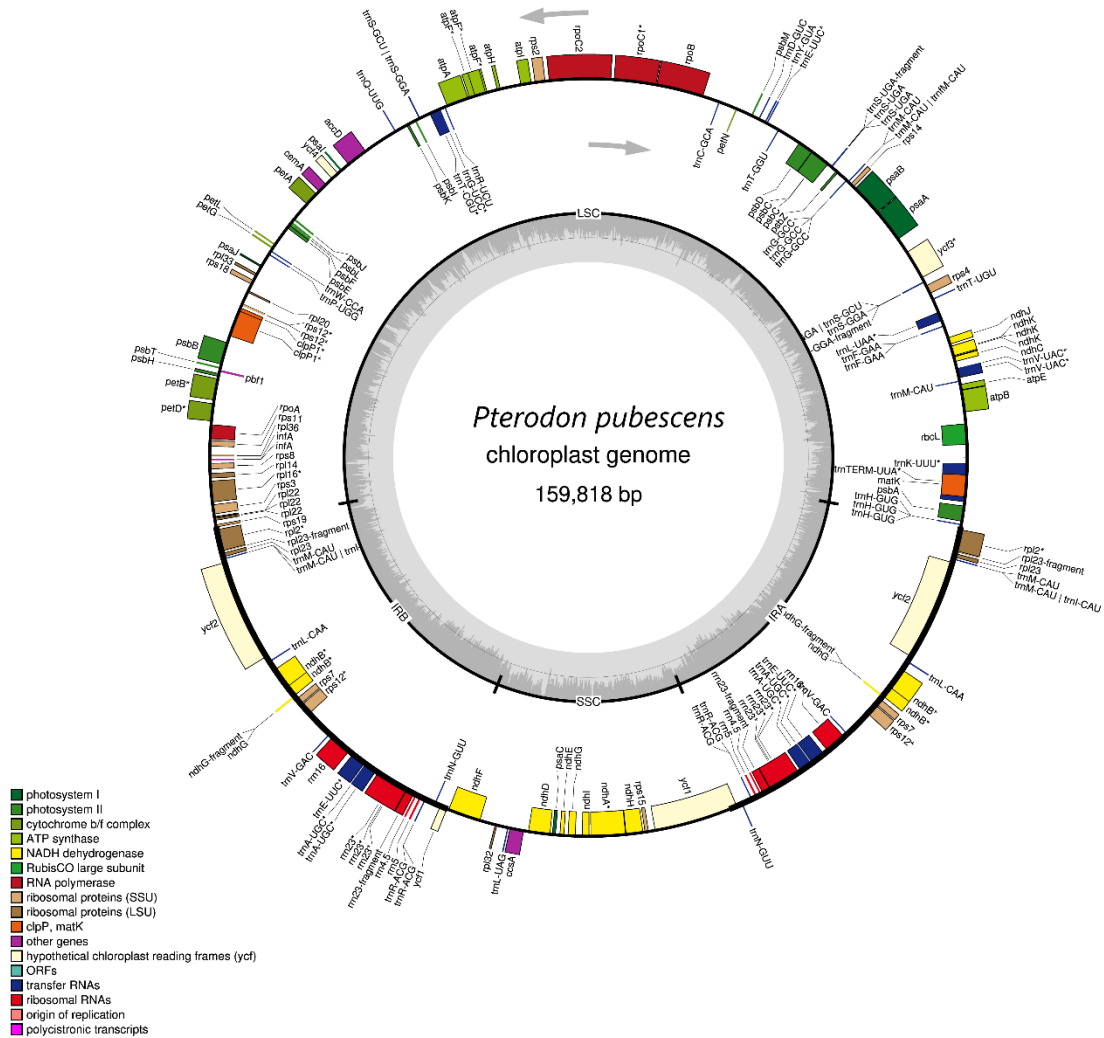
B

Figura 3B. Mapas circulares dos genomas cloroplastidiais de *Pterodon pubescens*. Os genes que aparecem na parte externa da figura são transcritos no sentido antihorário, enquanto os genes na parte interna são transcritos no sentido horário. Apresenta a cópia única longa (LSC), as regiões invertidas (IRs) e cópia única curta (SSC). O conteúdo do GC é representado em cinza no interior da figura. A cor da legenda indica a função genética do gene.

4.2 Anotação gênica e regiões microssatélites identificadas nos genomas cloroplastidiais de *P. pubescens* e *P. emarginatus*

Foram preditos 127 genes nos cloroplastos das duas espécies, dentre os quais 109 são de cópia única e 18 duplicados. Os genes preditos estão envolvidos em diversas vias metabólicas como fotossíntese, auto-replicação e biossíntese. Todos os genes exibem alta

similaridade de sequências de nucleotídeos e posição nos genomas dos cloroplastos de *P. emarginatus* e *P. pubescens*, ou seja, alta sintenia. Na porção SSC foram verificados 12 genes, sendo apenas um tRNA (*trnL-UAG*) e o restante genes codificadores de proteínas, já a região LSC apresenta 78 genes diferentes no total, incluindo 21 tRNAs (Tabela 3).

Tabela 3. Genes anotados, com suas respectivas categorias funcionais putativas, nos genomas de cloroplastos de *Pterodon emarginatus* e *P. pubescens*.

Categoria	Grupo de gene	Nome do gene	
Fotossíntese	Fotossistema I	<i>psaA, psaB, psaC, psaI, psaJ</i>	
	Fotossistema II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>	
	Desidrogenase NADH	<i>ndhA, ndhB(x2), ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>	
	Complexo citocromo b / f	<i>petA, petB, petD, petG, petL, petN</i>	
	Síntese de ATP	<i>atpA, atpB, atpE, atpF, atpH, atpI</i>	
	Subunidade grande rubisco	<i>rbcL</i>	
Autorrepliação	RNA polimerase	<i>rpoA, rpoB, rpoC1, rpoC2</i>	
	Proteína da subunidade ribossômica grande	<i>rpl2(x2), rpl14, rpl16, rpl20, rpl23(x2), rpl32, rpl33, rpl36</i>	
	Proteína da subunidade ribossômica pequena	<i>rps2, rps3, rps4, rps7(x2), rps8, rps11, rps12(x2), rps14, rps15, rps18</i>	
	RNAs ribossomais	<i>rrn5(x2), rrn16(x2), rrn23(x2), rrn4.5(x2)</i> <i>trnA-UGC(x2), trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnFM-CAU, trnG-UCC, trnH-GUG,</i>	
	RNAs transferases		<i>trnI-CAU(x2), trnI-GAU(x2), trnK-UUU, trnL-CAA(x2), trnL-UAA, trnL-UAG, trnM-CAU,</i> <i>trnN-GUU(x2), trnP-UGG, trnP-GGG, trnQ-UUG, trnR-ACG(x2), trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA,</i> <i>trnT-UGU, trnT-GGU, trnV-GAC(x2) trnV-UAC, trnW-CCA, trnY-GUA</i>
Biosíntese	Maturase	<i>matK</i>	
	Protease	<i>clpP</i>	
	Proteína membrana de envelope	<i>cemA</i>	
	Carboxilase Acetil-CoA	<i>accD</i>	
	Gene de síntese citocromo tipo C	<i>Ccsa</i>	
Outros		<i>ycf1(x2), ycf2(x2), ycf3, ycf4</i>	

(x2) equivale aos genes que se encontram duplicados nas regiões Invertidas.

Na região IR foram anotados dezoito genes duplicados, dentre eles estão os quatro genes de rRNAs: *rrn5, rrn16, rrn23 e rrn4.5*, os genes *rpl2, rpl23, rps7 e rps 14*,

ndhB, que são codificadores de proteínas, o gene *ycf2* e um fragmento do gene *yf1* e, além de sete genes tRNAs (*trnA-UGC*, *trnI-CAU*, *trnI-GAU*, *trnL-CAA*, *trnN-GUUt*, *trnR-ACG* e *trnV-GAC*).

Foram identificadas as regiões intrônicas nos genes *trnV-UAC*, *trnL-UAA*, *ycf3*, *ropC1*, *atpF*, *clpP*, *rpl2*, *ndhB*, *trnA-UGC*, *trnI-GAU* e *ndhA*. Apenas *ycf3* e *clpP* apresentaram 2 íntrons, enquanto todos os outros possuem um. O gene *trnL-UAA* expressou o menor íntron com 539 pb e o maior íntron foi identificado em *ndhA*, com 1290 pb com uma distância média de 710 pb entre os éxons (Tabela 4). Quanto à posição dos genes com íntrons no genoma, observou-se que apenas um gene (*ndhA*) está alocado na região SSC, seis estão em LSC (*trnV-UAC*, *trnL-UAA*, *ycf3*, *rpoC1*, *atpF* e *clpP*), e quatro dispostos nas regiões IR (*rpl2*, *ndhB*, *trnA-UGC* e *trnI-GAU*).

Tabela 4. Informação do conteúdo intrônico dos genes nos genomas cloroplastidiais das espécies de *Pterodon emarginatus* e *P. pubescens*, com a localização no genoma e o comprimento dos éxons e íntrons em pares de bases.

<i>P. emarginatus</i>	Região	Éxon I	Íntron I	Éxon II	Íntron II	Éxon III	<i>P. pubescens</i>	Região	Éxon I	Íntron I	Éxon II	Íntron II	Éxon III
<i>trnV-UAC</i>	LSC	38	590	36			<i>trnV-UAC</i>	LSC	38	590	36		
<i>trnL-UAA</i>	LSC	49	539	36			<i>trnL-UAA</i>	LSC	49	539	36		
<i>ycf3</i>	LSC	125	744	227	759	152	<i>ycf3</i>	LSC	125	744	227	759	152
<i>rpoC1</i>	LSC	431	815	1616			<i>rpoC1</i>	LSC	431	814	1617		
<i>atpF</i>	LSC	143	728	415			<i>atpF</i>	LSC	143	729	415		
<i>clpP</i>	LSC	227	629	290	815	68	<i>clpP</i>	LSC	227	629	290	815	68
<i>rpl2</i>	IR	434	666	389			<i>rpl2</i>	IR	434	666	389		
<i>ndhB</i>	IR	755	686	722			<i>ndhB</i>	IR	755	686	722		
<i>trnI-GAU</i>	IR	41	939	34			<i>trnI-GAU</i>	IR	41	939	34		
<i>trnA-UGC</i>	IR	37	801	34			<i>trnA-UGC</i>	IR	37	801	34		
<i>ndhA</i>	SSC	551	1290	539			<i>ndhA</i>	SSC	551	1290	539		

Na Tabela 5 é possível observar as repetições do tipo microssatélite, separadas por tipo de motivo de repetição. No geral, as regiões de microssatélites cloroplastidiais são similares e conservadas entre as duas espécies, com 141 regiões em *P. emarginatus* e 140 em *P. pubescens*. O tipo de microssatélite cloroplastidial mais comum nas duas espécies é o mononucleotídeo, sendo aproximadamente 75% com repetição da base A. Os números de microssatélites dos tipos tetra, penta e hexanucleotídeo foram iguais em ambas as espécies.

Foram identificadas nas duas espécies, 17 regiões microssatélites presentes em sequências gênicas dos cloroplastos, sendo eles: *atpB*, *atpF*, *clpP*, *ndhA*, *ndhF*, *psbC*, *rpoB*, *rpoC1*, *rpoC2*, *trnL-UAA* e *ycf3*.

Tabela 5. Comparação dos diferentes tipos e números de regiões microssatélites para cada motivo de repetição nos genomas cloroplastidiais.

Tipos de motivos	<i>Pterodon emarginatus</i>	<i>Pterodon pubescens</i>
Mononucleotídeos	82	79
Dinucleotídeos	39	40
Trinucleotídeos	10	11
Tetranucleotídeos	6	6
Pentanucleotídeos	3	3
Hexanucleotídeos	1	1
Total no genoma completo	141	140
Quantidade nas regiões codantes	17	17

Utilizando o programa *repeat masker* não foram identificadas repetições do tipo satélite e elementos transponíveis. Enquanto o *software* REputer, analisando os 50 melhores resultados, revelou os mesmos valores para as duas espécies na região IR, para as repetições do tipo *forward*, *reverse*, palíndromo e complementar, corroborando com os estudos que afirmam ser mais conservadas. Os valores anotados para as regiões LSC e SSC foram próximos como está demonstrado na Tabela 6.

Tabela 6. Repetições do tipo *Forward*, *Reverse*, Palindrômica e Complementar identificada com a utilização do *software* REputer para *Pterodon emarginatus* e *P. pubescens*.

Repetições	<i>P. emarginatus</i>			<i>P. pubescens</i>		
	IR	LSC	SSC	IR	LSC	SSC
<i>Forward</i>	25	17	15	25	18	16
<i>Reverse</i>	15	23	22	15	25	21
Palíndromo	9	8	11	9	7	11
Complementar	1	2	2	1	0	2

Com base nas regiões microssatélites identificadas e nos critérios estabelecidos, foram desenhados 10 pares de *primers* para amplificar as regiões microssatélites cloroplastidiais nas duas espécies de *Pterodon* (Tabela 7).

Tabela 7. *Primers* com potencial para desenvolvimento de marcadores moleculares.

Espécie	Motivo	Posição no genoma	Número de repetições	Sequência dos primers	Tamanho do primer	Tm	Tamanho do produto da PCR
<i>P. emarginatus</i>	CT	3696-3706	6	Forward TGCGATACAGTCAAAACAAGG	21	58.83	179
				Reverse AGAGCGATTGGGTTGAGAAA	22	59.81	
<i>P. emarginatus</i>	CT	8170-8179	5	Forward CAAAAACAATGTCTCCTTGCAT	22	59.13	183
				Reverse GAGCGGGATCTACTTTTTGG	20	58.79	
<i>P. emarginatus</i>	GAT	1047-1057	3	Forward AGGGCATATCGGTTGAAGTAGA	22	59.99	155
				Reverse CATGCGGTA CTTCATTGTGC	20	60.14	
<i>P. emarginatus</i>	AATC	3883-3893	3	Forward CACATTTATCATCACCCATAA	22	57.26	165
				Reverse AAGCAACGAGCTTTCATTTTT	21	58.2	
<i>P. emarginatus</i>	TTTCT	14719-14733	3	Forward GGTTTCCATACCAAGGCTCA	20	59.93	159
				Reverse TGAATTCAACAGTTCGGCATA	21	59.17	
<i>P. pubescens</i>	AG	24149-24159	6	Forward TGAAATCCCTTCTCTCCCACT	21	60.06	150
				Reverse CCGTCTCCACTGGATCTGTT	20	60.11	
<i>P. pubescens</i>	CA	10852-10861	5	Forward GGGTATCCTGAGCAATTTCAA	21	59.03	162
				Reverse TGA CTGGACGAAACCAAGAA	20	59.26	
<i>P. pubescens</i>	CAG	969-980	4	Forward GAGGCATACCATCAGAAAAGC	21	58.8	151
				Reverse GCTTGTTACATGGGTCTGTA	20	59.57	
<i>P. pubescens</i>	AAGA	10568-10578	3	Forward TGGTAGTGCTCTTAGATGGGAAT	23	59.17	153
				Reverse AAGACCCAACGTGCATTTTT	20	59.48	
<i>P. pubescens</i>	CTTTA	14027-14040	3	Forward TTCTCGACGTTTTATGGGAAG	21	59.21	213
				Reverse TGGCCTCTATGGGCTTTCT	19	59.78	

4.3 Comparação dos genomas cloroplastidiais de *Pterodon emarginatus* e *P. pubescens* com outros de Leguminosae

A diversidade nucleotídica para as regiões codantes foi a seguinte: em LSC 0,062, SSC 0,086 e na região IR 0,036 (Figura 4). A região IR apresenta menor diversidade. Os picos de genes com maiores diversidades foram: *trnG-UCC*, *trnK-UUU* e *ycf4* presentes na região LSC com valores de 0,23, 0,22 e 0,16, e na região SSC os genes, *rpl32* apresentando 0,15 de diversidade e *ycf1* com diversidade de 0,2, sendo que este gene apresenta uma parte na região IR.

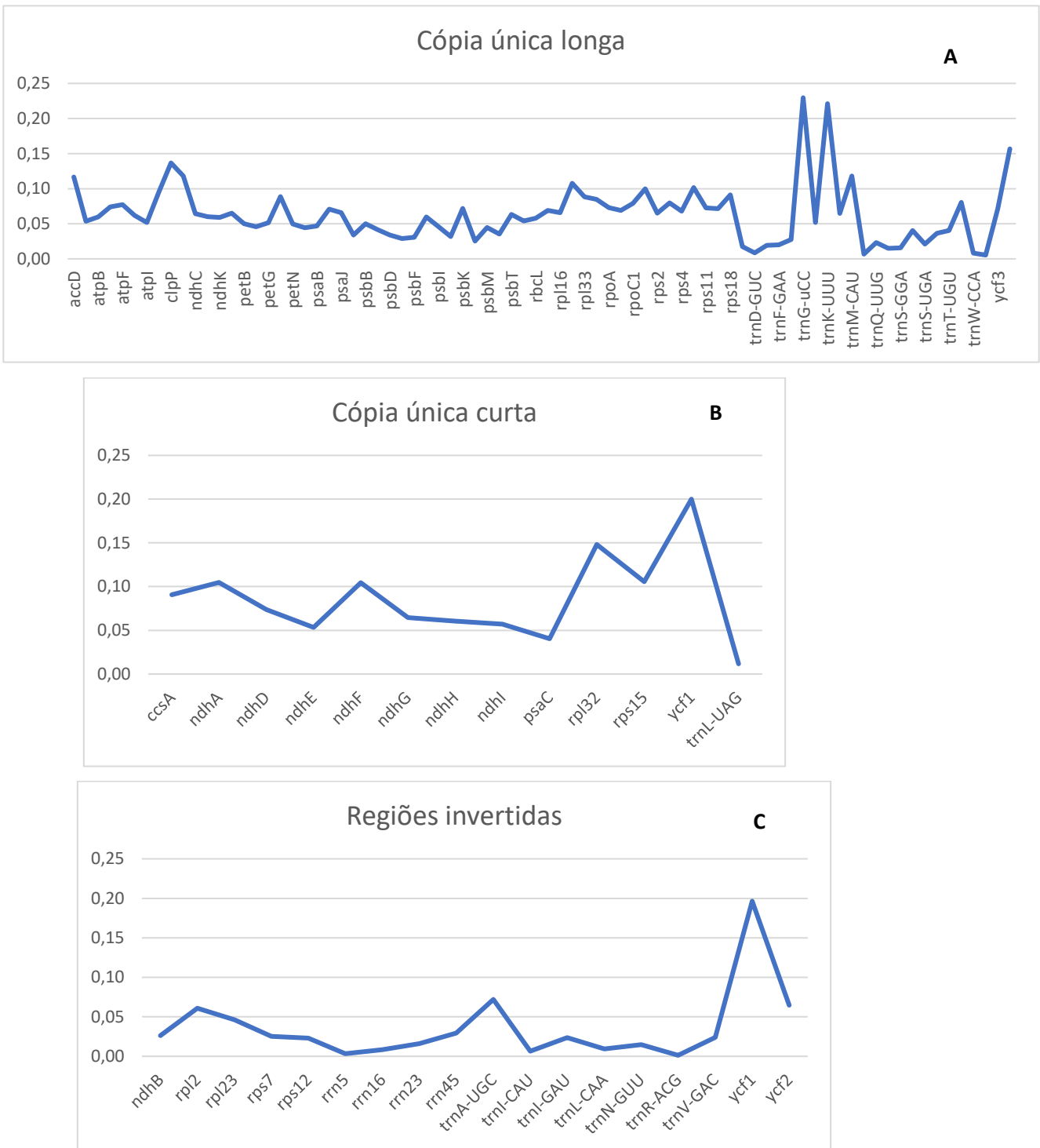
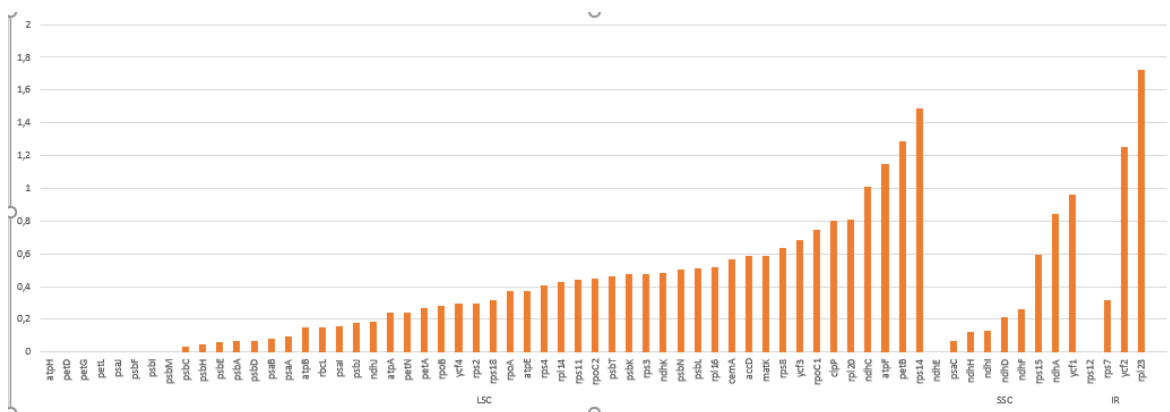


Figura 4. Comparação da diversidade nucleotídica nas regiões de cópia única longa (A), cópia única curta (B) e regiões invertidas (C) entre vinte espécies da família Leguminosae.

A razão (K_a / K_s) entre as taxas de substituições não sinônimas (K_a) para sinônimas (K_s) não identificou nenhum gene com valor igual ou superior a 1 entre *P. emarginatus* e *P. pubescens*. Os genes *AtpA*, *psbC*, *rbcL*, *rpoC2* encontrados na região LSC,

ndhF e *ycf1* na região SSC, apresentaram valores inferiores a 1, o que indica seleção negativa ou purificadora.

A análise de Ka/Ks feita para as cinco espécies do clado ADA apresentou entre os 65 genes que codificam proteínas, 48 com seleção negativa ou purificadora. Para 6 genes, a razão Ka / Ks indicou seleção positiva, são eles, os genes *ndhC*, *atpF*, *petB*, *rps14*, *ycf2* e *rpl23* (Figura 5).



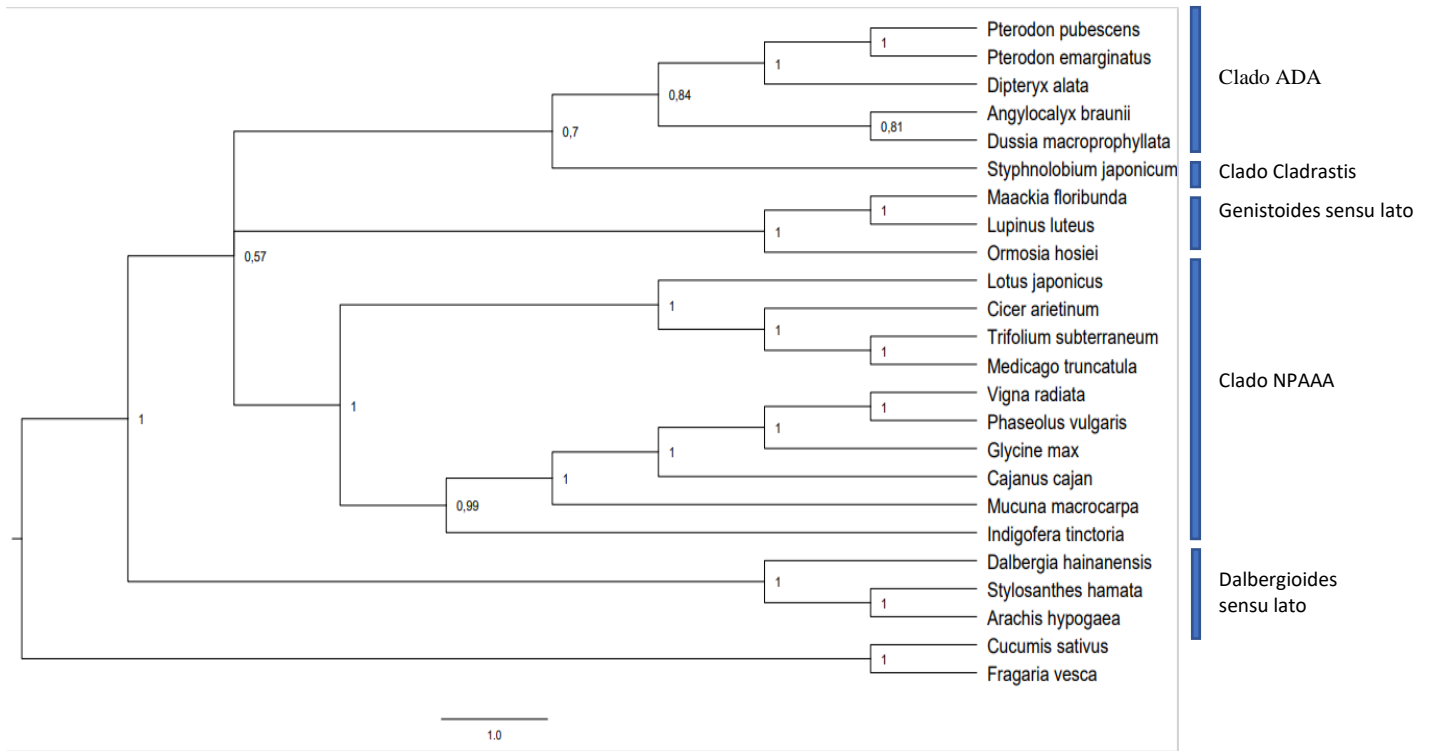


Figura 7. Árvore filogenética construída com o método de máxima verossimilhança, com 1000 replicatas para suporte *bootstrap*.

5. DISCUSSÃO

A montagem dos genomas cloroplastidiais das duas espécies de *Pterodon* apresentou estrutura típica comum às Angiospermas. Estrutura quadripartida, com uma região longa, outra curta separada por duas regiões invertidas (Daniell et al., 2016; Xiao-Ming et al., 2017). O tamanho desses genomas foi similar entre as espécies estudadas, *P. emarginatus* (159.823pb) e *P. pubescens* (159.818 pb), e se aproximam dos tamanhos dos genomas cloroplastidiais de outras espécies de Leguminosae, tais como *D. alata* com 158,6 kb (Antunes et al., 2020), *S. japonicum* com 158,8 kb (Lu et al., 2018) e 151,8 kb em *L. luteus* (Martin et al., 2014).

A quantidade total de genes presente em genomas de cloroplasto é em torno de 130 genes, como o observado, por exemplo, nas espécies *Cedrela odorata* (MG724915), *Entandrophragma cylindricum* (KY923074.1), *Khaya senegalensis* (KX364458.1) e *Carapa guianensis* (MF401522.1) (Mader et al., 2018). Neste estudo, a quantidade de genes nessas duas espécies de *Pterodon* foi de 127 genes e representa apenas um a mais que o número presente em *D. alata* (Antunes et al., 2020) e um a menos que *S. japonicum* (Lu et al., 2018). Comparando os tRNAs das duas espécies de *Pterodon* com *D. alata* e *S. japonicum*, ambos apresentam o gene *trnQ-UUG* que não foi anotado em *D. alata* e em *S. japonicum* o gene *trnA^P-GGG* não foi identificado. Quanto aos genes duplicados nas regiões IR, foi observado apenas um a menos que os identificados em *A. thaliana* (Park; Xi; Kim, 2020) e *Avena sativa* (Liu et al., 2020), ambas apresentam nesta região o gene *rps19* que está ausente em *P. emarginatus* e *P. pubescens*.

O número de genes varia entre as espécies, em decorrência de perdas durante o processo evolutivo (Qian et al., 2013; Yi et al., 2016; Antunes et al., 2020). O gene *rps19*, um dos responsáveis pela formação das proteínas da subunidade ribossômica pequena, encontra-se incompleto nos dois genomas de *Pterodon*. Antunes et al. (2020) em sua pesquisa com *D. alata*, afirma que o referido gene também não foi identificado, sugerindo que houve uma perda de parte deste gene no processo evolutivo de espécies do clado ADA (Antunes et al., 2020; Guo et al., 2007).

Neste estudo, o tamanho da região invertida em ambas as espécies foi de 25.584 pb, enquanto *D. alata* pertencente ao mesmo clado apresentou tamanho de 24.948 pb (Antunes et al., 2020). Estas regiões IR evidenciam um maior conteúdo GC, comparado às outras regiões do genoma, possivelmente explicado pela distribuição dos oito genes de RNA ribossomal (rRNA), *rrn5*, *rrn23*, *rrn16* e *rrn4.5*, com 52%, 54%, 56% e 47% de conteúdo GC, localizados nestas regiões. Assim, a média do conteúdo GC para os genes rRNA foi de 52%, enquanto em *D. alata* (Antunes et al., 2020) o resultado foi de 55%. Este valor seria maior, se não fosse pela presença de um trecho do gene *ycf1* que apresenta conteúdo GC de 34% no segmento presente nesta região.

O gene *ycf1* está inserido em uma região de fronteira entre as regiões SSC e IR, sendo que 504 pb estão repetidos nas extremidades das regiões IR, resultado semelhante identificado em *P. vulgaris*, em que este trecho do gene possui 505 pb (Guo et al., 2007). Em *G. max* a repetição apresenta 478 pb (Saski et al., 2005). Este gene apresenta alto valor de diversidade, este resultado representa a maior variação de todos os genes analisados nas regiões invertidas, o que promoveu um aumento da diversidade desta região.

Os genes de rRNA presentes nas regiões invertidas apresentam diversidade nucleotídica média de 0,014, se tratando de regiões mais conservadas do genoma, e neste sentido, contribuem com o valor baixo de diversidade na IR. Os genes com as maiores diversidade foram: *ycf1*, *trnG-UCC*, *trnK-UUU*, *ycf4*, *rpl32* e *ClpP*. No estudo com *Cersis chuniana*, Liu et al. (2018) encontraram as maiores diversidades nas regiões gênicas *trnT-trnL*, *psbZ-trnG*, *rpl32*, *rps3-rps19* e *ycf1* com valores acima de 0,12. Enquanto (Antunes et al., 2020) identificaram os maiores valores de diversidade nos genes *clpP* (0.142), *trnI-CAU* (0.1794) e *trnG-UCC* (0.23672) em *D. alata*.

Os resultados da análise entre as espécies do gênero *Pterodon* indicaram que os valores não sinônimos (Ka) / sinônimos (Ks) da maioria dos genes foi 0. Esses valores ocorreram normalmente quando as sequências alinhadas possuem 100% de correspondências entre os genes ou nos casos em que os valores de Ks são consideravelmente baixos (ANTUNES et al., 2020)(Li et al., 2020).

Souza et al. (2019) em seu estudo com *S. adstringens*, reportou a presença de três genes com valores de Ka/Ks acima de 1, indicando seleção positiva sendo eles os genes *rps16*, *psbH* e *clpP*. Enquanto Liu et al. (2018), em leguminosas cesalpinoídes identificaram quatro genes: *ndhD*, *ycf1*, *infA* e *rpl23*. Já as análises da razão Ka / Ks > 1 em espécies do gênero *Vicia*, foram identificados nove genes (*CcsA*, *clpP*, *rpl32*, *rpl33*, *rpoC1*, *rps15*, *rps2*,

rps4 e *rps7*). Liang et al. (2020) em estudo com 14 espécies de Cúrcuma encontraram os genes *atpF*, *ndhC* e *rps8* com seleção positiva, enquanto no presente estudo, os genes que apresentaram na seleção positiva foram os *ndhC*, *atpF*, *petB*, *rps14*, *ycf2* e *rpl23*. As causas e consequências destas diferentes taxas evolutivas entre genes codificadores de proteínas é um tema que merece atenção. Explicações potenciais para estes eventos surgem como disparidades no tempo de geração, seleção relaxada, nível de expressão genética e função genética (Gao et al., 2018).

O gene *rps16* que expõe a função de codificar a proteína ribossomal S16 está presente no genoma do cloroplasto da maioria das plantas superiores. Contudo, é comum identificar a perda múltipla deste gene, devido a uma consequência do duplo direcionamento da cópia nuclear *rps16* para o plastídio e para a mitocôndria, sugerindo que o *rps16* codificado pelo cloroplasto já foi silenciado e tornou-se um pseudogene pelo *rps16* codificado pelo núcleo. Deste modo, estratégias de sequenciamento, além de PCR mais rigorosas se fazem necessárias para a confirmação destes eventos (Souza et al., 2019; Jansen et al., 2008). O gene *rps16* encontra-se presente nos genomas das outras espécies do clado ADA que possuem sequências em bancos de dados, entretanto, estão ausentes nas espécies de *Pterodon*, assim como, também estão ausentes em outras espécies das Leguminosas como por exemplo, em *C. arietinum* (Jansen et al., 2008), *P. vulgaris* (Guo et al., 2007), gênero *Lupinus* (Keller et al., 2017). Em *Vigna radiata* provavelmente não é funcional, pois contém três códons de parada internos e seu códon de início é AGA (Tangphatsornruang et al., 2010).

A similaridade entre as duas espécies em estudo foi de 99% enquanto com *D. alata*, *S. japonicum* e *L. luteus*, foi de 95%, 92% e 86% respectivamente, o que sugere a proximidade filogenética entre os Clados ADA e Cladrastis (Antunes et al., 2020), resultado também evidenciado na árvore filogenética. As sequências gênicas cloroplastidiais no geral são muito conservadas, principalmente entre espécies filogeneticamente próximas, o que corrobora a alta similaridade entre os genes cloroplastidiais das duas espécies em estudo.

Quanto as regiões repetitivas no genoma, Antunes et al. (2020) identificaram 131 microssatélites no genoma cloroplastidial de *D. alata* e Souza et al. (2019) identificaram 137 microssatélites em *S. adstringens*. Estes valores são próximos aos identificados nas espécies em estudo e são considerados relativamente baixos. Os elementos repetitivos podem contribuir com a diversidade genética. Esta baixa quantidade de microssatélites pode indicar uma maior conservação destes genomas cloroplastidiais (Li; Su; Wang, 2018; Saski et al., 2005; Antunes et al., 2020).

6. CONCLUSÕES

O tamanho, a localização e composição dos genes, assim como das regiões microssatélites dos genomas cloroplastidiais completos de *P. emarginatus* e *P. pubescens* são bastante semelhantes entre as duas espécies estudadas. Primers estão disponibilizados para amplificar as regiões microssatélites com potencial uso em futuros trabalhos de desenvolvimento de marcadores. A análise mostra a perda de genes que são comuns em outras espécies, perdas essas, possivelmente associadas a processos evolutivos. A diversidade nucleotídica corrobora com outros trabalhos, a região IR apresenta maior conservação em comparação com as outras regiões do genoma. O resultado obtido nesta pesquisa é um significativo aporte para maior entendimento sobre o gênero *Pterodon*, expandindo assim, as possibilidades de pesquisa genética, para futuros trabalhos de conservação e melhoramento das espécies.

7. REFERÊNCIAS BIBLIOGRÁFICAS

- AFONSO, A. P. **Biologia reprodutiva e polinização de *Pterodon pubescens* Benth e *Pterodon polygalaeflorus* Benth (Fabaceae-Papilionoideae) em áreas de cerrado do Distrito Federal.** [s.l.] Universidade de Brasília, 1997.
- ALMEIDA, S. P. DE; PROENÇA, C. E. B.; SANO, S. M.; RIBEIRO, J. F. **Cerrado: espécies vegetais úteis.** Planaltina: Embrapa - CPAC, 1998.
- ALMEIDA, D. et al. Plantas visitadas por abelhas e polinização. **Série Produtor Rural**, p. 40, 2003.
- ALTSCHUL, S. F. et al. Basic local alignment search tool. **Journal of Molecular Biology**, v. 215, n. 3, p. 403–410, 1990.
- ALVES, S. F. Estudo da composição química, de atividades biológicas e microencapsulação do óleo essencial dos frutos de *Pterodon emarginatus* Vogel – Fabaceae (“sucupira”) Estudo da composição química , de atividades biológicas e microencapsulação do óleo essencial. p. 198, 2012.
- ALVES, S. F. et al. Microencapsulation of Essential Oil from Fruits of *Pterodon emarginatus* Using Gum Arabic and Maltodextrin as Wall Materials: Composition and Stability. **Drying Technology**, v. 32, n. 1, p. 96–105, 2014.
- AMAR, D. et al. Evaluation and integration of functional annotation pipelines for newly sequenced organisms: The potato genome as a test case. **BMC Plant Biology**, v. 14, n. 1, p. 329, 5 dez. 2014.
- ANDREWS, S. **Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data.** Disponível em: <<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>>. Acesso em: 19 mar. 2021.
- ANTUNES, A. M. et al. The chloroplast genome sequence of *Dipteryx alata* Vog. (Fabaceae: Papilionoideae): genomic features and comparative analysis with other legume genomes. **Revista Brasileira de Botânica**, v. 43, n. 2, p. 271–282, 1 jun. 2020.
- BANKEVICH, A. et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. **Journal of Computational Biology**, v. 19, n. 5, p. 455–477, 1 maio 2012.
- BARROSO GM; MORIM MP; PEIXOTO AL; ICHASO CLF (ED.). **Frutos e sementes. Morfologia aplicada à sistemática de dicotiledôneas.** . Viçosa: Editora UFV, 1999.
- BAVARESCO, O. S. A. et al. Utilização popular da *Pterodon* spp no tratamento de doenças reumáticas. **Revista de Investigação Biomédica**, v. 8, n. 1, p. 81, 2016.
- BEIER, S. et al. MISA-web: a web server for microsatellite prediction. **Bioinformatics (Oxford, England)**, v. 33, n. 16, p. 2583–2585, 15 ago. 2017.
- BENDICH, A. J. Why do chloroplasts and mitochondria contain so many copies of their genome? **BioEssays**, v. 6, n. 6, p. 279–282, 1987.
- BERGER, I. J. **Sequenciamento e análise de fragmentos cloroplastidiais de tomate para construção de vetores de transformação de cloroplastos via recombinação homóloga.** Piracicaba: Biblioteca Digital de Teses e Dissertações da Universidade de São Paulo, 19 dez.

2019.

BOLGER, A. M.; LOHSE, M.; USADEL, B. Trimmomatic: A flexible trimmer for Illumina sequence data. **Bioinformatics**, v. 30, n. 15, p. 2114–2120, 1 ago. 2014.

BROUARD, J. S. et al. The exceptionally large chloroplast genome of the green alga *Floydiella terrestris* illuminates the evolutionary history of the Chlorophyceae. **Genome Biology and Evolution**, v. 2, n. 1, p. 240–256, 2010.

CAICEDO, A. L.; PURUGGANAN, M. D. **Comparative plant genomics. Frontiers and prospects** *Plant Physiology* American Society of Plant Biologists, , 1 jun. 2005. Disponível em: <www.plantphysiol.org/cgi/doi/10.1104/pp.104.900148>. Acesso em: 19 mar. 2021

CARDOSO, D. et al. Reconstructing the deep-branching relationships of the papilionoid legumes. **South African Journal of Botany**, v. 89, p. 58–75, 1 nov. 2013.

CARDOSO, D. et al. Filling in the gaps of the papilionoid legume phylogeny: The enigmatic Amazonian genus *Petaladenium* is a new branch of the early-diverging Amburaneae clade. **Molecular Phylogenetics and Evolution**, v. 84, p. 112–124, 1 mar. 2015.

CARVALHO, C. S. **Flora do Brasil - Pterodon Vogel**. 2020. Acesso em: 19 mar. 2021.

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

DANIELL, H. et al. **Chloroplast genomes: Diversity, evolution, and applications in genetic engineering** *Genome Biology* BioMed Central Ltd., , 23 jun. 2016. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/27339192/>>. Acesso em: 19 mar. 2021

DE ARAÚJO, L. A. et al. Angiogenic activity of sucupira (*Pterodon emarginatus*) oil. **Scientia Medica**, v. 25, n. 2, p. 20351, 10 set. 2015.

DELANNOY, E. et al. Rampant Gene Loss in the Underground Orchid *Rhizanthella gardneri* Highlights Evolutionary Constraints on Plastid Genomes. **Molecular Biology and Evolution**, v. 28, n. 7, p. 2077–2086, 1 jul. 2011.

DOYLE, J. J.; DOYLE, J. L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. **PHYTOCHEMICAL BULLETIN**, 1987.

DUTRA, R. C. et al. CARACTERIZAÇÃO FITOQUÍMICA E CITOGENÉTICA DAS SEMENTES DE *Pterodon emarginatus* VOGEL. **Revista Técnico Científica do IFSC**, v. 1, n. 2, p. 99, 2012.

FILARDI, F. L. R. et al. Padrões de distribuição geográfica de espécies arbóreas de Leguminosae ocorrentes no cerrado. **Revista Brasileira de Biociências**, v. 5, n. 2, p. 1116–1118, 2007.

FRAZER, K. A. et al. VISTA: Computational tools for comparative genomics. **Nucleic Acids Research**, v. 32, n. WEB SERVER ISS., p. W273–W279, 1 jul. 2004.

GAO, X. et al. Comparative chloroplast genomes of *Paris Sect. Marmorata*: Insights into repeat regions and evolutionary implications. **BMC Genomics**, v. 19, n. Suppl 10, 31 dez. 2018.

GRAVELEY, B. R. **Alternative splicing: Increasing diversity in the proteomic world** *Trends in Genetics* Elsevier, , 1 fev. 2001. Disponível em: <<http://www.cell.com/article/S0168952500021764/fulltext>>. Acesso em: 19 mar. 2021

GUERRERO, R.; MARGULIS, L.; BERLANGA, M. Symbiogenesis: The holobiont as a unit of evolution. **International Microbiology**, v. 16, n. 3, p. 133–143, 2013.

- GUO, X. et al. Rapid evolutionary change of common bean (*Phaseolus vulgaris* L) plastome, and the genomic diversification of legume chloroplasts. **BMC Genomics**, v. 8, n. 1, p. 228, 10 jul. 2007.
- HANSEN, D. et al. 14, 15-epoxygeranylgeraniol and extracts isolated from *Pterodon emarginatus* Vog. fruits: antitumor activity on glioblastoma cells. **Journal of Applied Pharmaceutical Science**, v. 2, n. 9, p. 20–024, 2012.
- HANSEN, D.; HARAGUCHI, M.; ALONSO, A. **Pharmaceutical properties of “sucupira” (*Pterodon spp*)** **Brazilian Journal of Pharmaceutical Sciences** Faculdade de Ciências Farmaceuticas (Biblioteca), , 1 dez. 2010. Disponível em: <<https://www.revistas.usp.br/bjps/article/view/10817>>. Acesso em: 19 mar. 2021
- HOSCHEID, J.; CARDOSO, M. L. C. Sucupira as a Potential Plant for Arthritis Treatment and Other Diseases. **Arthritis**, v. 2015, p. 1–12, 3 nov. 2015.
- JANSEN, R. K. et al. Complete plastid genome sequence of the chickpea (*Cicer arietinum*) and the phylogenetic distribution of rps12 and clpP intron losses among legumes (Leguminosae). **Molecular Phylogenetics and Evolution**, v. 48, n. 3, p. 1204–1217, set. 2008.
- JEON, J. H.; KIM, S. C. Comparative analysis of the complete chloroplast genome sequences of three closely related east-asian wild roses (*Rosa* sect. *synstylae*; rosaceae). **Genes**, v. 10, n. 1, 1 jan. 2019.
- JIN, D.-P.; CHOI, I.-S.; CHOI, B.-H. Plastid genome evolution in tribe Desmodieae (Fabaceae: Papilionoideae). **PLOS ONE**, v. 14, n. 6, p. e0218743, 24 jun. 2019.
- KELLER, J. et al. The evolutionary fate of the chloroplast and nuclear rps16 genes as revealed through the sequencing and comparative analyses of four novel legume chloroplast genomes from *Lupinus*. **DNA Research**, v. 24, n. 4, p. 343–358, 1 ago. 2017.
- KERSTEN, B. et al. Genome sequences of *Populus tremula* chloroplast and mitochondrion: Implications for holistic poplar breeding. **PLoS ONE**, v. 11, n. 1, 1 jan. 2016.
- KHAN, A. et al. First complete chloroplast genomics and comparative phylogenetic analysis of *Commiphora gileadensis* and *C. Foliacea*: Myrrh producing trees. **PLoS ONE**, v. 14, n. 1, 1 jan. 2019.
- KOLODNER, R.; TEWARI, K. K. Physicochemical characterization of mitochondrial DNA from pea leaves. **Proceedings of the National Academy of Sciences of the United States of America**, v. 69, n. 7, p. 1830–1834, 1972.
- KUMAR, S. et al. MEGA X: Molecular evolutionary genetics analysis across computing platforms. **Molecular Biology and Evolution**, v. 35, n. 6, p. 1547–1549, 1 jun. 2018.
- LAGESSEN, K. et al. RNAMmer: Consistent and rapid annotation of ribosomal RNA genes. **Nucleic Acids Research**, v. 35, n. 9, p. 3100–3108, 1 maio 2007.
- LANGMEAD, B. et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. **Genome Biology**, v. 10, n. 3, p. R25, 4 mar. 2009.
- LI, D. M. et al. Complete chloroplast genomes of *Zingiber montanum* and *Zingiber zerumbet*: Genome structure, comparative and phylogenetic analyses. **PLoS ONE**, v. 15, n. 7 July, 1 jul. 2020.
- LI, H.; DURBIN, R. Fast and accurate short read alignment with Burrows-Wheeler transform. **Bioinformatics**, v. 25, n. 14, p. 1754–1760, jul. 2009.
- LI, J.; SU, Y.; WANG, T. The repeat sequences and elevated substitution rates of the

- chloroplast accd gene in cupressophytes. **Frontiers in Plant Science**, v. 9, 20 abr. 2018.
- LIANG, H. et al. The Complete Chloroplast Genome Sequences of 14 Curcuma Species: Insights Into Genome Evolution and Phylogenetic Relationships Within Zingiberales. **Frontiers in Genetics**, v. 11, 23 jul. 2020.
- LIMA, J. R.; TOZZI, A. M. G. A.; MANSANO, V. F. A checklist of woody Leguminosae in the South American Corridor of Dry Vegetation. **Phytotaxa**, v. 207, n. 1, p. 1–38, 2015.
- LIMA, S. Filogeografia de *Pterodon emarginatus* E *Pterodon pubescens* (LEGUMINOSAE). v. TESE, 2019.
- LIU, Q. et al. Comparative chloroplast genome analyses of *Avena*: Insights into evolutionary dynamics and phylogeny. **BMC Plant Biology**, v. 20, n. 1, p. 406, 2 set. 2020.
- LIU, W. et al. Complete chloroplast genome of *cercis chuniana* (Fabaceae) with structural and genetic comparison to six species in caesalpinioideae. **International Journal of Molecular Sciences**, v. 19, n. 5, 1 maio 2018.
- LOHSE, M. et al. Mercator: A fast and simple web server for genome scale functional annotation of plant sequence data. **Plant, Cell and Environment**, v. 37, n. 5, p. 1250–1258, 1 maio 2014.
- LORENZI, H. **Árvores Brasileiras Volume 1 - 5ª Edição Manual de Identificação e C.** Disponível em: <<https://www.editoraufv.com.br/produto/arvores-brasileiras-vol-1-8-edicao/1109405>>. Acesso em: 19 mar. 2021.
- LOWE, T. M.; EDDY, S. R. tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. **Nucleic Acids Research**, v. 25, n. 5, p. 955–964, 1 mar. 1997.
- LU, Y. et al. The complete chloroplast genome sequence of *Sophora japonica* var. *violacea*: gene organization and genomic resources. **Conservation Genetics Resources**, v. 10, n. 1, p. 1–4, 1 mar. 2018.
- LUO, R. et al. **SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler** *GigaScience* Oxford University Press, , 27 dez. 2012. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/23587118/>>. Acesso em: 26 mar. 2021
- MADER, M. et al. Complete chloroplast genome sequences of four meliaceae species and comparative analyses. **International Journal of Molecular Sciences**, v. 19, n. 3, 1 mar. 2018.
- MARDIS, E. R. Next-Generation DNA Sequencing Methods. **Annual Review of Genomics and Human Genetics**, v. 9, n. 1, p. 387–402, 3 set. 2008.
- MARTIN, G. E. et al. The first complete chloroplast genome of the Genistoid legume *Lupinus luteus*: Evidence for a novel major lineage-specific rearrangement and new insights regarding plastome evolution in the legume family. **Annals of Botany**, v. 113, n. 7, p. 1197–1210, 2014.
- MELO, P. Z. TAMANHO GENÔMICO E MARCADORES MICROSSATÉLITES PARA *Pterodon pubescens* E *Pterodon emarginatus* (Leguminosae). **Programa De Pós-Graduação Em Genética E Biologia Molecular**, v. Dissertação, p. 1–61, 2018.
- MENDONÇA, R. **Flora Vascular Do Bioma Cerrado: Checklist Com 12.356 Espécies** Embrapa Informação Tecnológica, , 2008.
- METZKER, M. L. **Sequencing technologies the next generation** *Nature Reviews Genetics* Nature Publishing Group, , 8 jan. 2010. Disponível em:

<<https://www.nature.com/articles/nrg2626>>. Acesso em: 19 mar. 2021

MILLER, J. R.; KOREN, S.; SUTTON, G. **Assembly algorithms for next-generation sequencing data** Genomics Academic Press, , 1 jun. 2010.

MOHANTA, T. K. et al. Gene loss and evolution of the plastome. **Genes**, v. 11, n. 10, p. 1–23, 1 out. 2020.

MUSCHNER, V. C. FILOGENIA MOLECULAR, TAXAS EVOLUTIVAS, TEMPO DE DIVERGÊNCIA E HERANÇA ORGANELAR EM *Passiflora L.* (PASSIFLORACEAE). **Tempo**, 2005.

NEGRI, G.; MATTEI, R.; MENDES, F. R. Antinociceptive activity of the HPLC- and MS-standardized hydroethanolic extract of *Pterodon emarginatus* Vogel leaves. **Phytomedicine**, v. 21, n. 8–9, p. 1062–1069, 2014.

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.

OLIVEIRA, A. E. M. F. M. et al. Development of a Larvicidal Nanoemulsion with *Pterodon emarginatus* Vogel Oil. **PLOS ONE**, v. 11, n. 1, p. e0145835, 7 jan. 2016.

PAN, I. C. et al. Complete chloroplast genome sequence of an orchid model plant candidate: *Erycina pusilla* apply in tropical *oncidium* breeding. **PLoS ONE**, v. 7, n. 4, 4 abr. 2012.

PARK, J.; XI, H.; KIM, Y. The Complete Chloroplast Genome of *Arabidopsis thaliana* Isolated in Korea (Brassicaceae): An Investigation of Intraspecific Variations of the Chloroplast Genome of Korean *A. thaliana*. **International Journal of Genomics**, v. 2020, 2020.

PASCOA, H. et al. Microemulsion based on *pterodon emarginatus* oil and its anti-inflammatory potential. **Brazilian Journal of Pharmaceutical Sciences**, v. 51, n. 1, p. 117–126, 1 jan. 2015.

QIAN, J. et al. The Complete Chloroplast Genome Sequence of the Medicinal Plant *Salvia miltiorrhiza*. **PLoS ONE**, v. 8, n. 2, p. e57607, 27 fev. 2013.

QUAIL, M. A. et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. **BMC Genomics**, v. 13, n. 1, 24 jul. 2012.

RATAN, A. ASSEMBLY ALGORITHMS FOR NEXT-GENERATION SEQUENCE DATA. **Aspectos Generales De La Planificación Tributaria En Venezuela**, v. 2009, n. 75, p. 31–47, 2009.

ROCHA, D. M. S. Aspectos taxonômicos, genéticos e reprodutivos de *Pterodon pubescens* (Benth.) Benth. e *Pterodon emarginatus* Vog. (Leguminosae, Dipteryxaceae). n. September, 2006.

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, 1 dez. 2017.

SANTANA, A. R. DE. UNIVERSIDADE FEDERAL DE GOIÁS INSTITUTO DE CIÊNCIAS BIOLÓGICAS PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR TRANSFERIBILIDADE E DESENVOLVIMENTO DE SISTEMA MULTIPLEX DE GENOTIPAGEM DE MARCADORES MICROSSATÉLITES PARA *Pterodon emarginatus* Vogel (FABACEAE). [s.l: s.n.].

SANTOS, A. P. et al. Composição química, atividade antimicrobiana do óleo essencial e

- ocorrência de esteróides nas folhas de *Pterodon emarginatus* Vogel, Fabaceae. **Brazilian Journal of Pharmacognosy**, v. 20, n. 6, p. 891–896, dez. 2010.
- SASKI, C. et al. Complete chloroplast genome sequence of glycine max and comparative analyses with other legume genomes. **Plant Molecular Biology**, v. 59, n. 2, p. 309–322, set. 2005.
- SCHATZ, M. C.; DELCHER, A. L.; SALZBERG, S. L. **Assembly of large genomes using second-generation sequencing** *Genome Research*, , set. 2010. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/20508146/>>. Acesso em: 26 mar. 2021
- SCHLEBUSCH, S.; ILLING, N. Next generation shotgun sequencing and the challenges of de novo genome assembly Affiliation. **Review Article S Afr J Sci**, v. 108, n. 11, 2012.
- SCHMUTZ, J. et al. Genome sequence of the palaeopolyploid soybean. **Nature**, v. 463, n. 7278, p. 178–183, 14 jan. 2010.
- SCHMUTZ, J. et al. A reference genome for common bean and genome-wide analysis of dual domestications. **Nature Genetics**, v. 46, n. 7, p. 707–713, 8 jun. 2014.
- SHINOZAKI, K. et al. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. **The EMBO Journal**, v. 5, n. 9, p. 2043–2049, 1 set. 1986.
- SOUZA, U. J. B. DE et al. The complete chloroplast genome of *Stryphnodendron adstringens* (Leguminosae - Caesalpinioideae): comparative analysis with related Mimosoid species. **Scientific Reports**, v. 9, n. 1, p. 1–12, 1 dez. 2019.
- TANG, H. et al. An improved genome release (version Mt4.0) for the model legume *Medicago truncatula*. **BMC Genomics**, v. 15, n. 1, p. 312, 27 abr. 2014.
- TANGPHATSORNRUANG, S. et al. The chloroplast genome sequence of mungbean (*Vigna radiata*) determined by high-throughput pyrosequencing: Structural organization and phylogenetic relationships. **DNA Research**, v. 17, n. 1, p. 11–22, fev. 2010.
- TAQUARY, A. M. A. Tamanho, montagem de novo e anotação do genoma de *Dipteryx alata* (Leguminosae). **Universidade Federal de Goiás**, p. 1–137, 2017.
- THOMPSON, J. D.; HIGGINS, D. G.; GIBSON, T. J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. **Nucleic Acids Research**, v. 22, n. 22, p. 4673–4680, 11 nov. 1994.
- TILLICH, M. et al. GeSeq - Versatile and accurate annotation of organelle genomes. **Nucleic Acids Research**, v. 45, n. W1, p. W6–W11, 3 jul. 2017.
- UNTERGASSER, A. et al. Primer3-new capabilities and interfaces. **Nucleic Acids Research**, v. 40, n. 15, ago. 2012.
- VARSHNEY, R. K. et al. Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. **Nature Biotechnology**, v. 31, n. 3, p. 240–246, 1 mar. 2013.
- WAMBUGU, P. W. et al. Relationships of wild and domesticated rices (*Oryza* AA genome species) based upon whole chloroplast genome sequences. **Scientific Reports**, v. 5, 10 set. 2015.
- WANG, W. et al. Assembly of chloroplast genomes with long- and short-read data: A comparison of approaches using *Eucalyptus pauciflora* as a test case. **BMC Genomics**, v. 19, n. 1, p. 977, 29 dez. 2018.
- WYMAN, S. K.; JANSEN, R. K.; BOORE, J. L. Automatic annotation of organellar

- genomes with DOGMA. **Bioinformatics**, v. 20, n. 17, p. 3252–3255, 22 nov. 2004.
- 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, 1 dez. 2017.
- XU, J. H. et al. Dynamics of chloroplast genomes in green plants. **Genomics**, v. 106, n. 4, p. 221–231, 1 out. 2015.
- XUE, S. et al. Comparative analysis of the complete chloroplast genome among *Prunus mume*, *P. armeniaca*, and *P. salicina*. **Horticulture Research**, v. 6, n. 1, 1 dez. 2019.
- YANG, Z. **Adaptive Molecular Evolution**. [s.l: s.n.].
- YI, D. K. et al. The complete chloroplast genome sequence of *Abies nephrolepis* (Pinaceae: Abietoideae). **Journal of Asia-Pacific Biodiversity**, v. 9, n. 2, p. 245–249, 1 jun. 2016.
- ZHANG, Y. et al. Complete chloroplast genome sequences of *Praxelis (Eupatorium catarium Veldkamp)*, an important invasive species. **Gene**, v. 549, n. 1, p. 58–69, 1 out. 2014.
- ZHANG, Y. et al. The complete chloroplast genome of greater duckweed (*Spirodela polyrhiza* 7498) using PacBio long reads: Insights into the chloroplast evolution and transcription regulation. **BMC Genomics**, v. 21, n. 1, p. 76, 28 jan. 2020.
- ZIMIN, A. V. et al. The MaSuRCA genome assembler. **Bioinformatics**, v. 29, n. 21, p. 2669–2677, 1 nov. 2013.
- ZOSCHKE, R.; LIERE, K.; BÖRNER, T. From seedling to mature plant: *Arabidopsis* plastidial genome copy number, RNA accumulation and transcription are differentially regulated during leaf development. **The Plant Journal**, v. 50, n. 4, p. 710–722, 25 abr. 2007.

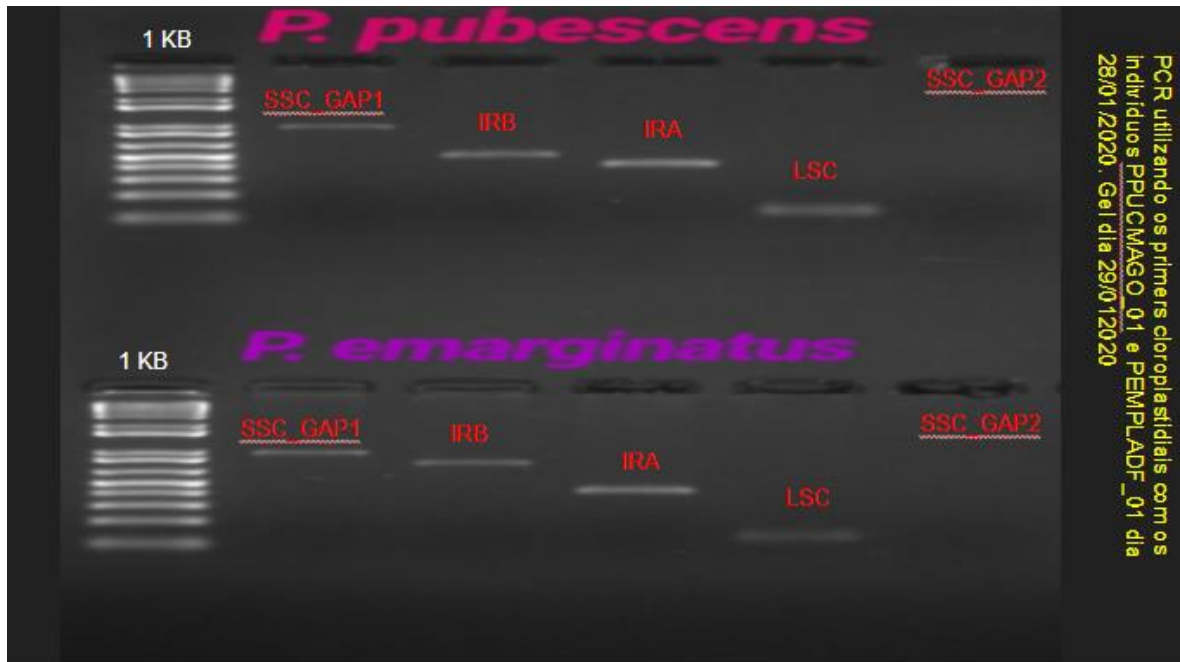


Figura 1. Eletroforese horizontal, em gel de agarose 3%, com produtos de Reações em Cadeia da Polimerase (PCR) das regiões de gap para indivíduos de *Pterodon pubescens* e *P. emarginatus*. Protocolo de PCR segundo Melo (2018).