UNIVERSIDADE FEDERAL DE GOIÁS

INSTITUTO DE CIÊNCIAS BIOLÓGICAS

PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

Citogenômica de *Pterodon pubescens* e citogenética comparativa com *P. emarginatus* (Leguminosae)

Victória Borges Albernaz

Nível: Mestrado

Orientadora: Profa. Dra.Thannya Nascimento Soares

Coorientador: Prof. Dr. Luiz Gustavo Rodrigues Souza

GOIÂNIA-GO

2020

UFG

UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS

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

E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG

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

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

1. Identificação do material bibliográfico

[x] Dissertação [ ] Tese

2. Nome completo do autor

Victória Borges Albernaz

3. Título do trabalho

Citogenômica de *Pterodon pubescens* e citogenética comparativa com *P. emarginatus* (Leguminosae)

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

Concorda com a liberação total do documento [X] SIM [ ] NÃO[1]

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. Após esse período, a possível disponibilização ocorrerá apenas mediante:
a) consulta ao(à) autor(a) e ao(à) orientador(a);
b) novo Termo de Ciência e de Autorização (TECA) assinado e inserido no arquivo da tese ou dissertação.
O documento não será disponibilizado durante o período de embargo.
Casos de embargo:
- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;
- Publicação da dissertação/tese em livro.

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

Documento assinado eletronicamente por **Thannya Nascimento Soares**, **Coordenadora de Curso**, em 17/06/2020, às 17:28, 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 **VICTORIA BORGES ALBERNAZ**, **Discente**, em 19/06/2020, às 19:13, 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 **1382612** e o código CRC **BD176C0D**.

Referência: Processo nº 23070.005664/2020-99 SEI nº 1382612

VICTÓRIA BORGES ALBERNAZ

Citogenômica de *Pterodon pubescens* e citogenégica comparativa com *P. emarginatus* (Leguminosae)

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

Orientadora: Prof. Dr. Thannya Nascimento Soares
Co-orientador: Prof. Dr. Luiz Gustavo Rodrigues Souza

GOIÂNIA-GO
2020

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

Borges Albernaz, Victória
Citogenômica de Pterodon pubescens e citogenética comparativa com P. emarginatus (Leguminosae) [manuscrito] / Victória Borges Albernaz. - 2020.
lxxvi, 76 f.: il.

Orientador: Profa. Dra. Thannya Nascimento Soares; co orientador Dr. Luiz Gustavo Rodrigues Souza.
Dissertação (Mestrado) - Universidade Federal de Goiás, Instituto de Ciências Biológicas (ICB), Programa de Pós-Graduação em Genética e Biologia Molecular, Goiânia, 2020.
Bibliografia.
Inclui siglas, mapas, abreviaturas, tabelas.

1. DNAsat. 2. FISH. 3. NGS. 4. sequências repetitivas. 5. sucupira-branca. I. Nascimento Soares, Thannya, orient. II. Título.

CDU 575

UNIVERSIDADE FEDERAL DE GOIÁS

INSTITUTO DE CIÊNCIAS BIOLÓGICAS

ATA DE DEFESA DE DISSERTAÇÃO

Ata nº **082** da sessão de Defesa de Dissertação de **VICTÓRIA BORGES ALBERNAZ**, que confere o título de Mestre(a) em **GENÉTICA E BIOLOGIA MOLECULAR**, na área de concentração em **Genética de Populações e Evolução Molecular**.

Ao/s **cinco dias do mês de março de dois mil e vinte**, a partir das **8:30 horas**, no **miniauditório** do **CERCOMP**, realizou-se a sessão pública de Defesa de Dissertação intitulada **"Citogenômica comparativa de *Pterodon pubescens* e *P. emarginatus* (Leguminosae)"**. Os trabalhos foram instalados pela Orientadora, Professora Doutora **Thannya Nascimento Soares (ICB/UFG)** com a participação dos demais membros da Banca Examinadora: Professora Doutora **Andrea Pedrosa-Harand (CB/UFPE)**, membro titular externo, cuja participação ocorreu através de videoconferência e a Doutora **Rosana Pereira Vianello (EMBRAPA)**, membro titular externo. Durante a arguição os membros da banca **fizeram** sugestão de alteração do título do trabalho para **Citogenômica de *Pterodon pubescens* e citogenética comparativa com *P. emarginatus* (Leguminosae)**. 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 Professora Doutora **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, ao(s) **cinco dias do mês de março de dois mil e vinte**.

TÍTULO SUGERIDO PELA BANCA

Citogenômica de *Pterodon pubescens* e citogenética comparativa com *P. emarginatus* (Leguminosae)

Documento assinado eletronicamente por **Andrea Pedrosa Harand**, **Usuário Externo**, em 05/03/2020, 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.

Documento assinado eletronicamente por **Thannya Nascimento Soares**, **Coordenadora de Curso**, em 05/03/2020, 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.

Documento assinado eletronicamente por **Rosana Pereira Vianello**, **Usuário Externo**, em 05/03/2020, às 13:52, 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 **1204280** e o código CRC **10622236**.

Referência: Processo nº 23070.005664/2020-99 SEI nº 1204280

Agradecimentos

Agradeço a minha orientadora Thannya pela oportunidade, orientação, convívio, compreensão, ensinamentos que foram importantes para minha formação e principalmente por sempre acreditar no meu potencial. Eu a admiro muito como profissional e pessoa, muito dedicada e atenta as nossas dificuldades pessoais.

Ao meu co-orientador Gustavo por todo apoio, ajuda no desenvolvimento do meu trabalho mesmo de longe.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa de estudo para o desenvolvimento do trabalho de mestrado. Aos órgãos de fomento que financiaram o desenvolvimento desse trabalho, sendo eles CNPq (chamada MCTIC/CNPq Nº 28/2018 - Universal, proc. 431320/2018-7), PRONEM (chamada pública: Nº. 07/2016, proc. 201710267000539). O presente trabalho foi desenvolvido no contexto do Instituto Nacional de Ciência e Tecnologia (INCT) em Ecologia, Evolução e Conservação da Biodiversidade - EECBio, apoiado pelo MCTIC/CNPq e FAPEG (proc. 465610 / 2014-5).

Aos colegas do Laboratório de Genética & Biodiversidade: Ramilla, Rhewter, Jordana, Lucas, Daniel, Ana Cristina, Aline, Eliane, Juliana, Amanda, Sara, Larissa, Ariany, Adriana, Rafael pelos momentos de estudo e trabalho.

Aos meus amigos da graduação: Thaynah Faria, Isabella Melo, Igor Dardenne e Victor Tedesco, mesmo cada um seguindo se caminho não perdemos a amizade e companheirismo.

Em especial gostaria de agradecer a Thaynah por se manter presente apesar da sua mudança para Uberlândia, mesmo que durante o mestrado nos vimos poucas vezes, mas essas poucas vezes foram memoráveis e cheias de alegria. Obrigada também por sempre atender minhas ligações de 2 horas de duração para ouvir as novidades e reclamações.

À Cíntia Targueta, obrigada por todas as discussões enriquecedoras, companhia e risadas. Você é uma excelente pesquisadora que tive o prazer de conhecer, espero continuar tendo momentos de aprendizagem.

Aos meus amigos Rodrigo, Mari, Leonardo Oliveira, Victor Felipe, Mariel, Leonardo Corvalan pelos momentos divertidos, companheirismo e por ouvir meus desabafos. Mesmo tendo conhecidos vocês no final do mestrado os considero muito!

Aos meus amigos de Recife: Mariela, Yennifer, Pablo, Bruna, Amanda, Erton, Paulo, Natália, Amália e Rayssa, que fiz durante o estágio de três meses no laboratório de

Citogenética e Evolução Vegetal (UFPE). Nesse curto período me senti acolhida e que fazia parte do laboratório, foi uma experiência incrível e cheia de aprendizados! Em especial quero agradecer a Mariela e Yennifer pela paciência e pelo excelente treinamento, aprendi muito com as duas.

Aos professores responsáveis pelo laboratório de Citogenética e Evolução Vegetal (UFPE), Marcelo Guerra, Andrea Pedrosa-Harand e Luiz Gustavo Rodrigues Souza por estarem sempre disponíveis a auxiliar os alunos.

À minha amiga de longa data Renata Freitas, que ao passar dos anos nossa amizade fortalece cada vez mais, mesmo quando não podemos nos ver frequentemente, todos os nossos encontros são inesquecíveis e cheios de alegria. Obrigada por sempre acreditar no meu potencial e por estar presente nos momentos alegres e difíceis.

À minha família que sempre me incentivou e esteve ao meu lado em todas as minhas pequenas e grandes conquistas. Especialmente a minha mãe, Sonia, que me ajudou a ser a pessoa que sou hoje, com capacidade de correr atrás dos meus objetivos e sonhos, e com liberdade para viver a vida que desejo e buscar sempre minha independência.

Aos meus irmãos: Amanda, Filipe e Vinicius, que às vezes acreditam mais em mim do que eu mesma, que compreendem meus nervosismos e ansiedade.

Obrigada por tudo!

Sumário

Resumo

O gênero *Pterodon* Vogel (Leguminosae) possui apenas quatro espécies, das quais *P. pubescens* (Benth.) Benth. e *P. emarginatus* Vogel (ambas conhecidas como "sucupira branca") são as mais próximas filogeneticamente. Estas espécies apresentam ampla distribuição geográfica no Brasil e possuem o número cromossômico 2*n* = 16 com cromossomos pequenos e morfologicamente similares. O objetivo do presente trabalho foi realizar uma análise comparativa do tamanho do genoma, o padrão de bandeamento e a composição de elementos repetitivos nos cromossomos de *P. pubescens* e *P. emarginatus*, visando aprofundar os conhecimentos citogenômico e evolutivo destas espécies. Para isso, foi realizada a caracterização citogenética pela análise do número e morfologia cromossômica, bandeamento CMA e DAPI, FISH com DNAr 5S e 35S e determinação do tamanho do genoma por citometria de fluxo. Além disso, utilizamos o sequenciamento do genoma de *P. pubescens* por NGS (*Next Generation Sequencing*) usando a plataforma Illumina para caracterizar as frações repetitivas genômicas, utilizando uma plataforma Galaxy/RepeatExplorer-Elixir. Os elementos mais abundantes do genoma de *P. pubescens* foram localizados nos cromossomos por hibridização fluorescente *in situ* (FISH) e transferidos para *P. emarginatus*. As espécies apresentaram cariótipos muito semelhantes com: (i) bandas $CMA^+/DAPI^-$ na região terminal de dois pares de cromossomos e na região pericentromérica de todos os cromossomos; (ii) dois pares de sítios de DNAr 35S co-localizados com as bandas CMA^+ terminais; (iii) um par de sítios de DNAr 5S localizados na região proximal de um par cromossômico. O tamanho do genoma de *P. pubescens* e *P. emarginatus* também foi semelhante, 1C = 0,665 pg e 1C = 0,620 pg, respectivamente. A fração repetitiva representou 26,4% do genoma obtido de *P. pubescens*, sendo que os retrotransposons Ty3-Athila (24,24%), Ty3-Tekay (21,93%) e Ty1-Ale (3,37%) são os elementos mais abundantes. DNAs satélites de baixa abundância foram identificados: PubSat1-254 (2,09%), PubSat2-76 (2,06%), PubSat3-216 (0,58%), PubSat4-138 (0,23%). A hibridização *in situ* revelou que todas as sequências repetitivas analisadas estão colocalizadas com as bandas CMA^+ proximais em ambas espécies, exceto o retroelemento Ty1-Ale, que também está disperso na eucromatina dos cromossomos de *P. pubescens*. A similaridade citomolecular observada aqui sugere que os genomas de *P. pubescens* e *P. emarginatus* apresentam frações repetitivas altamente semelhantes, o que corrobora sua proximidade filogenética. No entanto, a recente expansão do

elemento Ty1-Ale no genoma de *P. pubescens* sugere algum grau de diferenciação nas frações repetitivas desses genomas.

Palavras-chave: DNAsat, FISH, NGS, sequências repetitivas, sucupira-branca, retroelementos

Abstract

The genus *Pterodon* Vogel (Leguminosae) has only four species, of which *P. pubescens* (Benth.) Benth. and *P. emarginatus* Vogel (both known as "white sucupira") are the closest phylogenetically. These species have a wide geographical distribution in Brazil and have chromosome number 2*n* = 16 with small and morphologically similar chromosomes. The objective of the present work was to carry out a comparative analysis of the genome size, the banding pattern and the composition of repetitive elements in the chromosomes of *P. pubescens* and *P. emarginatus*, aiming to enhance the cytogenomic and evolutionary knowledge of these species. For this, cytogenetic characterization was performed by analyzing the number and chromosomal morphology, CMA and DAPI banding, hybridization with DNAr 5S and 35S and determining the genome size by flow cytometry. In addition, we used *P. pubescens* genome sequencing by NGS (Next Generation Sequencing) using the Illumina platform to characterize repetitive genomic fractions, using a Galaxy/RepeatExplorer-Elixir platform. The most abundant elements of the *P. pubescens* genome were located on the chromosomes by fluorescent in situ hybridization (FISH) and transferred to *P. emarginatus*. The species showed very similar karyotypes with: (i) CMA^{+}/$DAPI^{-}$ bands in the terminal region of two chromosome pairs and pericentromeric region of all chromosomes; (ii) two pairs of 35S rDNA sites co-located with the terminal CMA^{+} bands; (iii) a pair of 5S rDNA sites located in the proximal region of a chromosomal pair. The genome size of *P. pubescens* and *P. emarginatus* was also similar, 1C = 0.665 pg and 1C = 0.620 pg, respectively. The repetitive fraction represented 26,4% of the *P. pubescens* sequenced genome, with Ty3-Athila Ty3-Athila (24,24%), Ty3-Tekay (21,93%) and Ty1-Ale (3,37%) retrotransposons being the most abundant elements. Low abundance satellite DNAs were identified: PubSat1-254 (2,09%), PubSat2-76 (2,06%), PubSat3-216 (0,58%), PubSat4-138 (0,23%). *In situ* hybridization reveal that all analyzed repeats were enriched in proximal CMA^{+} heterochromatin in both species, except for the Ty1-Ale retroelement, which was also dispersed also in the euchromatin of *P. pubescens* chromosomes. The cytomolecular similarity observed here suggests that the genomes of *P. pubescens* and *P. emarginatus* have highly similar repetitive fractions, which corroborates their phylogenetic proximity. However, the recent expansion of the Ty1-

Ale element in the *P. pubescens* genome suggests some degree of differentiation in the repetitive fractions of these genomes.

Keywords: FISH, NGS, satDNA, repetitive sequences, "white sucupira**”**

Introdução Geral

Família Leguminosae

Leguminosae é a terceira maior família de plantas, com aproximadamente 770 gêneros e mais de 19.500 espécies, sendo as famílias Orchidaceae e Asteraceae as únicas que a superam em número de espécies (Cardoso et al. 2012; LPWG 2017). A família Leguminosae apresenta uma importância econômica e ecológica (Lewis et al. 2005; Yahara et al. 2013), perdendo apenas para a família das gramíneas em importância econômica (Lewis et al. 2013). Economicamente, essa família é valorizada pela extensa produção de diferentes tipos de alimentos, madeiras, combustíveis, medicamentos e produtos químicos (Lewis et al. 2005; Cardoso et al. 2012; Yahara et al. 2013; LPWG 2017). As espécies que compõem a família Leguminosae estão presentes em quase todos os biomas da Terra, desde florestas tropicais até desertos, e ecologicamente, a maioria das leguminosas se destacam por serem responsáveis pela fixação de nitrogênio atmosférico através de bactérias simbióticas presentes nas raízes (McKey 1994; Yahara et al. 2013; DRYFLOR 2016; LPWG 2017).

A família Leguminosae foi recentemente dividida em seis subfamílias (Cercidoideae, Detarioideae, Duparquitioideae, Dialioideae, Caesalpinioideae e Papilionoideae) (LPWG 2017). As leguminosas pertencentes à subfamília Papilionoideae formam um grupo diverso, contendo aproximadamente 14.000 espécies e 484 gêneros (Lewis et al. 2005; Lewis et al. 2013; Cardoso et al. 2015). Essa diversificação dentro desse grupo pode ser evidenciada pela morfologia floral, que inclui não apenas as flores papilionadas, mas também exemplos de flores não-papilionadas, de perda de pétalas, de pétalas indiferenciadas, numerosos estames livres, cálice fechado e simetria radial (Leite et al. 2014b). A subfamília Papilionoideae (Leguminosae) inclui três tribos distintas, Angylocalyx, Dipterygeae e Amburana, que compõem o clado ADA (Cardoso et al. 2012; Cardoso et al. 2013; Cardoso et al. 2015). Dipterygeae é um grupo monofilético e compreende quatro gêneros: *Dipteryx* Schreb., *Monopteryx* Spruce ex Benth, *Pterodon* Vogel e *Taralea* Aubl. (Cardoso et al. 2012; Cardoso et al. 2015).

Pterodon emarginatus e *P. pubescens*

O gênero *Pteroron* pode ser distinguido por apresentar características marcantes, tais como fruto em formato de criptosâmara com glândulas de óleo no epicarpo, folhas com raque marginada, semente lisa com rafe aparente, hilo na posição lateral coberto

por um arilo e um embrião liso (Leite et al. 2014a). Ao longo dos anos esse gênero apresentou diferentes números de espécies variando entre seis, três e duas espécies dependendo da base de dados que fosse consultada (Rocha 2006). Essa discrepância em relação ao número de espécies do gênero é causada principalmente pela delimitação das espécies *Pterodon pubescens* (Benth.) Benth. e *Pterodon emarginatus* Vogel, por já terem sido consideradas como uma espécie apenas, possuindo duas morfos (Rocha 2006). Segundo o REFLORA (Carvalho, em construção), o gênero Pterodon possui quatro espécies *Pterodon. abruptus* (Moric.) Benth, *Pterodon. apparicioi* Pedersoli, *P. pubescens* e *P. emarginatus*.

As espécies *P. pubescens* e *P. emarginatus* são as únicas do gênero que não são endêmicas do Brasil, mas são amplamente distribuídas pelo Brasil, ocorrendo nos estados de Rondônia, Tocantins, Bahia, Maranhão, Piauí, Goiás, Distrito Federal, Mato Grosso, Mato Grosso do Sul, Minas Gerais e São Paulo (Carvalho, em construção). Ambas as espécies ocorrem no Cerrado (*lato sensu*), Floresta Ciliar ou Galeria, exceto Floresta Estacional Semidecidual onde somente *P. emarginatus* é encontrada (Carvalho, em construção).

Em Afonso 1997, foi relatado períodos distintos de floração (*P. emarginatus* com pico em agosto enquanto *P. pubescens* em setembro), mas em campo é observada a floração simultânea, com capacidade de formar híbridos naturais (Rocha 2006). No estudo de Rocha 2006, foram encontrados três indivíduos com características morfológicas intermediárias entre as duas espécies. Nesse estudo foram utilizados marcadores RAPD e caracteres morfológicos para verificar se a variação encotrada dentro da espécie *P. emarginatus* (uma morfo com flores rosas e a outra com flores roxas) poderia ser um indicativo de que *P. emarginatus* engloba mais de uma entidade taxonômica. Os resultados moleculares e morfológicos suportam a premissa de que são na verdade duas espécies distintas. As diferenças morfológicas entre *P. pubescens* e *P. emarginatus* são discretas, de folhas pusbescentes com 6 a 19 folíolos ou folhas glabras com 4 a 10 folíolos, respectivamente. Os indivíduos de *P. emarginatus* apresentam, em geral, folíolos grandes e mais largos, botões florais com base mais estreita do que o ápice, que é arredondado (Rocha 2006). Em contrapartida, os botões florais em *P. pubescens* são alongados, com a base da mesma largura que o ápice (Rocha 2006). A diferença morfológica de maior destaque são as flores de coloração do róseo claro, quase branco, a róseo escuro em *P. pubescens* e roxo, em *P. emarginatus* (Rocha 2006).

A polinização é realizada por abelhas e as sementes são dispersas pelo vento (Almeida et al. 2003). Em *P. emarginatus*, as flores são polinizadas por *Bombus morio*, enquanto as flores de *P. pubescens* são polinizadas por *Bombus attratus* (Afonso 1997). As diferenças de polinizadores e picos florais, corroboram a delimitação como espécies distintas, mas é provável a hibridização entre *P. pubescens* e *P. emarginatus*, principalmente em regiões que ocorrem as duas espécies. No estudo de Lima 2019, foi evidenciado uma região com essas espécies em condições de provável simpatia, no município de Itacaiú-GO e na região da bacia do Araguaia (entre latitudes -14,70 e -15,80).

Estudos com o gênero *Pterodon* têm sido relacionados principalmente ao potencial medicinal do óleo extraído de sementes. Em *P. pubescens*, o óleo apresentou atividade anti-*Trypanosoma cruzi* (Menna-Barreto et al. 2008), bem como anti-inflamatória (Hoscheid et al. 2013) e antinociceptiva (Coelho et al. 2005). Atividade anti-inflamatória e antinociceptiva também foram confirmadas para *P. emarginatus* (de Moraes et al. 2009). Trabalhos que abrangem outras áreas da ciência têm sido realizados, como genética de populações para acessar à diversidade genética e estrutura em populações naturais de *P. emarginatus* (Pinto 2017). Nesse estudo de Pinto 2017 foram testados 25 iniciadores transferidos de *Dipteryx alata* Vogel em 12 populações naturais foram genotipadas para 6 locos microssatélites. Os resultados desse estudo permitem concluir que as populações apresentaram nível intermediário de diversidade genética, maior variabilidade dentro de populações do que entre populações e a espécie apresentam um sistema misto de acasalamento. No estudo de Melo (2018), 27 marcadores microssatélites foram desenvolvidos para *P. pubescens* e 26 marcadores polimórficos foram transferidos para espécies de *P. emarginatus*, considerando que serão úteis para estudos de genética de populações devido ao alto potencial de discriminação individual encontrado. Um estudo filogeográfico comparado entre *P. pubescens* e *P. emarginatus* concluiu que as espécies estão em processo de separação recente, por isso foi encontrado um alto compartilhamento de haplótipos (Lima 2019). Esse resultado segue o padrão de recente diversificação evidenciado para outras leguminosas: *Astragalus* (Scherson et al. 2008), *Abarema* (Iganci et al. 2016), *Canavalia* (Snak et al. 2016), *Chamaecrista* ser. Coriaceae (Rando et al. 2016), *Lupinus* (Drummond 2008) e *Phaseolus* (Delgado-Salinas et al. 2007).

O gênero *Pterodon*, assim como o clado ADA, é pouco estudado do ponto de vista citogenético, sendo conhecido apenas o número cromossômico de duas espécies, *Pterodon pubescens* (Benth.) Benth. e *Pterodon polygalaeflorus* (Benth.) Benth. (= *Pterodon emarginatus* Vogel) ambos com 2*n* = 16 e cromossomos pequenos (Bandel 1974; Coleman e DeMenezes 1980). Para algumas espécies do clado ADA também são conhecidos os números cromossômicos, como *Angylocalyx oligophyllus* (Baker) Baker (2*n* = 26), Dipteryx odorata (Aubl.) Willd. (2*n* = 32), Dipteryx alata (2*n* = 16, Taquary 2017), Amburana cearensis (Allemao) A. C. Sm. (2*n* = 22), Myrospermum frutescens Jacq. (2*n* = 26), Myrospermum sousanum A. Delgado & M. C. Johnst. (2*n* =26) Myroxylon balsamiferum Harms (2*n* = 26), Myroxylon pereirae Klotzsch (2*n* = 28), Myroxylon pereirae Royle (2*n* = 28) (Rice et al. 2015).

Citogenética clássica, molecular e Citogenômica

A Citogenética consiste no estudo de cromossomos isolados ou em conjunto, condensado ou distendido, tanto no que se refere a morfologia, organização, função, estrutura e replicação (Guerra 2004). Estes cromossomos, na maioria das vezes, são cromossomos mitóticos de eucariotos. A caracterização do cariótipo diz respeito a análise de características cromossômicas como: número cromossômico, tamanho dos cromossomos, posição do centrômero e presença de constrições secundárias (Guerra 2004). Essa caracterização pode ser feita utilizando técnicas convencionais, gerando fórmulas cariotípicas para as espécies (Malik and Thomas 1966; Weiss et al. 2003).

A fórmula cariotípica para alguns gêneros de plantas que possuem número cromossômico variável, permite a diferenciação entre espécies (Stace 1978; Martel et al. 1997). No entanto, existem gêneros que possuem espécies com número cromossômico e morfologia similares (Siljak-Yakovlev et al. 2003; de Souza Almeida et al. 2007; Viana and Souza 2012). Esses casos necessitam de técnicas de coloração diferencial, sendo o bandeamento com fluorocromos CMA (cromomicina A3) e DAPI (4', 6-diamidino- 2-fenilindol), os mais utilizados para o bandeamento cromossômico em plantas. Os fluorocromos CMA e DAPI marcam as regiões ricas em GC e AT, respectivamente, e são empregados para análise da heterocromatina (Guerra 2000). Esses fluorocromos são úteis para diferenciar espécies com número cromossômico estável (da Costa Silva et al. 2015; Mendonça de Almeida et al. 2016; Van-Lume et al. 2017; Ortiz et al. 2017).

As técnicas convencionais, como CMA/DAPI, permitem a verificação da localização da heterocromatina, mas não se pode definir quais os tipos de elementos repetitivos estão presentes. No entanto, com a integração de dados genômicos e citogenéticos foi originada a citogenômica (Talukdar and Sinjushin 2015), que possibilitou a caracterização dos elementos repetitivos presentes. A identificação da composição e proporção de elementos da fração repetitiva de genomas eucarióticos que estão geralmente associados à heterocromatina, como elementos transponíveis (TEs), DNA satélite (DNAsat) e DNA ribossomal (DNAr) (Belyayev et al. 2018, Pamponét et al. 2019; Bolsheva et al. 2019; Samoluk et al. 2019; Van-Lume et al. 2019), foi facilitada pelo sequenciamento de nova geração (NGS) (Ansorge 2009) juntamente com análises bioinformáticas apropriadas (*RepeatExplorer*, Novák et al. 2013). Após a identificação dos elementos presentes no genoma, a verificação da distribuição desses elementos nos cromossomos é feita pela técnica de hibridização *in situ* fluorescente (FISH), técnica convencional da Citogenética molecular. A Citogenética molecular e Citogenômica forneceram novas possibilidades de estudo na evolução cromossômica e organização do genoma, o que também contribui para uma melhor caracterização do cariótipo (Siljak-Yakovlev et al. 2014).

A técnica de FISH consiste na hibridização de uma sequência repetitiva ou não repetitiva de DNA (sonda) à uma região específica do cromossomo (Guerra 2004). As sequências repetitivas de DNA representam uma grande parte do genoma de plantas, podendo representar mais que 80% do genoma total [*Triticum aestivum* L. (Li et al. 2004); *Zea mays* L. (Mehrotra and Goyal 2014)]. O DNA repetitivo pode ser classificado geralmente pelos seus padrões de sequência e localização. Os DNA satélites, por exemplo, são formados por sequências repetitivas (monômeros) em tandem, geralmente gênero-específicas, e são um dos principais componentes das regiões heterocromáticas, podendo estar localizados nas regiões centroméricas e subteloméricas, mas também em posições intercalares (Mehrotra e Goyal 2014). Como as regiões centroméricas são os locais onde se encontram mais agrupamentos de sequências repetitivas em tandem, tanto para plantas quanto para animais (Melters et al. 2013), a função mais aceita do DNAsat está relacionada ao centrômero, que possui um papel essencial na segregação correta dos cromossomos na mitose e meiose (Plohl et al. 2014).

O DNAsat é um dos principais componentes da heterocromatina, e diferenças na composição dessa região podem estar relacionadas com eventos de especiação e isolamento reprodutivo (Bachmann et al. 1989; Plohl et al. 2014). A região do

centrômero acumula diferenças, pois, apesar da sua função ser crítica e conservada, o DNA centromérico subjacente (e.g DNA satélite e LTR-retrotransposons) evoluem rapidamente (Plohl et al. 2014; Rosin and Mellone 2017), levando à uma compatibilidade reduzida em cromossomos homólogos de híbridos (Henikoff 2001). Esse evento de redução de compatibilidade pode gerar um isolamento pós-zigótico, desencadeando a especiação (Henikoff 2001; Fishman and Saunders 2008).

Os DNA ribossomais (DNAr) também fazem parte da fração repetitiva do genoma e são formados por sequências repetitivas em tandem altamente conservadas que codificam RNA ribossomais (RNAr) e por regiões espaçadoras intergênicas (Volkov et al. 2017). Existem dois tipos de DNAr: 35S e 5S, porém existem quatro tipos regiões codificadoras de RNAr (18S, 5.8S, 26S e 5S) que são os constituintes essenciais do ribossomo de todos eucariotos (Hemleben V, Volkov RA, Zentgraf U 2004). Os locos do DNAr 45S consistem em unidades repetidas em tandem das sequências 18S, 5.8S e 26S, sequências transcritas internas (ITS1 e ITS2) e espaçadores gênicos (IGS) compõem um loco gênico (Siljak-Yakovlev et al. 2014). Os genes transcritos por essas regiões formam a região organizadora no nucléolo (NOR), e podem ser reconhecidas por análises cromossômicas citológicas como constrições secundárias na meiose/mitose (Hemleben V, Volkov RA, Zentgraf U 2004). Ao contrário do conhecimento limitado que temos sobre a maioria das sequências repetitivas, cujas funções ainda não estão bem definidas, os DNArs 5S e 35S são vitais para os organismos por fornecerem os rRNAs necessários para a montagem de ribossomos funcionais, que representam mais de 90% de todo o RNA transcrito de uma célula (Volkov et al. 2017). Nos eucariotos, a quantidade de cópias de DNAr excede o número necessário de RNA transcritos, por isso grande parte das cópias é silenciada transcricionalmente (Volkov et al. 2017).

Os elementos transponíveis, por sua vez, estão geralmente dispersos por todo o genoma e são classificados com base nos domínios enzimáticos conservados e estrutura (Pamponét et al. 2019). Retrotransposons (classe 1) e transposons de DNA (classe 2) são os dois tipos presentes nos genomas eucarióticos. Os retrotransposons são caracterizados por apresentarem RNA intermediário transcrito reversamente e um mecanismo de "*copy and paste*", enquanto os transposons possuem DNA intermediário e frequentemente um mecanismo de transposição "*cut and paste*", porém existem outros mecanismos a depender da família do retrotransposon (Finnegan 1989; Feschotte e Pritham 2007). Dentre os retrotransposons existem duas subclasses, aquelas flanqueadas por repetições terminais longas (RT-LTR) e aquelas que podem ou não ter repetições

terminais curtas (não RT-LTR) (Finnegan 1989; Gaiero et al. 2019). Os retrotransposons LTR-RT quando autônomos possuem pelo menos cinco domínios de proteína: GAG, protease (PROT), transcriptase reversa (RT), ribonuclease H (RH) e integrasse (INT). Esses retrotransposons estão agrupados em superfamílias e vários estudos mostram que as superfamílias Ty3/*gypsy* e Ty1/*copia* são as mais abundantes em plantas (Won et al. 2018; Said et al. 2018; González et al. 2018; Liu et al. 2019; Hloušková et al. 2019). Além dos RT-LTRT representarem tipicamente a maior fração da maioria dos genomas vegetais, há indícios de que seja o principal determinante da variação no tamanho dos genomas (Wessler 2006; Zuccolo et al. 2007; Macas et al. 2015), devido às novas cópias de elementos geradas que, após a integração, aumentam o tamanho do genoma (Galindo-González et al. 2017; Neumann et al. 2019). O tamanho genômico de espécies de plantas pode variar de 63 a 149.000 Mb (Heslop-Harrison and Schwarzacher 2011; Bennett and Leitch 2011).

Os retrotransposons comumente são encontrados dispersos pelo genoma, mas também podem estar concentrados nas regiões centroméricas e pericentroméricas de cromossomos de plantas. O grupo denominado de cromovírus, por exemplo, podem ser encontrados nos centrômeros. Esse grupo é formado pelos elementos Chlamyvir, Tcn1, Tekay, Reina, CRM e Galadriel pertencentes à superfamília Ty3/*gypsy* e possuem cromodomínios (CHD e CHDCR) (Hřibová et al. 2010; Neumann et al. 2019). A presença de cromodomínio nesse grupo é o que os diferencia do restante dos elementos denominados como não-cromovírus (Phygy, Selgy, Athila, Tat, Ogre, Retand)(Hřibová et al. 2010; Neumann et al. 2019).

Em estudos citogenéticos comparativos, a caracterização da distribuição física da heterocromatina em genomas de plantas utilizando técnicas de bandeamento cromossômico tem contribuído para diferenciação de espécies, especialmente em grupos com números cromossômicos estáveis (de Souza Almeida et al. 2007; da Costa Silva et al. 2015; Van-Lume et al. 2017). A caracterização da composição da heterocromatina pode ajudar a aprofundar as análises comparativas, identificando algumas classes de repetições que enriquecem as bandas heterocromáticas (Belyayev et al. 2018; González et al. 2018; Van-Lume et al. 2019; Peška et al. 2019). O conhecimento sobre a distribuição e abundância dessas sequências repetitivas (DNAsat, DNAr e TEs) pode contribuir para a compreensão sobre a evolução de genomas, como processos expansão e retração dos genomas, reestruturação de cromossomos (fusão, fissão, translocação e inversão), assim como processos de especiação e hibridização (Mehrotra e Goyal 2014;

Garrido-Ramos 2015). No entanto, observa-se que os tipos de sequências de DNA que compreendem bandas heterocromáticas ainda são pouco conhecidos na maioria dos estudos que exploraram a distribuição física da heterocromatina (Bolsheva et al., 2019; Gaiero et al. 2019). Portanto, não é claro se essas bandas localizadas em posições cromossômicas semelhantes dentro e entre as espécies são compostas de repetições semelhantes (Liu et al. 2019).

Considerando a proximidade filogenética entre *P. pubescens* e *P. emarginatus*, a análise citogenômica comparativa pode contribuir para o entendimento do relacionamento evolutivo destas espécies.

Objetivos

Objetivo geral

Realizar uma análise comparativa do tamanho do genoma, o padrão de bandeamento e a composição de elementos repetitivos nos cromossomos de *P. pubescens* e *P. emarginatus*, visando aprofundar os conhecimentos citogenômico e evolutivo destas espécies.

Objetivos específicos:

- Conhecer o tamanho genômico de *P. pubescens* e *P. emarginatus*;
- Selecionar as sequências repetitivas mais abundantes do genoma de *P. pubescens* para desenvolver sondas a hibridação *in situ* por fluorescência (FISH) e transferi-las para *P. emarginatus*;
- Determinar a distribuição física e composição da heterocromatina por meio da coração CMA/DAPI e pelo uso de sondas de regiões repetitivas para *P. pubescens* e *P. emarginatus*;
- Comparar os padrões de distribuição dos elementos repetitivos nos cariótipos de *P. pubescens* e *P. emarginatus*.

Manuscrito a ser submetido à publicação (**revista Planta: FI=3.347, Qualis A1**)

Cytogenomic analysis of *Pterodon pubescens* (Benth.) Benth. and comparative cytogenetics with *P. emarginatus* Vogel (Leguminosae, Papilionoideae) reveals similarity in their genomes repetitive fraction

Victória B Albernaz [1], Yennifer Mata-Sucre [2], Mariela A Sader [2], Pablo E Rodriguez [2], Cíntia P Targueta [1], Adriana M Antunes[1], Rafael B Pinto[1], Gustavo Souza [2], Thannya N Soares [1]

[1]Laboratory of Genetics and Biodiversity, Department of Genetics, Institute of Biological Sciences, Federal University of Goiás, Goiânia, Brazil.

[2] Laboratory of Plant Cytogenetics and Evolution, Department of Botany, Federal University of Pernambuco, Recife, Pernambuco, Brazil.

Corresponding author:

Thannya N Soares

Universidade Federal de Goiás

Instituto de Ciências Biológicas

Departamento de Genética

Laboratório de Genética e Biodiversidade

Estrada do Campus, s/n, Campus Universitário

74690-900 Goiânia GO Brasil

Tel: + 55 62 3521-1752, +55 62 3521-1731

e-mail: tnsoares@ufg.br

Main Conclusion: Comparative analysis of *P. pubescens* and *P. emarginatus* reveals highly similar chromosomal repeats distribution. However, expansion of Ty1/*copia*-Ale in *P. pubescens* euchromatin represents a genomic mark that differentiates their karyotypes.

The genus *Pterodon* presents currently four species, of which *P. pubescens* and *P. emarginatus* (both known as “sucupira branca") are more morphologically and phylogenetically related. These species have a wide geographical distribution in Brazil and both present the karyotype 2*n* = 16 with small and morphologically similar metacentric/ submetacentric chromosomes. The objective of the present work was to characterize the repetitive fraction of *P. pubescens* by genome skimming, aiming to perform a comparative cytogenomic analysis between this species and *P. emarginatus*. The species presented similar karyotypes with: (i) CMA^{+}/$DAPI^{-}$ bands in the terminal region of two chromosome pairs and proximal region of all chromosomes; (ii) two pairs of 35S rDNA sites co-localized with the terminal CMA^{+} bands; and (iii) one pair of 5S rDNA sites located in the proximal region of one chromosomal pair. The genome size of *P. pubescens* was similar to *P. emarginatus* - 1C = 0.665 pg and 1C = 0.620 pg, respectively. The repetitive fraction represented 26.4% of the *P. pubescens* sequenced genome, with Ty3/*gypsy*-Athila (24.24%), Ty3/*gypsy*-Tekay (21.93%), Ty3/*gypsy*-Ogre (6.72%) and Ty1/*copia*-Ale (3.37%) retrotransposons being the most abundant elements. Additionally, low abundance of satellite DNAs were identified: PubSat1-254 (2.09%), PubSat2-76 (2.06%), PubSat3-216 (0.58%), and PubSat4-138 (0.23%). *In situ*

hybridization revealed that all analyzed repeats were enriched in proximal CMA^+ heterochromatin in both species, except the Ale retroelement, which was dispersed also in the euchromatin of *P. pubescens* chromosomes. The cytomolecular similarity observed here shows that the genomes of *Pterodon* species share highly similar repetitive fractions, which corroborates their phylogenetic proximity. However, the Ty1/*copia*-Ale dynamic reinforce the idea that different families of LTR-RTs have independent activity on genomes and chromosomes being a genomic mark that differentiates karyotypes from *P. pubescens* and *P. emarginatus* and corroborate the delimitation of both species.

Keywords: FISH, NGS, satDNA, repetitive sequences, "sucupira branca"

Abbreviations

CMA: Chromomicin A3

Cy3-dUTP: 5-amino-propargyl-2′-deoxyuridine 5′- triphosphate coupled to red cyanine fluorescent dye

DAPI: 4',6-diamidino-2-phenylindole

FISH: Fluorescent *in situ* hybridization

NGS: Next-generation sequencing

rDNA: Ribosomal DNA

TEs: Transposable elements

RT: Retrotransposons

LTR: Long terminal repeat

Introduction

Repetitive DNA sequences represent large fractions of plant nuclear genomes, which can be divided into two large groups (e.g. de Souza et al. 2018; Bolsheva et al. 2019). The first group consists of elements dispersed throughout the genome, called transposable elements (TEs). There are two classes of TEs: retrotransposons (class I) and transposons (class II) (Finnegan 1989; Bourque et al. 2018; Won et al. 2018). Retrotransposons (RTs) are divided into two subclasses; those flanked by long terminal repeats (LTR-RT) and those that may or may not have short terminal repeats (non-LTR-RT). LTR retrotransposons Ty3/*gypsy* and Ty1/*copia* are the most abundant superfamilies in plants nuclear genome (Won et al. 2018; Said et al. 2018; González et al. 2018; Liu et al. 2019; Hloušková et al. 2019). The second repetitive group consists of tandem-organized sequences (as satellite DNA), with copies located adjacent to each other, and predominantly concentrated in functionally important regions in the chromosomes: centromeres, telomeres, and sometimes chromosomal interstitial regions (e.g. Garrido-Ramos 2015; Belyayev et al. 2018; Bolsheva et al. 2019).

Repetitive elements of the genome can modify the karyotype patterns and plant genome through genetic changes generated mainly by the activity of TEs (Oliver et al. 2013; Horváth et al. 2017). In general, the repetitive elements are enriched in heterochromatic regions (González et al. 2018; Li et al. 2019; Samoluk et al. 2019) and it has been observed that TEs expansion have a high impact on genome size and that it can cover half, or even more, of the total amount of nuclear DNA of many plant (Fu et al. 2019), (Schnable et al. 2009), (Oliver et al. 2013) and (Mehra et al. 2015).

In addition to classical cytogenetic parameters (Zoldos et al. 1999; Carvalho and Guerra 2002; Pinto et al. 2016), TEs have been used in comparative cytogenomic studies to understand inter and intraspecific variability in plants (Deng et al. 2019; Tomás et al. 2016; Samoluk et al. 2019 ; Bolsheva et al. 2019; Van-Lume et al. 2019;

Siljak-Yakovlev et al. 2003; Almeida et al. 2007; Viana and Souza 2012; da Costa Silva et al. 2015; Van-Lume et al. 2017; Ortiz et al. 2017; Mata-Sucre et al. 2020).)

Comparative cytogenetic studies have usually been performed using chromosome banding (e.g. C-banding and double staining with the fluorochromes chromomycin A3 [CMA] and 4', 6-diamidino-2-phenylindole [DAPI]) and/or fluorescent *in situ* hybridization (FISH) with ribosomal DNA (rDNA) probes, respectively, evaluating the distribution of heterochromatin and the number and position of ribosomal loci along the chromosomes in plant karyotypes. These techniques are powerful for comparative cytogenetic studies, especially for plant groups with slight differentiation in chromosome morphology and stable chromosome numbers (Siljak-Yakovlev et al. 2003; De Souza Almeida et al. 2007; Viana and Souza 2012; da Costa Silva et al. 2015; Van-Lume et al. 2017; Ortiz et al. 2017; Mata-Sucre et al. 2020). However, many studies involving close related species were not able to distinguish karyotypes by these conventional banding techniques (Zoldos et al. 1999; De Carvalho and Guerra 2002). NGS sequencing technologies has recently emerged as a versatile source of sequencing genomic data, specially enabling the genome skimming approach (Dodsworth et al. 2019). Hence, the advent of genome skimming allowed coupled with the development of bioinformatic analysis (e.g. RepeatExplorer, Novák et al. 2013) has integrated genomics and cytogenetics (cytogenomics) allowing to determine the composition of the repetitive fraction of the genome and the association of repetitive elements with heterochromatin

Pterodon Vogel is a legume genus of the subfamily Papilionoideae which can be distinguished from other members of the tribe Dipterygeae by conspicuous morphological features. Among these, we can highlight the cryptosamara-shaped fruit with oil glands on the epicarp, the vented rachis leaf, the seed testa smooth, the apparent

raphe with the hilum in the lateral position covered by an aryl, and a smooth embryo (Leite et al. 2014; Pinto et al. 2014). The species of the genus presents medicinal potential through the oil extracted from its seeds (Oliveira et al. 2016; Faria et al. 2017; Outuki et al. 2018). The genus is currently composed by only four species: *P. abruptus* (Moric.) Benth, *P. apparicioi* Pedersoli, *P. pubescens* (Benth.) Benth. and *P. emarginatus* Vogel. The last two are not endemic to Brazil, nor Brazilian savannah, but occur in Bolivia and are widely distributed throughout Brazil, occurring in the states of Rondônia, Tocantins, Bahia, Maranhão, Piauí, Goiás, Distrito Federal, Mato Grosso, Mato Grosso do Sul, Minas Gerais and São Paulo (Flora do Brasil 2020, constantly updated).

Despite the easy differentiation when species are blooming, for many years it was difficult to delimit the four species of the genus. Over the years, this genus has taxonomic inconstancies, showing different species numbers varying from six, three and two species depending on the database (Rocha 2006). This taxonomic discrepancy is mostly due to the delimitation of *P. pubescens* and *P. emarginatus*, as they have been considered as one species, with two morphotypes (Rocha 2006). Apparently, there is no hybridization between natural populations of *P. pubescens* (with pink flowers and hairy leaves) and *P. emarginatus* (with purple flowers and glabrous leaves), although three individuals carrying intermediate morphological characteristics were observed (Rocha 2006). The possibility of natural hybridization may be linked to the recent divergence of the species (Lima 2019), and can be better studied with additional tools to classical taxonomy, such as cytogenetics and genomics approaches (see Chase et al. 2003; Chester et al. 2010; Da Silva et al. 2017; Marques et al. 2018; Silva et al. 2018) .

Pterodon has been poorly studied from a cytogenetic point of view. The chromosome number is the only cytogenetic feature described for the species *P.*

pubescens and *Pterodon polygalaeflorus* (Benth.) Benth. (= *P. emarginatus*), both having 2*n* = 16 and presenting small chromosomes (Bandel 1974; Coleman and DeMenezes 1980). Therefore, the two *Pterodon* species, *P. pubescens* and *P. emarginatus*, represent an interesting case to test comparative cytogenomic to evaluate further genomic characteristics that can support the circumscription of both species.

Thus, we aim to analyze the abundance and composition of repetitive elements in *P. pubescens* and perform a comparative cytogenetic analysis between this species and *P. emarginatus*. Here we address three questions: (1) Do the usual cytomolecular techniques (CMA/DAPI banding, 5S and 35S rDNA and genome size) allow distinguishing the two species karyotypes? (2) What kind of repetitive elements are located along the chromosomes of these species? (3) How similar is the heterochromatic composition of these genomes?

Material and Methods

Genome size estimative

The nuclear DNA content of *P. pubescens* e *P. emarginatus* was estimated by flow cytometry. Seeds of *P. pubescens* were collected in Goiânia (LAT -16.577/ LON -49.273, the plant material collected was deposited at the Herbarium of the Federal University of Goiás - Voucher: 61013) and *P. emarginatus* of Alvorada do Norte (LAT -14.586/ LON -46.609), Goiás, Brazil. All seeds were germinated, grown, and maintained in the experimental garden of the Laboratory of Plant Cytogenetics and Evolution, Recife, Pernambuco, Brazil. Young leaves were collected and nuclei suspensions were prepared using the one-step protocol adapted from Doležel et al. (2007). Approximately 25–50 mg of leaf tissue was chopped with a razor blade on a

Petri dish (kept on ice) containing 1 mL of isolation buffer Marie (Marie and Brown 1993). The solution was filtered through a 30 mm mesh filter and mixed with 50 μL propidium iodide (1 mg / mL) for nuclei staining. The absolute DNA amount of a sample was calculated based on the values of the G1 peak means: "[2C sample DNA content (pg)] = [(sample G1 /standard G1)] x [2C standard DNA content (pg)]", where "G1" refers to the mean fluorescence value emitted by nuclei in the G1 stage of interphase and "2C standard" refers to the absolute DNA content of the internal standard used in the measurement (Doležel and Bartoš 2005). Leaves of *Solanum lycopersicum* L. was used as internal standard (Doležel et al. 1998).

Flow cytometry measurements were performed using a Cyflow Space flow cytometer (SYSMEX, Norderstedt, Germany) equipped with a green laser (532 nm). Three independent estimates of DNA using three individuals were performed on different days. Relative fluorescence histograms were analyzed using FloMax 2.3 software (PARTEC, Muster-Germany), and measurements that generated peaks with a coefficient of variation (CV) greater than 5% were discarded.

Repeat Identification and Annotation

The DNA used for Next-Generation Sequencing (NGS) was extracted from leaves of *P. pubescens* collected in Goiânia (LAT -16.577/ LON -49.273, the plant material collected was deposited at the Herbarium of the Federal University of Goiás - Voucher: 61013), using the cetyltrimethylammonium bromide (CTAB) method, described by Doyle and Doyle (1987). NGS of the *P. pubescens* genome was done using Illumina platform. DNA libraries were constructed using Nextera DNA Library prep kit and sequencing was performed on Illumina MiSeq using 600-cycle Miseq Reagent v3 kit (2× 250bp, paired reads). A fasta file containing 3,337,638 reads was used as input in

cluster analysis with the RepeatExplorer pipeline (Novak et al. 2013, http://repeatexplorer.umbr.cas.cz/). In order to identify and classify repetitive sequences in the *P. pubescens* genome, paired reads were analyzed using graph-based comparative clustering. Reads were assigned to the same clusters if they shared 90% similarity in a 55% minimum sequence overlap. The clusters were classified based on similarity search in Viridiplantae version 3.0 database implemented in RepeatExplorer. The TAREAN tool was used to identify and characterize satellite DNA sequences (Novak et al. 2017). Clusters that potentially had satellite DNA sequences were examined manually using the Dot-plot genomic similarity tool in Geneious Prime 2019.2.1 (https://www.geneious.com) (Kearse et al. 2012). Protein domains were identified using NCBI's conserved domain database (Derbyshire et al. 2015), to identify putative TE domains elements.

Primers design, PCR-amplification of satellites DNAs and retrotransposons families

Primers were designed using the Primer 3 tool (Untergasser et al. 2012) included in Geneious version 7.2.1. (Kearse et al. 2012). For satellite DNA, primers were designed based on K-mer information from most conserved regions and in facing towards leaving a maximum of 4-5 bp distance between forward and reverse (Table 1). In addition, three LTR-RT (Ty3/*gypsy*-Tekay, Ty3/*gypsy*-Athila and Ty1/*copia*-Ale) specific primers were designed for the integrase domain. All primers had produced successfully amplified products in PCR reaction (Online Resource 1), and a DNA fragment of expected size from each region were used for sequence confirmation by Sanger sequencing on the ABI-3500 platform (Applied Biosystems).

For the PCR-amplification, the DNA used was extracted from leaves of *P. pubescens* collected in Pontalina (LAT -17.520 / LON -49.444), Goiás, Brazil. The

samples were stored in freezer at -80 °C and subsequently the DNA was extracted using the cetyltrimethylammonium bromide (CTAB) method described by Doyle and Doyle (1987). PCR reactions were carried out with final concentration of 20 ng DNA, 0.4 μM forward and reverse primers, 0.1 μM from each dNTPs, 1x PCR buffer, 2 μM $MgCl_2$, 0.025 U/μL Taq polymerase and ultrapure water to complete the final volume of 25 μL. DNA amplification was performed on the Applied Biosystems® Veriti® thermal cycler, which was programmed for PCR reactions of: 94°C for 3 min, 30 × (94°C 1 min, 55°C 1 min, 72°C 1 min) and 72°C for 10 min. For probe synthesis, amplification tests were used to determine the best annealing temperature for *P. pubescens.* The probes were produced and transferred for *P. emarginatus*.

Slide preparation and chromosome banding

Seeds of *P. pubescens* were collected in Goiânia (LAT -16.577/ LON -49.273) and seeds of *P. emarginatus* were collected in Alvorada do Norte (LAT -14.586/ LON -46.609) and Planaltina (LAT -15.600/ LON -47.658), Brasília-DF, Brazil. The material collected in Goiânia and Planaltina were deposited at the Herbarium of the Federal University of Goiás (Voucher: 61013 and 68411, respectively). Seeds of these two species are enclosed by the woody wrap of the fruit and a layer of oily glands. Sowing is most effective when fruits are manually opened, but the germination rate is still low. After removal of this woody wrap, seeds were washed once with a 4% hypochlorite solution for 10 min, and twice with distilled water for 5 min. Root tips obtained from germinated seeds were pretreated with 2 mM 8-hydroxyquinoline for 24 h at 10 °C, fixed in ethanol-acetic acid (3:1 v/v) and stored at -20 °C.

Root tips were digested using a solution containing 2% cellulase and 20% pectinase (w/v) for 120 minutes at 37 °C and mitotic preparations were performed

according to Carvalho and Saraiva (1993). The CMA/DAPI double staining technique was used for fluorochrome banding, following Souza et al. (2015). The slides were stained with CMA (0.1 mg/mL) for 30 min and mounted in DAPI (2 µg/mL): McIlvaine pH 7.0 buffer with glycerol (1:1) and aged for 3 days before analysis in a Leica DMLB epifluorescence microscope. Metaphases cells images were captured with a Cohu CCD video camera using Leica QFISH software and were optimized for better contrast and brightness using Adobe Photoshop CS6.

In situ *hybridization*

PCR products were labeled by *Nick translation* (Invitrogen or Roche Diagnostics) with Cy3-dUTP (GE). Labeled probes were used to locate repetitive elements by FISH technique following the protocol described by Ribeiro et al. (2017). The hybridization mixture consisted of 50% (v/v) formamide, 10% (v/v) dextran sulfate, 2× SSC and 5 ng/mL of each probe. The slides were denatured at 75 °C for 5 min. The final stringency was of approx. 76% and some hybridizations were made with low stringency. Slides were counterstained in 2 µg/mL DAPI in Vectashield (Vector). Images of the best cells were captured as described above. The *Lotus japonicas* (Regel) K. Larsen D2 (Pedrosa et al. 2002) and *Triticum aestivum* pTa71 (Gerlach and Bedbrook 1979) probes, labeled with Cy3-dUTP and digoxigenin-11-dUTP (Roche), were used to locate 5S and 35S rDNA sites, respectively.

Ty1/copia-Ale Sequencing

Since repeat Ty1/*copia*-Ale is the most divergent in the cytogenomic comparison (see Results), a sequencing approach was used to estimate the integrase sequence variability of the Ty1/*copia*-Ale elements in natural populations of *P. emarginatus* and *P.*

pubescens. Sanger sequencing was performed on the ABI-3500 platform (Applied Biosystems) including 19 populations for *P. pubescens* and 11 for *P. emarginatus* [one individual per population was used] encompassing a wide distribution of species (Online Resource 2). Each individual sequence (forward and reverse) was analyzed in software BioEdit version 7.0.5.3 (Hall 1999), then a consensus sequence was assembled for each individual and all sequences were aligned using ClustalW.

Estimating the age of **Pterodon**

To estimate the age of *P. emarginatus* and *P. pubescens* divergence, we performed a molecular phylogeny of the clade Dipterygeae based on the *mat*K gene, the plastid spacer *trn*L–*trn*F, and the ITS (ITS1– 5.8S–ITS2) of nuclear rDNA. We sampled three of the genus *Pterodon*, eight species of the genus *Dipteryx* and as outgroup *Taralea rigida*. All sequences were obtained from GenBank. All the sequences were aligned using MUSCLE (Edgar 2004) as a plugin in Geneious v.9.1.8 (Kearse et al. 2012) with subsequent manual adjustments. We used jModelTest v.2.1.6 to assess the best model of DNA substitution for each individual locus (Darriba et al. 2012) through the Akaike information criterion (Akaike 1974). The best fitting model was GTR for all loci. Phylogenetic relationships were inferred using the BI approach implemented in MrBayes v.3.2.6. (Ronquist et al. 2012). All analyses were performed for each region separately. Four independent runs with four Markov Chain Monte Carlo (MCMC) runs were conducted, sampling every 1000 generations for 3 000 000 generations. Each run was evaluated in Tracer v.1.6 (Rambaut et al. 2014) to determine that the estimated sample size (ESS) for each relevant parameter was >200, and a burn-in of 25 % was applied. The majority rule consensus tree and posterior probability (PP) were visualized and edited in FigTree v.1.4.2 (Rambaut 2014).

Divergence time estimates were performed in BEAST v.1.8.3 (Drummond and Rambaut 2007; Drummond et al. 2012) fixing the tree topology of the Bayesian analyses. An uncorrelated relaxed lognormal clock (Drummond et al. 2006) and a Yule Process speciation model (Gernhard 2008) were applied. Two independent runs of 10,000,000 generations each were performed, sampling every 10,000 generations for consensus *mat*K+ *trn*L–*trn*F+ ITS loci. In order to verify the effective sampling of all parameters and assess the convergence of independent chains, we examined their posterior distributions in Tracer v.1.6, and the MCMC sampling was considered sufficient at an ESS >200. After removing 25% of samples as burn-in, the independent runs were combined and a maximum clade credibility (MCC) tree was constructed using TreeAnnotator v.1.8.2. (Drummond et al. 2012). Calibrations were performed using the secondary calibrations of Kumar et al. (2017) for the *Pterodon*/*Dipteryx* divergence approx. 4.21 million years ago (Mya).

Results

Genome size and characterization of the repetitive genomic fraction of P. pubescens

Considering the genome size of *P. pubescens* as 1C = 650.37 Mbp (0.665 pg) (Online Resource 3) we obtained a 0.2× coverage. A total of 266 clusters, containing 135 to 9,669 reads, and corresponding to at least 0.01% of the genome were used for annotation of moderately to highly abundant repetitive elements. The repetitive fraction represented 26.4% of the total genome sequenced. The analysis revealed that the retrotransposons lineages Ty3/*gypsy*-Athila (24.24%), Ty3/*gypsy*-Tekay (21.93%) and Ty1/*copia*-Ale (3.37%) were the most abundant of Ty3/*gypsy* and Ty1/*copia* superfamilies, respectively (Fig. 1, Table 1). A total of 2.55% of the repetitive fraction was represented by 5S and 35S rDNA. Four clusters showed a characteristic ring graph

pattern (Online Resource 4), which is used to classify the repeat families as DNA satellites. These satellite DNA showed an abundance of 4.96% in the genome and were named PubSat1-254, PubSat2-76, PubSat3-216 and PubSat4-138 (Fig. 1, Table 1).

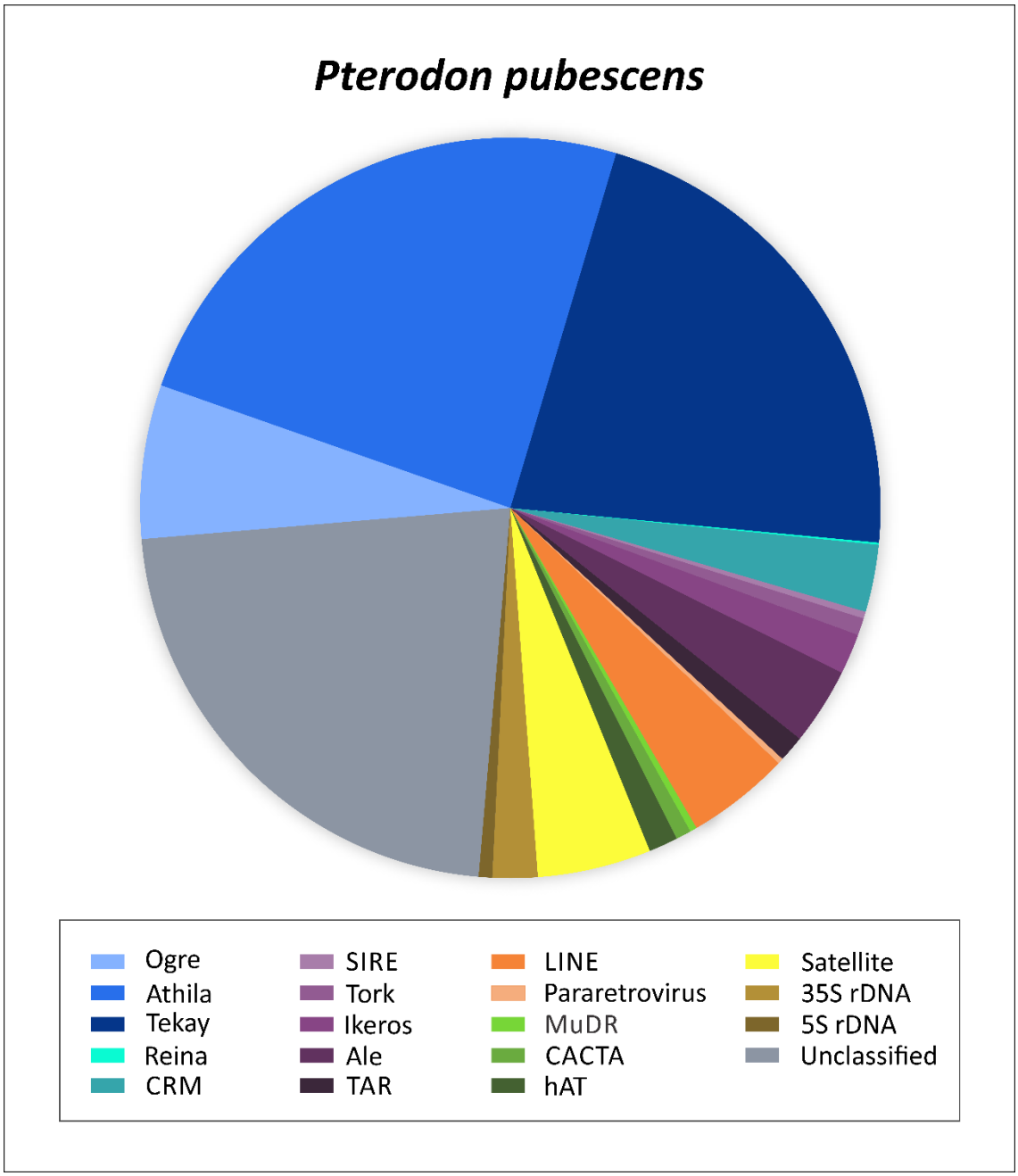


Figure 1: Relative abundances of different kinds of repetitive DNA sequences in the genome of *P. pubescens*, based on the graph-based clustering analysis

Table 1. Information of primers sequences designed for FISH and PCR amplification of Satellite DNAs and TE fragments identified in the genome of *P.pubescens*

Analysis type	Name	Monomer/ FS (pb)	Genome proportion	Classification	Primers Forward/Reverse
TAREAN	PubSat1-254	254	2.09%	Satellite	GGGCGAAAGTTATCAACCTA TGGTTGATAACTGCATGTGA
TAREAN	PubSat2-76	76	2.06%	Satellite	GTTGGGGCTTGAAAAGAAAA CCCCCATTTTCATTGTTCTG
TAREAN	PubSat3-216	216	0.58%	Satellite	ACTTGTGTCTTCTGGTTGAT ACACACCAGTAGCAGAAAAT
TAREAN	PubSat4-138	138	0.23%	Satellite	CCAACCTTCCAAGTAAGCAA TTGCACACACTACAACTTGT
Protein domains tools	Ty3/*gypsy*-Athila	329	24.24%	Ty3/*gypsy*-Athila	CATTTGTCTGAGGGTGGTAA GGGGTATTGATTTCATGGGT
Protein domains tools	Ty3/*gypsy*-Tekay	215	21.93%	Ty3/*gypsy*-Tekay	GTGAAGGCTGAACATCAAAG TGGGCATACTGAGATACACT
Protein domains tools	Ty1/*copia*-Ale	242	3.37%	Ty1/*copia*-Ale	AGCAGAGTCTCCTTTTTCAA CGCAAAATACGAATGGTCTT

FS = fragment size or monomer obtained after PCR for probe used in FISH experiments

Comparative genome size, CMA/DAPI and rDNA mapping

The genome sizes of *P. pubescens* and *P. emarginatus* were similar, with 1C = 0.665 pg and 1C = 0.620 pg, respectively (Online Resource 3). *P. pubescens* and *P. emarginatus* showed 2*n* = 16 with chromosome morphology very similar, predominantly meta/submetacentric. Only one acrocentric pair (VIII) was observed in *P. pubescens* and two (VII and VIII) in *P. emarginatus* (Fig. 2a-b). The $CMA^+/DAPI^-$ bands were observed in the terminal region of two chromosome pairs (I and VIII) and in the pericentromeric regions of all chromosomes (Fig. 2). In *P. pubescens*, the chromosome pairs III and V showed CMA^+ band extending along most of the long and short arms, respectively (Fig. 2a). One pair of 5S rDNA site, located in the proximal region of pair III, and two pairs of 35S rDNA, located in the terminal region of pairs I and VIII, were observed in both species (Fig. 2). This site on pairs I and VIII were colocalized with terminal $CMA^+/DAPI^-$ bands (Fig. 2a-b).

Chromosomal distribution of repeats

The three RT probes (Ty3/*gypsy*-Tekay, Ty3/*gypsy*-Athila and Ty1/*copia*-Ale) were *in situ* hybridized and showed labels in the chromosomes of both species (Fig. 2). In general, the retrotransposons were enriched in centromeric and pericentromeric CMA^+ heterochromatin on most chromosomes (Fig. 2). For instance, in *P. pubescens* and *P. emarginatus*, Ty3/*gypsy*-Tekay signals were concentrated in the proximal region of all chromosomes, with weak labelling on chromosome pair VI, which also showed a weaker CMA^+ band (Fig. 2).

In situ hybridization of Ty3/*gypsy*-Athila revealed small dot-like signals in the proximal CMA^+ heterochromatin (Fig. 2). The short arm of the acrocentric pair VIII

showed a stronger Ty3/*gypsy*-Athila label, revealing a higher richness of this element in this chromosomal pair on both species and colocalized with 35S rDNA signal. In *P. emarginatus*, absence of Ty3/*gypsy*-Athila signals was observed in chromosomal pair VI (Fig. 2b). The chromosomal distribution of Ty1/*copia*-Ale was the most dissimilar between the karyotypes of *P. pubescens* and *P. emarginatus*. As expected based in the chromosomal location of other repeats, *P. emarginatus* showed a Ty1/*copia*-Ale distribution restricted to the centromeric and pericentromeric region of all chromosomes (Fig. 2b). On the other hand, in *P. pubescens* chromosome pairs I, II and VIII were completely labeled, revealing a expansion of the labeling for the euchromatin (Fig. 2a). Chromosomal pairs VI, characterized by low amount of proximal CMA^+ heterochromatin, also showed weak or absent Ty1/*copia*-Ale signals (Fig. 2a). We investigated whether this expansion of Ty1/*copia*-Ale in *P. pubescens* was accompanied sequence diversification. Despite the difference in the distribution of Ty1/*copia*-Ale, the sequencing of the Ty1/*copia*-Ale element in a wide populational sample including individuals of both species did not reveal any polymorphisms, indicating that the expansion of these elements in the *P. pubescens* genomes was not accompanied by incremental sequence-level variability, at least, in the integrase domain. Comparative *in situ* hybridizations have been performed using satellite probes too, and the only two showed signals: DNA PubSat2-76 and PubSat4-138 (Fig. 3). Like RTs, satellite DNA were enriched in CMA^+ pericentromeric heterochromatin in both species (Fig. 3).

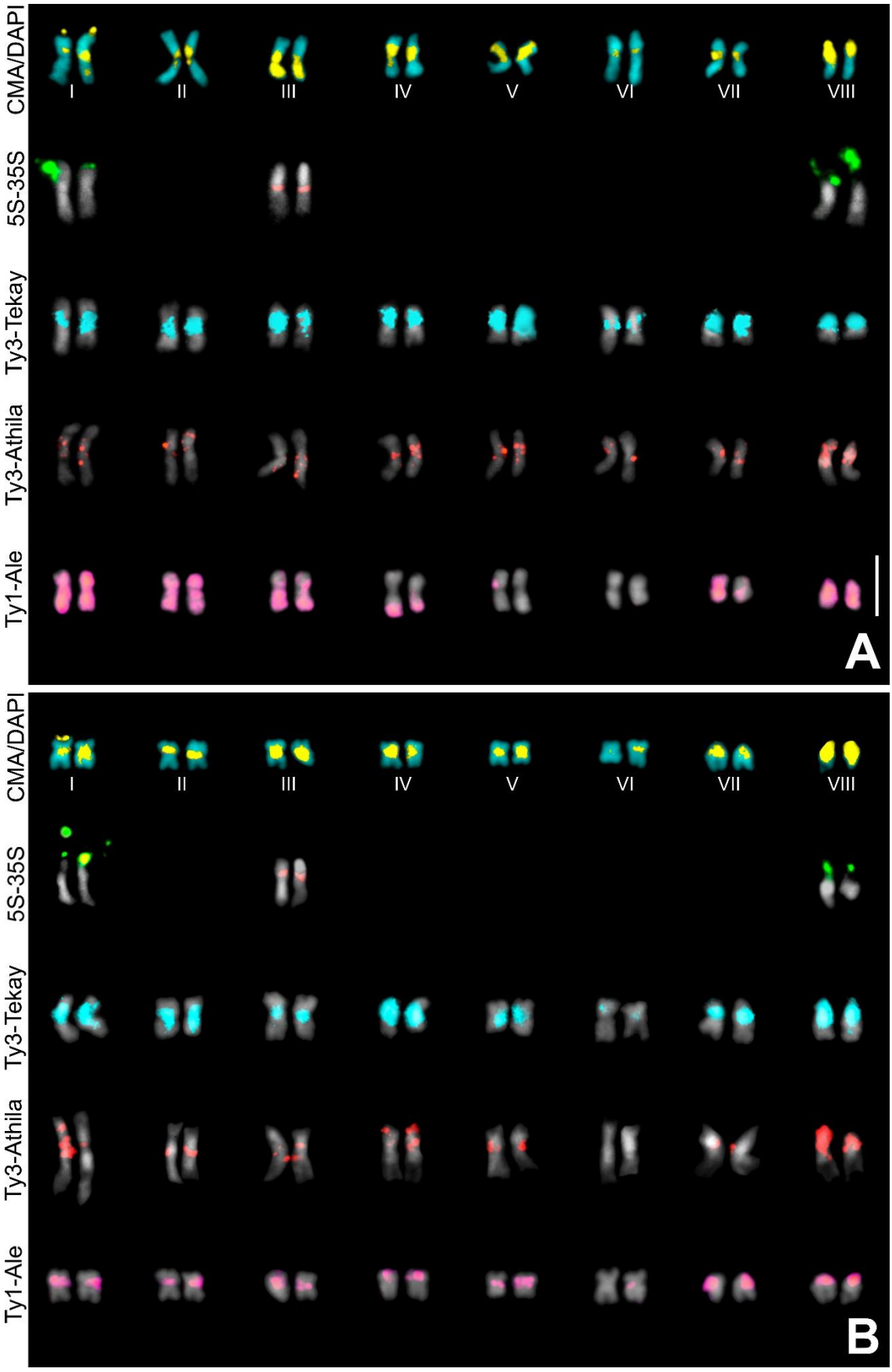

CMA/DAPI
5S-35S
Ty3-Tekay
Ty3-Athila
Ty1-Ale
I
II
III
IV
V
VI
VII
VIII
A
B

Figure 2: Karyograms showing the relative chromosome length, position of centromeres, CMA/DAPI banding, *in situ* hybridization with 5S (red) and 35S (green) rDNA and three retrotransposons (Ty3/*gypsy*-Athila, Ty3/*gypsy*-Tekay, Ty1/*copia*-Ale) probes in *P. pubescens* (A) and *P. emarginatus* (B). Chromosome labelled with different probes are from different metaphase cells. Bar in A represents 5 μm

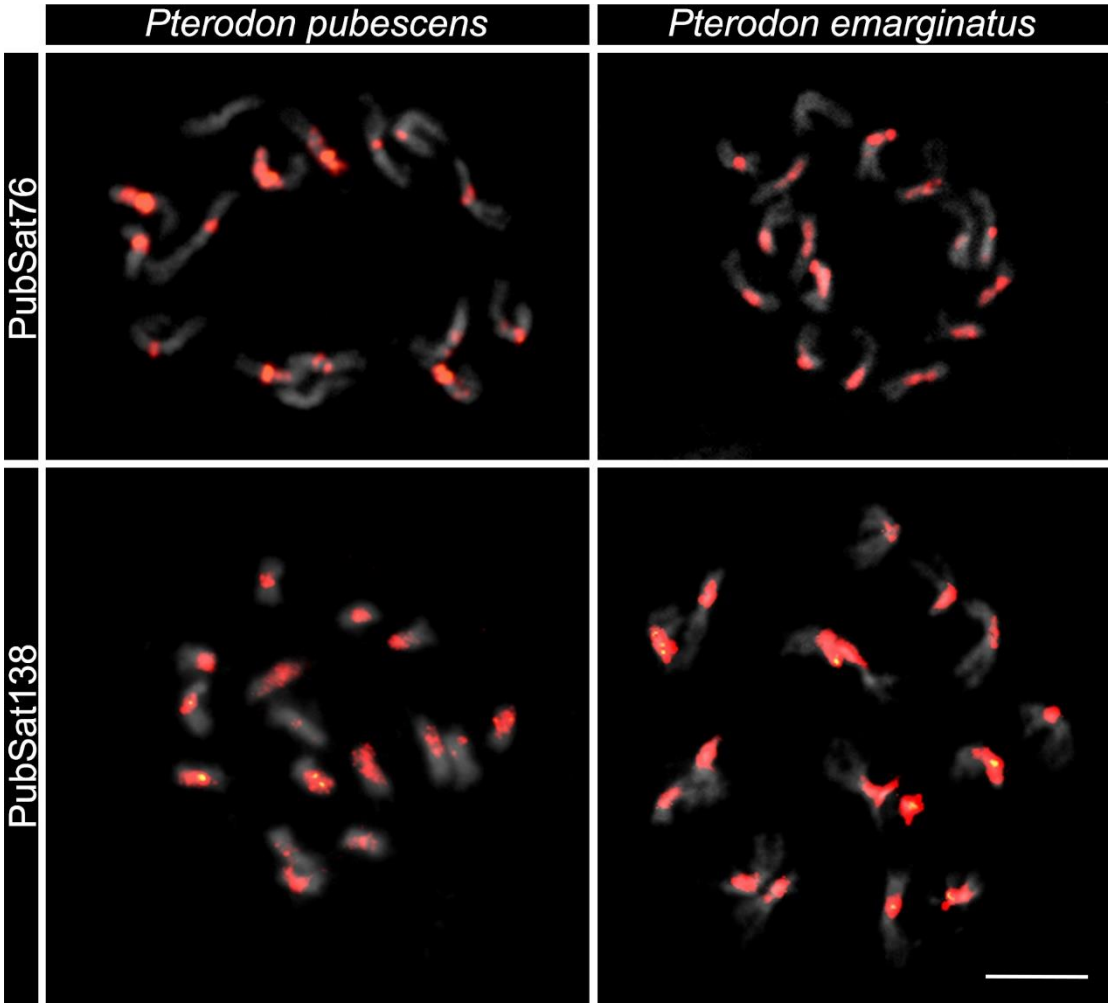


Figure 3: Distribution patterns of DNA satellite PubSat2-76 and PubSat4-138 on *P. pubescens* and *P. emarginatus*. The satellite DNA sequences were labeled with Cy3 (red signal) and the chromosomes were counterstained with DAPI. Bar represents 5 μm

Estimating the age of Pterodon

Because we were interested in having an estimation of the time of diversification of P. pubescens and P. emarginatus. We performed molecular dating with BEAST using ITS

and plastid trnL–trnF regions, with an estimed time divergence at approx.. in 0.905 Mya (0.519 - 1.593 Mya) (Online Resource 5).

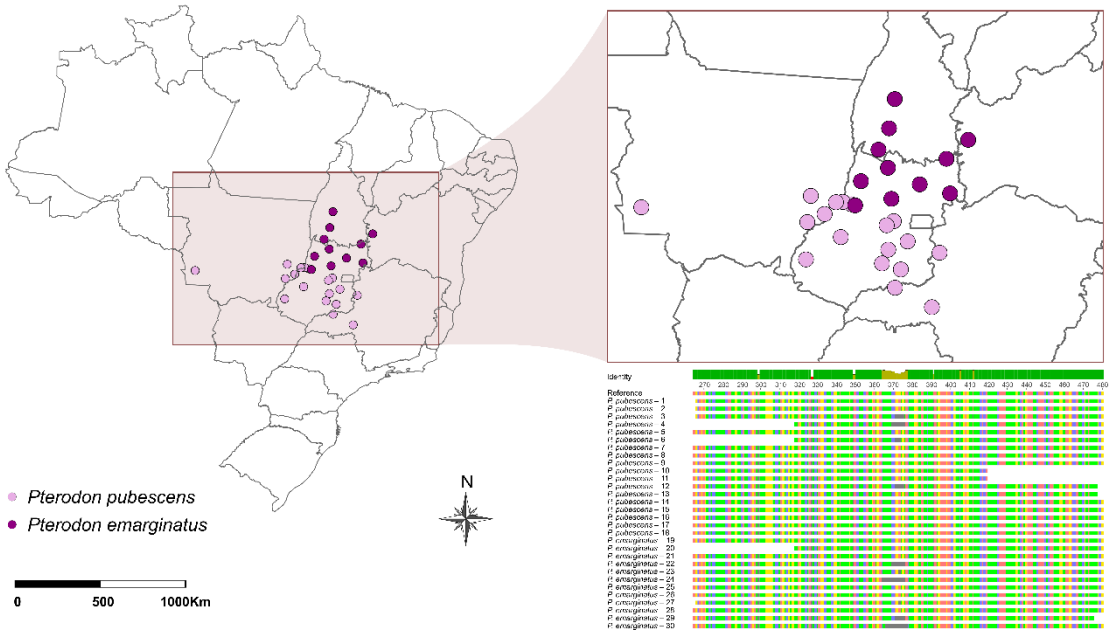


Figure 4 Sanger sequences of integrase region of retrotransposon Ty1/*copia*-Ale from native populations of *Pterodon pubescens* and *P. emarginatus*. Distribution map of the populations of *P. pubescens* and *P. emarginatus*

Discussion

Karyotypic characterization of the clade Dipterygeae

Our results revealed a high karyotype similarity between the analyzed species of the genus *Pterodon*. The previously reported karyotype 2*n* = 16, with small and undifferentiated chromosomes (Bandel 1974; Coleman and DeMenezes 1980) was confirmed here. This same chromosome number was reported also for *Dipteryx alata* Vogel and *D. magnifica* Ducke (Taquary 2017; Madrigal 2018), which together with *Pterodon*, *Monopteryx* Spruce ex Benth. and *Taralea* Aubl. compose the clade Dipterygeae (Cardoso et al. 2013). The genome sizes reported here for *P. pubescens* and *P. emarginatus* (1C = 0.665 pg and 1C = 0.620 pg, respectively) is about half of the

genome size of *P. abruptus* (1C = 1.300 pg, see Madrigal 2018) and can be also an indicative of phylogenetic proximity. This similarity could be also related to the recent diversification of this species, estimated here (Online Resource 5). In fact, *P. pubescens* and *P. emarginatus* seems to present the smallest genome size within the tribe so far, since the genera *Monopteryx* (1C ~ 1.125 pg), *Taralea* (1C ~ 1.255 pg) and *Dipteryx* (1C ~ 1.5605) presented higher values reported (Madrigal 2018). This difference suggests that *P. pubescens* and *P. emarginatus* suffered a strong genomic size reduction through evolutionary process.

The usual cytomolecular parameters do not differentiate* P. pubescens *and* P. emarginatus *karyotypes

Although chromosomal banding and *in situ* hybridization with rDNA probes have proved to be useful in differentiating plant karyotypes (de Souza Almeida et al. 2007; Barros and Silva et al. 2010; Van-Lume et al. 2017), our results do not allow a differentiation between *P. pubescens* and *P. emarginatus* with those markers. This may be related to the phylogenetic proximity of these species (Cardoso et al. 2013). The stability in legume karyotypes can be associated with the timing of diversification, reproduction strategies, and other factors related to the evolutionary history of each group (Soltis and Soltis 2009; Gu et al. 2016; Franco et al. 2019). Genome comparisons have shown conserved syntenic blocks between papilionoid genomes, especially among phylogenetically closely related species (Young and Bharti 2012). Groups of species with a recent diversification usually show evidences of strongly conserved karyotypes (Pinto et al. 2016). This stability of karyotypes found here was also observed in some groups of recent divergence was also identified in other papilionoid groups (Naganowska et al. 2005; Pinto et al. 2016; Susek et al. 2016; Zheng et al. 1993;

Almeida and Pedrosa-Harand 2013; Fonsêca and Pedrosa-Harand 2017; Franco et al. 2019).

***Retrotransposons and satellite DNA are enriched in the proximal heterochromatin of* P. pubescens *and* P. emarginatus**

The advent of NGS technologies has revolutionized cytogenetics, with much impact on the characterization of the heterochromatin (Liu et al. 2019; Bolsheva et al. 2019). The classic view that heterochromatic bands are composed of satellite DNA (Barnes et al. 1985; Pich et al. 1996; Barros and Silva et al. 2010; Kirov et al. 2017) has changed with cytogenomic data, which has revealed that heterochromatin is actually much more complex and heterogeneous in terms of composition (Peška et al. 2019; Hufnagel et al. 2019; Van-Lume et al. 2019). The literature indicates that retrotransposons generally accumulate in blocks or can be found dispersed across plant chromosomes, unlike satellite DNAs that form more defined blocks (Heslop-Harrison and Schmidt 2012; Santos et al. 2015; Ribeiro et al. 2017; Zhang et al. 2017; de Souza et al. 2018). We identified here by FISH the presence of three RTs (Ty3/*gypsy*-Tekay, Ty3/*gypsy*-Athila and Ty1/*copia*-Ale), two satellite DNA and 5S rDNA in the proximal heterochromatin of *P. pubescens* and *P. emarginatus*.

Concerning the RTs, the occurrence of Ty3/*gypsy* elements in the centromeric and pericentromeric regions of plant chromosomes has often been reported, particularly from those containing chromodomains (Neumann et al. 2011, 2019). A recent phylogenetic analysis of the protein-coding regions of LTR retrotransposons showed that such chromovirus-type Ty3/*gypsy* elements could be subdivided into six clades named Chlamyvir, Tcn1, Tekay, Reina, CRM and Galadriel, each one distinct from the remaining Ty3/*gypsy* retrotransposons that lacked chromodomains (e.g. Ogre, Tat and

Athila elements) (Hřibová et al. 2010; Neumann et al. 2019). Comprehensive analysis of the chromovirus elements has shown that genuinely centromeric-targeting retrotransposons actually represent only a fraction of the CRM lineage and are characterized by a distinctive type of chromodomain (CHDCR, Neumann et al. 2011, 2019). In contrast, Ty3/*gypsy*-Tekay elements have been shown to contain the more widespread CRD-type chromodomains previously classified as Type II by Neumann et al. (2011) and Gao et al. (2008). The presence of the chromodomain likely plays a role in targeting the insertion of Ty3/*gypsy* retrotransposons predominantly into heterochromatin since it can recognize and bind to methylated histone H3K9 (Jacobs and Khorasanizadeh 2002; Nielsen et al. 2002) and H3K27 (Fischle et al. 2003; Min et al. 2003) residues which are common in heterochromatic regions of the DNA (Lin et al. 2008; Gao et al. 2008). Certainly, our data show that Ty3/*gypsy*-Tekay retrotransposon are enriched in the pericentromeric region associated with a $CMA^{+/}DAPI^{-}$ heterochromatin bands, as observed in other legumes such as Caesalpinia group species (Van-Lume et al. 2019).

Previous studies have shown that heterochromatic bands can be composed of different types of repeats, although typically satellite DNA repeats are the most common (Barros and Silva et al. 2010; Samoluk et al. 2017; Zakrzewski and Schmidt 2017). Indeed, studies have shown centromeric and pericentromeric heterochromatin composed by combination of satellite DNA and retrotransposon, as reported here (Cheng et al. 2002; Mizuno et al. 2018; Van-Lume et al. 2019; Báez et al. 2019; Ribeiro et al. 2019). The results reported herein, reinforce the association between retrotransposons and heterochromatic CMA^{+} bands visible on *Pterodon* chromosomes.

Cytogenomic mapping of RTs Ty1/copia-Ale allows differentiation of* P. pubescens *and* P. emarginatus *genomes

The only repetitive element that presented clearly divergent chromosomal distribution between *P. pubescens* and *P. emarginatus* was Ty1/*copia*-Ale. Given the conserved distribution of all repeats analyzed in the proximal CMA^+ heterochromatin, our data suggest that there was an expansion of the Ty1/*copia*-Ale elements, which became dispersed in some chromosomal pairs of *P. pubescens*.

The remarkable stability at integrase sequence level in our comparative analysis of natural populations suggests that this expansion of the Ty1/*copia*-Ale elements in the *P. pubescens* genome was a recent event, considering that the genus *Pterodon* originated 4.21 Mya (Särkinen et al. 2012) (see Online Resource 5). Alternatively, the sequencing strategy used here (direct sequencing of PCR products) may be omitting a putative variability in these elements (see alignment of the sequences of retrotransposon Ty1/*copia*-Ale on Fig. 4). TEs expansion, detectable at chromosome level, has also been reported for *copia* and *gypsy* LTR-RT in species of the genus *Eleocharis* R. Br. (Cyperaceae) (de Souza et al. 2018). Herein, retrotransposons proliferated throughout the chromatin of only a few chromosomal pairs in *P. pubescens*. Probably, this difference between species is caused by such factors as the potential independent activity of each RT family in each genome, epigenetic controls in different genomes or chromosomal regions and the influence of neighboring sequences of other natures (de Souza et al. 2018). Our FISH results reinforce the idea that different families of LTR-RTs have independent activity on genomes and chromosomes, with different evolutionary histories and fates, being a genomic mark that differentiates karyotypes from *P. pubescens* and *P. emarginatus*.

Conclusions

The cytomolecular similarity observed here suggests that the genomes of *P. pubescens* and *P. emarginatus* have highly similar distribution of abundant repeats, which corroborates their phylogenetic proximity. *In situ* hybridization revealed that all repeats (RT and satellite DNA) were present in the proximal heterochromatin of the two species. The exception was the Ty1/*copia*-Ale retroelement, which was scattered across some chromosomal pairs of *P. pubescens*. The Ty1/*copia*-Ale results reinforce the idea that different families of LTR-RTs have independent activity on genomes and chromosomes, with different evolutionary histories and fates, being a genomic mark that differentiates karyotypes from *P. pubescens* and *P. emarginatus* and corroborate the delimitation of both species.

Acknowledgements

We are grateful to Coordenação de Aperfeiçoamento de Pessoal de Nível Superior for financial support and scholarship awarded to VB Albernaz. This paper is developed in the context of National Institutes for Science and Technology (INCT) in Ecology, Evolution and Biodiversity Conservation, supported by MCTIC/CNpq and FAPEG (process# 465610/2014-0, FAPEG (process #201810267000023), Programa de Apoio a Núcleos Emergentes PRONEM/FAPEG/CNPq (process #201710267000539/2016-07), CNPq Universal (process #431320/2018-07). TNS is funded by the research productivity grant from CNPq. We are grateful for Dr. Catarina de Carvalho for kindly provide some informations about the genus *Pterodon*. We are also grateful to Dr Dulce Maria Sucena da Rocha, Dr Mariana Pires de Campos Telles and Dr Lázaro José Chaves for kindly collecting the fruits.

References

Almeida C, Pedrosa-Harand A (2013) High macro-collinearity between lima bean (*Phaseolus lunatus* L.) and the common bean (*P. vulgaris* L.) as revealed by comparative cytogenetic mapping. Theor Appl Genet 126:1909–1916

Báez M, Vaio M, Dreissig S, et al (2019) Together But Different: The Subgenomes of the Bimodal Eleutherine Karyotypes Are Differentially Organized. Front Plant Sci 10:1–14. doi: 10.3389/fpls.2019.01170

Bandel G (1974) Chromosome numbers and evolution in the leguminosae. Caryologia 27:17–32. doi: 10.1080/00087114.1974.10796558

Barnes SR, James AM, Jamieson G (1985) The organization, nucleotide sequence, and chromosomal distribution of a satellite DNA from *Allium cepa*. Chromosoma 92:185–192. doi: 10.1007/BF00348692

Barros e Silva AE, Marques A, Dos Santos KGB, Guerra M (2010) The evolution of CMA bands in *Citrus* and related genera. Chromosom Res 18:503–514. Chromosom Res 18:503–514. doi: 10.1007/s10577-010-9130-2

Belyayev A, Paštová L, Fehrer J, et al (2018) Mapping of *Hieracium* (Asteraceae) chromosomes with genus-specific satDNA elements derived from next-generation sequencing data. Plant Syst Evol 304:387–396. doi: 10.1007/s00606-017-1483-y

Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res 27:573–580. doi:10.1093/nar/27.2.573

Bolsheva NL, Melnikova N V., Kirov I V., et al (2019) Characterization of repeated DNA sequences in genomes of blue-flowered flax. BMC Evol Biol 19:49

Cardoso D, Pennington RT, de Queiroz LP, et al (2013) Reconstructing the deep-branching relationships of the papilionoid legumes. South African J Bot 89:58–75. doi: 10.1016/j.sajb.2013.05.001

Carvalho C *Pterodon* in Flora do Brasil 2020 em construção. In: REFLORA. http://reflora.jbrj.gov.br/reflora/floradobrasil/FB29842. Accessed 28 Jul 2019

Carvalho CR, Saraiva LS (1993) An Air Drying Technique for Maize Chromosomes without Enzymatic Maceration AU - De Carvalho, Carlos Roberto. Biotech Histochem 68:142–145. doi: 10.3109/10520299309104684

Cheng Z, Dong F, Langdon T, et al (2002) Functional Rice Centromeres Are Marked by a Satellite Repeat and a Centromere-Specific Retrotransposon. Plant Cell 14:1691–1704. doi: 10.1105/tpc.003079

Coleman JR, DeMenezes EM (1980) Chromosome numbers in Leguminosae from the State of Sao Paulo. Rhodora 82:475–481

da Costa Silva S, Mendes S, Soares Filho W dos S, Pedrosa-Harand A (2015) Chromosome homologies between *Citrus* and *Poncirus*—the comparative cytogenetic map of mandarin (*Citrus reticulata*). Tree Genet Genomes 11:811. doi: 10.1007/s11295-014-0811-4

Derbyshire MK, Gonzales NR, Lu S, et al (2015) Improving the consistency of domain annotation within the Conserved Domain Database. Database 2015:1–8. doi: 10.1093/database/bav012

de Carvalho R, Guerra M (2002) Cytogenetics of *Manihot esculenta* Crantz (cassava) and eight related species. Hereditas 136:159–168. doi: 10.1034/j.1601-5223.2002.1360212.x

de Souza Almeida CCS, De Lemos Carvalho PC, Guerra M (2007) Karyotype differentiation among *Spondias* species and the putative hybrid Umbu-cajá

(Anacardiaceae). Bot J Linn Soc 155:541–547. doi: 10.1111/j.1095-8339.2007.00721.x

de Souza TB, Chaluvadi SR, Johnen L, et al (2018) Analysis of retrotransposon abundance, diversity and distribution in holocentric *Eleocharis* (Cyperaceae) genomes. Ann Bot 122:279–290. doi: 10.1093/aob/mcy066

Doležel J, Bartoš J (2005) Plant DNA flow cytometry and estimation of nuclear genome size. Ann Bot 95:99–110. doi: 10.1093/aob/mci005

Doležel J, Greilhuber J, Lucretti S, et al (1998) Plant Genome Size Estimation by Flow Cytometry: Inter-laboratory Comparison. Ann Bot 82:17–26. doi: 10.1093/oxfordjournals.aob.a010312

Doležel J, Greilhuber J, Suda J (2007) Estimation of nuclear DNA content in plants using flow cytometry. Nat Protoc 2:2233–2244. doi: 10.1038/nprot.2007.310

Doyle J, Doyle J (1987) Isolation of plant DNA from fresh tissue. Focus 12:13–15

Drummond CS (2008) Diversification of *Lupinus* (Leguminosae) in the western New World: Derived evolution of perennial life history and colonization of montane habitats. Mol Phylogenet Evol 48:408–421. doi: 10.1016/j.ympev.2008.03.009

Faria AM, Santos A da S, Ferreira LL, et al (2017) Acaricidal activity of ethanolic extracts of *Stryphnodendron adstringens* and *Lafoensia pacari* and oil resins from *Copaifera sp.* and *Pterodon emarginatus* against *Sarcoptes scabiei* var. suis. Pesqui Vet Bras 37:1411–1415. doi: 10.1590/S0100-736X2017001200009

Finnegan DJ (1989) Eukaryotic transposable elements and genome evolution. Trends Genet 5:103–107. doi: 10.1016/0168-9525(89)90039-5

Fischle W, Wang Y, Jacobs SA, et al (2003) Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by polycomb and HP1 chromodomains. Genes Dev 17:1870–1881. doi: 10.1101/gad.1110503

Fonsêca A, Pedrosa-Harand A (2017) Cytogenetics and Comparative Analysis of *Phaseolus* Species. In: Pérez de la Vega M., Santalla M. MF (eds) (ed) The Common Bean Genome. Springer, Cham, p pp 57-68

Franco AL, Figueredo A, Pereira L de M, et al (2019) [PROVISIONAL] Low cytomolecular diversification in the genus *Stylosanthes* Sw. (Papilionoideae, Leguminosae). Genet Mol Biol. doi: 10.1590/1678-4685-gmb-2018-0250

Fu J, Zhang H, Guo F, et al (2019) Identification and characterization of abundant repetitive sequences in *Allium cepa*. Sci Rep 9:16756. doi: 10.1038/s41598-019-52995-9

Gao X, Hou Y, Ebina H, et al (2008) Chromodomains direct integration of retrotransposons to heterochromatin. Genome Res 18:359–369. doi: 10.1101/gr.7146408

Garrido-Ramos MA (2015) Satellite DNA in plants: More than just rubbish. Cytogenet Genome Res 146:153–170. doi: 10.1159/000437008

Gerlach WL, Bedbrook JR (1979) Cloning and characterization of ribosomal RNA genes from wheat and barley. Nucleic Acids Res 7:1869–1885. doi: 10.1093/nar/7.7.1869

González ML, Chiapella JO, Urdampilleta JD (2018) Characterization of some satellite DNA families in *Deschampsia antarctica* (Poaceae). Polar Biol 41:457–468. doi: 10.1007/s00300-017-2205-1

Gu Y, Xing S, He C (2016) Genome-wide analysis indicates lineage-specific gene loss during Papilionoideae evolution. Genome Biol Evol 8:635–648. doi: 10.1093/gbe/evw021

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp 95–98. doi: 10.1039/c7qi00394c

Heslop-Harrison J (Pat), Schmidt T (2012) Plant Nuclear Genome Composition. eLS. doi: 10.1002/9780470015902.a0002014.pub2

Hloušková P, Mandáková T, Pouch M, et al (2019) The large genome size variation in the *Hesperis* clade was shaped by the prevalent proliferation of DNA repeats and rarer genome downsizing. Ann Bot 1–17. doi: 10.1093/aob/mcz036

Horváth V, Merenciano M, González J (2017) Revisiting the Relationship between Transposable Elements and the Eukaryotic Stress Response. Trends Genet 33:832–841. doi: 10.1016/j.tig.2017.08.007

Hřibová E, Neumann P, Matsumoto T, et al (2010) Repetitive part of the banana (*Musa acuminata*) genome investigated by low-depth 454 sequencing. BMC Plant Biol 10. doi: 10.1186/1471-2229-10-204

Hufnagel B, Marques A, Soriano A, et al (2019) Genome sequence of the cluster root forming white lupin. bioRxiv 1:708917. doi: 10.1101/708917

Jacobs SA, Khorasanizadeh S (2002) Structure of HP1 Chromodomain Bound to a Lysine 9-Methylated Histone H3 Tail. Science (80) 295:2080–2083. doi: 10.1126/science.1069473

Kearse M, Moir R, Wilson A, et al (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649. doi: 10.1093/bioinformatics/bts199

Kirov I V., Kiseleva A V., Van Laere K, et al (2017) Tandem repeats of Allium fistulosum associated with major chromosomal landmarks. Mol Genet Genomics 292:453–464. doi: 10.1007/s00438-016-1286-9

Leite VG, Mansano VF, Teixeira SP (2014a) Floral ontogeny in Dipterygeae (Fabaceae) reveals new insights into one of the earliest branching tribes in papilionoid legumes. Bot J Linn Soc 174:529–550. doi: 10.1111/boj.12158

Lima S (2019) Filogeografia de *Pterodon emarginatus* e *Pterodon pubescens* (LEGUMINOSAE). Thesis, Universidade Federal de Goiás

Li SF, Guo YJ, Li JR, et al (2019) The landscape of transposable elements and satellite DNAs in the genome of a dioecious plant spinach (*Spinacia oleracea* L.). Mob DNA 10:1–15. doi: 10.1186/s13100-019-0147-6

Lin CH, Li B, Swanson S, et al (2008) Heterochromatin Protein 1a Stimulates Histone H3 Lysine 36 Demethylation by the *Drosophila* KDM4A Demethylase. Mol Cell 32:696–706. doi: 10.1016/j.molcel.2008.11.008

Liu Q, Li X, Zhou X, et al (2019) The repetitive DNA landscape in *Avena* (Poaceae): Chromosome and genome evolution defined by major repeat classes in whole-genome sequence reads. BMC Plant Biol 19:1–17. doi: 10.1186/s12870-019-1769-z

Macas J, Novak P, Pellicer J, et al (2015) In depth characterization of repetitive DNA in 23 plant genomes reveals sources of genome size variation in the legume tribe Fabeae. PLoS One 10:1–23. doi: 10.1371/journal.pone.0143424

Madrigal JJP (2018) Next-Generation Sequencing Technologies in Tree Improvement and Conservation Genetics of *Dipteryx oleifera* Benth. PhD Thesis, North Carolina State University

Marie D, Brown SC (1993) A cytometric exercise in plant DNA histograms, with 2C values for 70 species. Biol Cell 78:41–51. doi: 10.1016/0248-4900(93)90113-S

Marques A, Moraes L, Dos Santos MA, et al (2018) Origin and parental genome characterization of the allotetraploid *Stylosanthes scabra* Vogel (Papilionoideae, Leguminosae), an important legume pasture crop. Ann Bot 122:1143–1159. doi: 10.1093/aob/mcy113

Mehrotra S, Goyal V (2014) Repetitive Sequences in Plant Nuclear DNA: Types, Distribution, Evolution and Function. Genomics Proteomics Bioinformatics 12:164–171. doi: 10.1016/j.gpb.2014.07.003

Min J, Zhang Y, Xu RM (2003) Structural basis for specific binding of polycomb chromodomain to histone H3 methylated at Lys 27. Genes Dev 17:1823–1828. doi: 10.1101/gad.269603

Mizuno H, Matsumoto T, Wu J (2018) Composition and structure of rice centromeres and telomeres. In: Sasaki T., Ashikari M. (eds) Rice Genomics, Genetics and Breeding. Springer, Singapore, pp 37–52

Naganowska B, Wolko B, Śliwińska E, et al (2005) 2C DNA variation and relationships among New World species of the genus *Lupinus* (Fabaceae). Plant Syst Evol 256:147–157. doi: 10.1007/s00606-005-0364-y

Neumann P, Navrátilová A, Koblížková A, et al (2011) Plant centromeric retrotransposons: A structural and cytogenetic perspective. Mob DNA 2:1–16. doi: 10.1186/1759-8753-2-4

Neumann P, Novák P, Hoštáková N, Macas J (2019) Systematic survey of plant LTR-retrotransposons elucidates phylogenetic relationships of their polyprotein domains and provides a reference for element classification. Mob DNA 10:1–17. doi: 10.1186/s13100-018-0144-1

Nielsen PR, Nietlispach D, Mott HR, et al (2002) Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. Nature 416:103–107. doi: 10.1038/nature722

Noé L, Kucherov G (2005) YASS: Enhancing the sensitivity of DNA similarity search. Nucleic Acids Res 33:540–543. doi: 10.1093/nar/gki478

Novák P, Neumann P, Pech J, et al (2013) RepeatExplorer: A Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. Bioinformatics 29:792–793. doi: 10.1093/bioinformatics/btt054

Novák P, Robledillo LÁ, Koblížková A, et al (2017) TAREAN: A computational tool for identification and characterization of satellite DNA from unassembled short reads. Nucleic Acids Res 45:e111-e111. doi: 10.1093/nar/gkx257

Oliver KR, McComb JA, Greene WK (2013) Transposable elements: Powerful contributors to angiosperm evolution and diversity. Genome Biol Evol 5:1886–1901. doi: 10.1093/gbe/evt141

Ortiz AM, Robledo G, Seijo G, et al (2017) Cytogenetic evidences on the evolutionary relationships between the tetraploids of the section Rhizomatosae and related diploid species (*Arachis*, Leguminosae). J Plant Res 130:791–807. doi: 10.1007/s10265-017-0949-x

Outuki PM, Kleinubing SA, Hoscheid J, et al (2018) The incorporation of *Pterodon pubescens* fruit oil into optimized nanostructured lipid carriers improves its effectiveness in colorectal cancer. Ind Crops Prod 123:719–730. doi: 10.1016/j.indcrop.2018.07.044

Pamponét VCC, Cayres P, Souza MM, et al (2019) Low coverage sequencing for repetitive DNA analysis in *Passiflora edulis* Sims: citogenomic characterization of transposable elements and satellite DNA 20:262. doi: 10.1186/s12864-019-5576-6

Pedrosa A, Sandal N, Stougaard J, et al (2002) Chromosomal map of the model legume *Lotus japonicus*. Genetics 161:1661–1672.

Peška V, Mandáková T, Ihradská V, Fajkus J (2019) Comparative dissection of three giant genomes: Allium cepa, Allium sativum, and Allium ursinum. Int J Mol Sci 20:733. doi: 10.3390/ijms20030733

Pinto RB, de Freitas Mansano V, Torke BM, Forni-Martins ER (2016) Evidence for a conserved karyotype in *Swartzia* (Fabaceae, Papilionoideae): Implications for the taxonomy and evolutionary diversification of a species-rich neotropical tree genus. Brittonia 68:93–101. doi: 10.1007/s12228-015-9395-z

Ribeiro T, Marques A, Novák P, et al (2017) Centromeric and non-centromeric satellite DNA organisation differs in holocentric *Rhynchospora* species. Chromosoma 126:325–335. doi: 10.1007/s00412-016-0616-3

Ribeiro T, Vasconcelos E, dos Santos KGB, et al (2019) Diversity of repetitive sequences within compact genomes of Phaseolus L. beans and allied genera Cajanus L. and Vigna Savi. Chromosom Res. doi: 10.1007/s10577-019-09618-w

Rocha DMS (2006) Aspectos taxômicos, genéticos e reprodutivos de *Pterodon pubescens* (Benth.) Benth. e *P. emarginatus* Vog. (Leguminosae, Dipteryxeae). Dissertation, Universidade Estadual de Campinas

Said M, Hřibová E, Danilova T V., et al (2018) The *Agropyron cristatum* karyotype, chromosome structure and cross-genome homoeology as revealed by fluorescence *in situ* hybridization with tandem repeats and wheat single-gene probes. Theor Appl Genet 131:2213–2227. doi: 10.1007/s00122-018-3148-9

Samoluk SS, Chalup LMI, Chavarro C, et al (2019) Heterochromatin evolution in *Arachis* investigated through genome-wide analysis of repetitive DNA. Planta 249:1405–1415. doi: 10.1007/s00425-019-03096-4

Samoluk SS, Robledo G, Bertioli D, Seijo JG (2017) Evolutionary dynamics of an at-rich satellite DNA and its contribution to karyotype differentiation in wild diploid *Arachis* species. Mol Genet Genomics 292:283–296. doi: 10.1007/s00438-016-1271-3

Santos FC, Guyot R, do Valle CB, et al (2015) Chromosomal distribution and evolution of abundant retrotransposons in plants: gypsy elements in diploid and polyploid Brachiaria forage grasses. Chromosom Res 23:571–582. doi: 10.1007/s10577-015-9492-6

Särkinen T, Pennington RT, Lavin M, et al (2012) Evolutionary islands in the Andes: Persistence and isolation explain high endemism in Andean dry tropical forests. J Biogeogr 39:884–900. doi: 10.1111/j.1365-2699.2011.02644.x

Schnable PS, Pasternak S, Liang C, et al (2009) The B73 Maize Genome: Complexity, Diversity, and Dynamics. Science 326:1112–1115. doi:10.1126/science.1178534

Siljak-Yakovlev S, Peccenini S, Muratovic E, et al (2003) Chromosomal differentiation and genome size in three European mountain *Lilium* species. Plant Syst Evol 236:165–173. doi: 10.1007/s00606-002-0240-y

Soltis PS, Soltis DE (2009) The Role of Hybridization in Plant Speciation. Annu Rev Plant Biol 60:561–588. doi: 10.1146/annurev.arplant.043008.092039

Souza G, Crosa O, Guerra M (2015) Karyological, morphological, and phylogenetic diversification in *Leucocoryne* Lindl (Allioideae, Amaryllidaceae). Plant Syst Evol 301:2013–2023. doi: 10.1007/s00606-015-1216-z

Susek K, Bielski WK, Hasterok R, et al (2016) A first glimpse of wild lupin karyotype variation as revealed by comparative cytogenetic mapping. Front Plant Sci 7:1–15. doi: 10.3389/fpls.2016.01152

Tetreault HM, Ungerer MC (2016) Long Terminal Repeat Retrotransposon Content in Eight Diploid Sunflower Species Inferred from Next-Generation Sequence Data. G3; Genes|Genomes|Genetics 6:2299–2308. doi: 10.1534/g3.116.029082

Torke B (2006) Systematics and evolutionary diversification in the species-rich neotropical tree genus *Swartzia* (Leguminosae - Papilionoideae). Thesis (PhD), Washington University

Untergasser A, Cutcutache I, Koressaar T, et al (2012) Primer3-new capabilities and interfaces. Nucleic Acids Res 40:1–12. doi: 10.1093/nar/gks596

Van-Lume B, Esposito T, Diniz-Filho JAF, et al (2017) Heterochromatic and cytomolecular diversification in the Caesalpinia group (Leguminosae): Relationships between phylogenetic and cytogeographical data. Perspect Plant Ecol Evol Syst 29:51–63. doi: 10.1016/j.ppees.2017.11.004

Van-Lume B, Mata Y, Mariana S, et al (2019) Evolutionary convergence or homology ? Comparative cytogenomics of Caesalpinia group species (Leguminosae) reveals diversification in the pericentromeric heterochromatic composition. Planta 250:2173–2186. doi: 10.1007/s00425-019-03287-z

Viana AJC, Souza MM (2012) Comparative cytogenetics between the species *Passiflora edulis* and *Passiflora cacaoensis*. Plant Biol 14:820–827. doi: 10.1111/j.1438-8677.2011.00557.x

Weiss-schneeweiss H, Leitch AR, Mccann J, et al (2015) Employing next generation sequencing to explore the repeat landscape of the plant genome. Next-Generation Seq Plant Syst 1–26. doi: 10.14630/000006

Wessler SR (2006) Teaching Old Genomes New Tricks. In: Implicit Genome. pp 138–165

Won SY, Hwang YJ, Jung JA, et al (2018) Identification of repetitive DNA sequences in the *Chrysanthemum boreale* genome. Sci Hortic (Amsterdam) 236:238–243. doi: 10.1016/j.scienta.2018.03.055

Young ND, Bharti AK (2012) Genome-Enabled Insights into Legume Biology. Annu Rev Plant Biol 63:283–305. doi: 10.1146/annurev-arplant-042110-103754

Zakrzewski F, Schmidt T (2017) Epigenetic Characterization of Satellite DNA in Sugar Beet (*Beta vulgaris*). In: Rajewsky N., Jurga S. BJ (eds) (ed) Plant Epigenetics. Springer, Cham, p pp 445-462

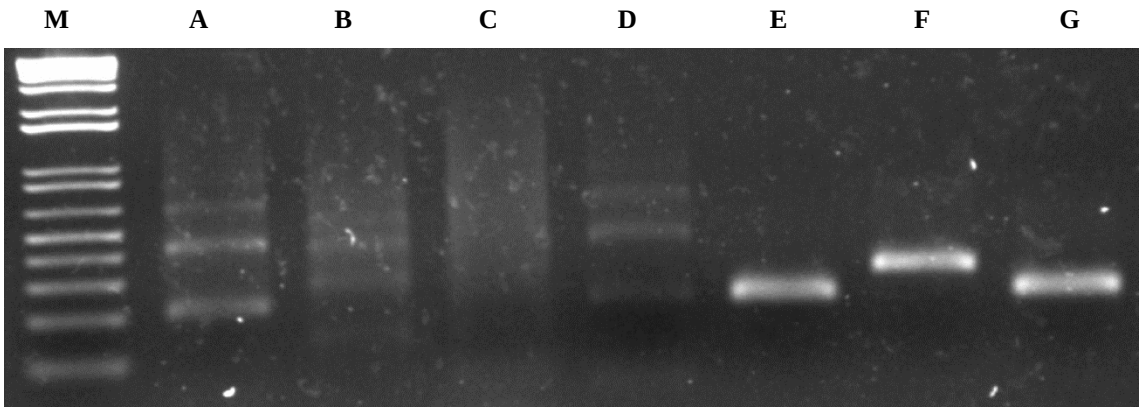
Zhang W, Zuo S, Li Z, et al (2017) Isolation and characterization of centromeric repetitive DNA sequences in Saccharum spontaneum. Sci Rep 7:1–12. doi: 10.1038/srep41659

Zheng JY, Nakata M, Irifune K, et al (1993) Fluorescent banding pattern analysis of eight taxa of *Phaseolus* and *Vigna* in relation to their phylogenetic relationships. Theor Appl Genet 87:38–43. doi: 10.1007/BF00223741

Zoldos V, Papes D, Cerbah M, et al (1999) Molecular-cytogenetic studies of ribosomal genes and heterochromatin reveal conserved genome organization among 11 *Quercus* species. Theor Appl Genet 99:969–977. doi: 10.1007/s001220051404

Zuccolo A, Sebastian A, Talag J, et al (2007) Transposable element distribution, abundance and role in genome size variation in the genus *Oryza*. BMC Evol Biol 7:1–15. doi: 10.1186/1471-2148-7-152

Supplementary material

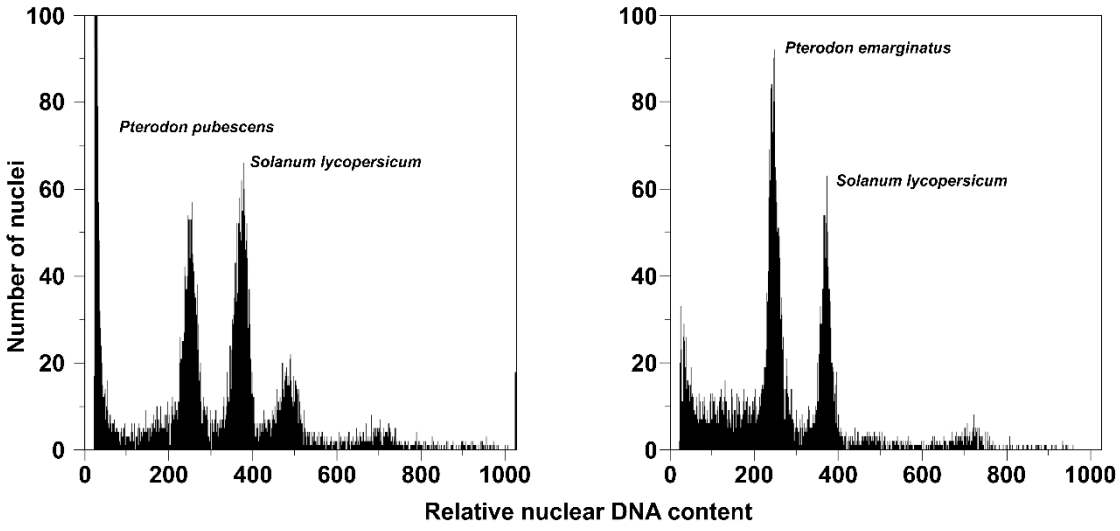


Online Resource 1 1% agarose gel showing the amplification of repetitive regions in *P. pubescens* (A) PubSat1-254, (B) PubSat4-138, (C) PubSat2-76, (K) PubSat3-216, (K) Ty1-Ale, (D) Ty3-Athila and (E) Ty3-Tekay

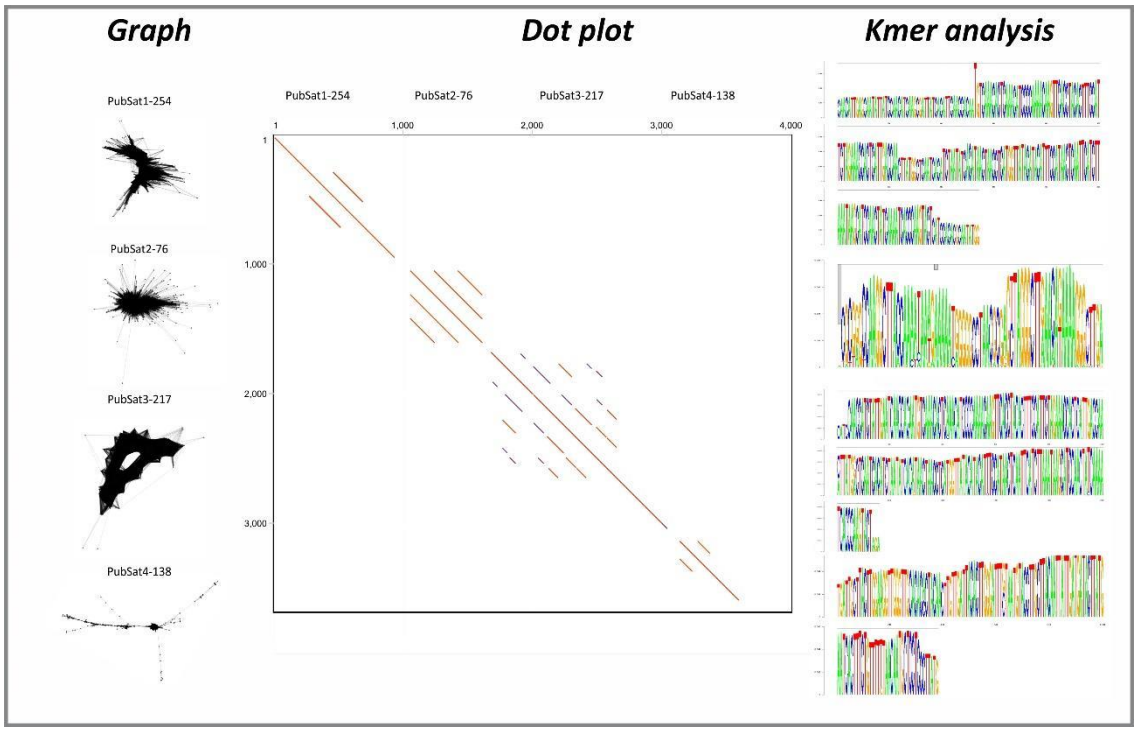
Online Resource 2 Information about the collected populations of *Pterodon pubescens* and *P. emarginatus* for Sanger sequences of integrase region of retrotransposon Ty1/*copia*-Ale, with geographical coordinates

Nº	Longitude	Latitude	Locality	Population Code	Species
1	-46,609800	-14,586420	Alvorada do Norte-GO	PeANOGO	*P. emarginatus*
2	-47,874230	-14,206910	Alto Paraíso-GO	PeAPAGO	*P. emarginatus*
3	-50,562940	-15,087270	Araguapaz-GO	PeAPZGO	*P. emarginatus*
4	-49,589200	-12,754200	Araguaçu-TO	PeARATO	*P. emarginatus*
5	-49,151080	-11,859240	Cariri-TO	PeCARTO	*P. emarginatus*
6	-45,855530	-12,339770	Luiz Eduardo-BA	PeLEMBA	*P. emarginatus*
7	-46,762120	-13,137970	Monte Alegre de Goiás-GO	PeMAGGO	*P. emarginatus*
8	-50,318970	-14,072610	Nova Crixás-GO	PeNCRGO	*P. emarginatus*
9	-48,911850	-10,634870	Nova Rosalândia-TO	PeNROTO	*P. emarginatus*
10	-49,195760	-13,514030	Porangatu-GO	PePORGO	*P. emarginatus*
11	-49,051850	-14,813040	Santa Rita do Novo Destino-GO	PeSRNGO	*P. emarginatus*
12	-50,721170	-15,040440	Araguapaz-GO	PpAPZGO	*P. pubescens*
13	-51,834090	-15,451030	Araguaiana-MT	PpARGMT	*P. pubescens*

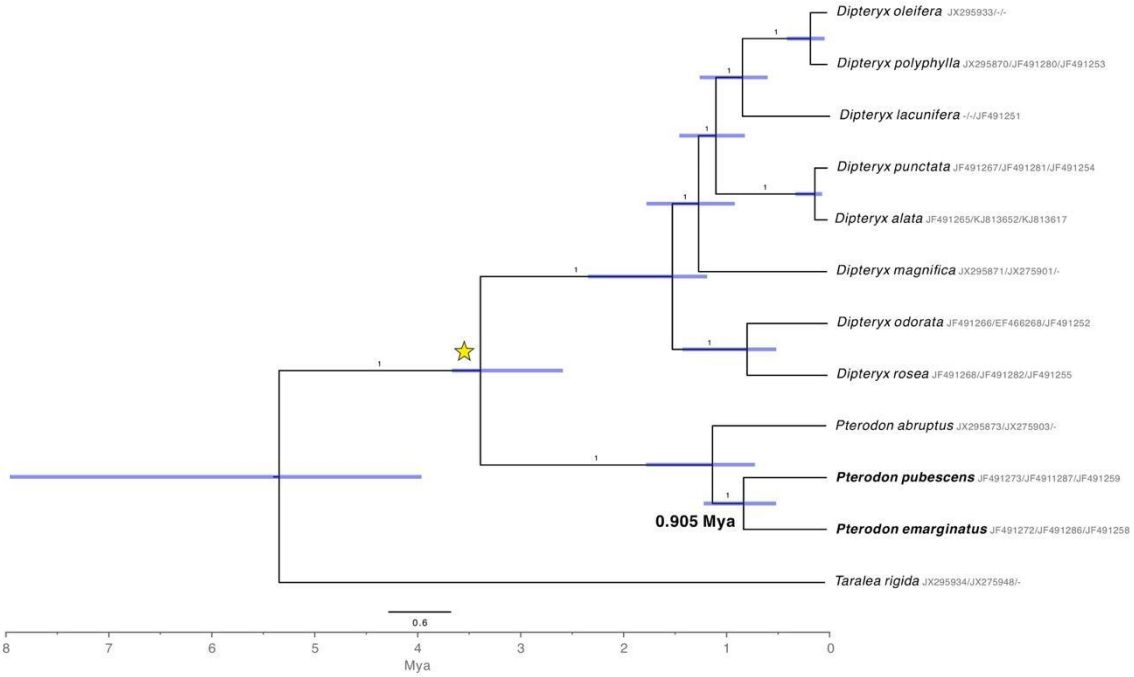
14	-51,078160	-14,940360	Aruanã-GO	PpARUGO	*P. pubescens*
15	-52,553110	-15,789990	Barra do Graças-MT	PpBAGMT	*P. pubescens*
16	-48.625278	-17.745278	Caldas Novas	PpCALGO	*P. pubescens*
17	-51,360290	-14,954970	Cocalinho-MT	PpCOCMT	*P. pubescens*
18	-49,183850	-16,94287	Hidrolândia-GO	PpHIDGO	*P. pubescens*
19	-51,157240	-16,416140	Ipora-GO	PpIPOGO	*P. pubescens*
20	-52,405612	-14,687747	Nova Xavantina-MT	PpNOXMT	*P. pubescens*
21	-47,366580	-19,350770	Perdizes-MG	PpPERMG	*P. pubescens*
22	-48,955280	-15,748170	Pirenópolis-GO	PpPIRGO	*P. pubescens*
23	-49,444200	-17,520200	Pontalina-GO	PpPONGO	*P. pubescens*
24	-52,607767	-17,350414	Portelândia-GO	PpPOTGO	*P. pubescens*
25	-47,045680	-17,069150	Paracatu-MG	PpPTUMG	*P. pubescens*
26	-48,660000	-17,772500	Serra de Caldas-GO	PpSECGO	*P. pubescens*
27	-49,251530	-15,920950	São Francisco de Goiás-GO	PpSFRGO	*P. pubescens*
28	-48,378230	-16,597670	Silvania-GO	PpSILGO	*P. pubescens*
29	-48,900820	-18,537490	Tupaciagura-MG	PpTUPMG	*P. pubescens*
30	-59,460050	-15,167570	Pontes e Lacerda-MT	PePOLMT	*P. pubescens*



Online Resource 3 Histograms of relative fluorescence intensities obtained by analyzing isolated nuclei from *Pterodon pubescens* — 1C = 0.665 pg DNA (left histogram) and, *Pterodon emarginatus* — 1C = 0.620 pg DNA (right histogram) and the internal reference standard *(Solanum lycopersicum* — 1C = 0.980 pg DNA)



Online Resource 4 Characterization of the putative satellite DNA repeats identified in *Pterodon pubescens* genome. (a) Graph layouts derived from RepeatExplorer output, (b) Dot-plot similarity comparison of all repeat sequences against themselves, showing that satellite DNA sequences from different clusters do not share similarity and (c) Monomer reconstruction using base frequency logo representation (WebLogo 3; http://weblogo.threeplusone.com)



Online Resource 5 Chronogram of Dipterygeae, with focus on the age of *Pterodon* species, based on BEAST analysis using nuclear and plastid consensus topology. Blue bars indicate 95 % highest posterior density intervals. GenBank accession numbers are given as grey letters on the side of each sample name in order: *mat*K/ *trn*L-*trn*F/ ITS

Considerações finais

- O tamanho do genoma de *P. pubescens* e *P. emarginatus* foram semelhantes.

- A fração repetitiva representou 26,4% do genoma obtido de *P. pubescens*, sendo as famílias Ty3-*gypsy* e Ty1-*copia* mais abundantes.

- As espécies analisadas apresentaram bandas heterocromáticas muito similares, concentrando-se principalmente na região pericentromérica de todos os cromossomos.

- As sondas de três retrotransposons do tipo LTR e de quatro DNA satélites foram desenvolvidas e transferidas para *P. emarginatus*. Porém apenas duas sondas de DNA satélites apresentaram sinal no experimento de FISH (PubSat2-76 e PubSat4-138).

- A hibridização *in situ* revelou que todas as sequências repetitivas analisadas estão colocalizadas com as bandas CMA^{+} pericentroméricas em ambas espécies, exceto o retroelemento Ty1-Ale, que também está disperso na eucromatina de quatro pares de cromossomos (I, II, III e VIII) de *P. pubescens*.

- A similaridade citomolecular observada aqui sugere que os genomas de *P. pubescens* e *P. emarginatus* possuem frações repetitivas altamente semelhantes.

Referências

Afonso AP (1997) Biologia reprodutiva e polinização de Pterodon pubescens Benth e Pterodon polygalaeflorus Benth (Fabaceae-Papilionoideae) em áreas de cerrado do Distrito Federal. Dissertation, Universidade de Brasília

Almeida C, Pedrosa-Harand A (2013) High macro-collinearity between lima bean (*Phaseolus lunatus* L.) and the common bean (*P. vulgaris* L.) as revealed by comparative cytogenetic mapping. Theor Appl Genet 126:1909–1916. doi: 10.1007/s00122-013-2106-9

Almeida D, Marchini LC, Sodré GS, et al (2003) Plantas visitadas por abelhas e polinização, Série Prod. Divisão de Biblioteca e Documentação – DIBD, Piracicaba: ESALQ

Ansorge WJ (2009) Next-generation DNA sequencing techniques. N Biotechnol 25:195–203. doi: 10.1016/j.nbt.2008.12.009

Bachmann L, Raab M, Sperlich D (1989) Satellite DNA and speciation: A species specific satellite DNA of Drosophila guanchel. J Zool Syst Evol Res 27:84–93. doi: 10.1111/j.1439-0469.1989.tb00333.x

Bandel G (1974) Chromosome numbers and evolution in the leguminosae. Caryologia 27:17–32. doi: 10.1080/00087114.1974.10796558

Belyayev A, Paštová L, Fehrer J, et al (2018) Mapping of *Hieracium* (Asteraceae) chromosomes with genus-specific satDNA elements derived from next-generation sequencing data. Plant Syst Evol 304:387–396. doi: 10.1007/s00606-017-1483-y

Bennett MD, Leitch IJ (2011) Nuclear DNA amounts in angiosperms: Targets, trends and tomorrow. Ann Bot 107:467–590. doi: 10.1093/aob/mcq258

Bolsheva NL, Melnikova N V., Kirov I V., et al (2019) Characterization of repeated DNA sequences in genomes of blue-flowered flax. BMC Evol Biol 19:49. doi: 10.1186/s12862-019-1375-6

Cardoso D, De Queiroz LP, Toby Pennington R, et al (2012) Revisiting the phylogeny of papilionoid legumes: New insights from comprehensively sampled early-branching lineages. Am J Bot 99:1991–2013. doi: 10.3732/ajb.1200380

Cardoso D, Pennington RT, de Queiroz LP, et al (2013) Reconstructing the deep-branching relationships of the papilionoid legumes. South African J Bot 89:58–75. doi: 10.1016/j.sajb.2013.05.001

Cardoso D, São-Mateus WMB, da Cruz DT, et al (2015) Filling in the gaps of the papilionoid legume phylogeny: The enigmatic Amazonian genus *Petaladenium* is a new branch of the early-diverging Amburaneae clade. Mol Phylogenet Evol 84:112–124.

Carvalho C *Pterodon* in Flora do Brasil 2020 em construção. In: REFLORA. http://reflora.jbrj.gov.br/reflora/floradobrasil/FB29842. Accessed 28 Jul 2019

Coelho LP, Reis PA, De Castro FL, et al (2005) Antinociceptive properties of ethanolic extract and fractions of *Pterodon pubescens* Benth. seeds. J Ethnopharmacol 98:109–116. doi: 10.1016/j.jep.2005.01.014

Coleman JR, DeMenezes EM (1980) Chromosome numbers in Leguminosae from the State of Sao Paulo. Rhodora 82:475–481.

da Costa Silva S, Mendes S, Soares Filho W dos S, Pedrosa-Harand A (2015) Chromosome homologies between *Citrus* and *Poncirus*—the comparative cytogenetic map of mandarin (*Citrus reticulata*). Tree Genet Genomes 11:811. doi: 10.1007/s11295-014-0811-4

de Moraes WF, De Matos LG, Mariano Nascimento MV, et al (2009) Anti-inflammatory and anti-nociceptive effects of *Pterodon emarginatus* stem bark alcohol extract. Pharm Biol 47:146–150. doi: 10.1080/13880200802436117

de Souza Almeida CCS, De Lemos Carvalho PC, Guerra M (2007) Karyotype differentiation among *Spondias* species and the putative hybrid Umbu-cajá (Anacardiaceae). Bot J Linn Soc 155:541–547. doi: 10.1111/j.1095-8339.2007.00721.x

Delgado-Salinas A, Bibler R, Lavin M (2007) Phylogeny of the Genus *Phaseolus* (Leguminosae): A Recent Diversification in an Ancient Landscape. Syst Bot 31:779–791. doi: 10.1600/036364406779695960

DRYFLOR (2016) Plant diversity patterns in neotropical dry forests and their conservation implications.

Finnegan DJ (1989) Eukaryotic transposable elements and genome evolution. Trends Genet 5:103–107. doi: 10.1016/0168-9525(89)90039-5

Fishman L, Saunders A (2008) Centromere-Associated Female Meiotic Drive Entails Male Fitness Costs in Monkeyflowers. Science (80) 322:1559–1562. doi: 10.1126/science.1161406

Gaiero P, Vaio M, Peters SA, et al (2019) Comparative analysis of repetitive sequences among species from the potato and the tomato clades. Ann Bot 123:521–532. doi: 10.1093/aob/mcy186

Galindo-González L, Mhiri C, Deyholos MK, Grandbastien MA (2017) LTR-retrotransposons in plants: Engines of evolution. Gene 626:14–25.

Garrido-Ramos MA (2015) Satellite DNA in plants: More than just rubbish. Cytogenet Genome Res 146:153–170. doi: 10.1159/000437008

González ML, Chiapella JO, Urdampilleta JD (2018) Characterization of some satellite DNA families in *Deschampsia antarctica* (Poaceae). Polar Biol 41:457–468. doi: 10.1007/s00300-017-2205-1

Guerra M (2004) FISH - Conceitos e Aplicações na Citogenética. SBG, Ribeirão Preto, SP

Guerra M (2000) Patterns of heterochromatin distribution in plant chromosomes. Genet Mol Biol 23:1029–1041. doi: 10.1590/S1415-47572000000400049

Hemleben V, Volkov RA, Zentgraf U MF (2004) Molecular Cell Biology: Organization and Molecular Evolution of rDNA, Nucleolar Dominance, and Nucleolus Structure. In: Esser K, Lüttge U, Beyschlag W MJ (ed) Progress in Botany 65. Springer Berlin Heidelberg

Henikoff S (2001) The Centromere Paradox: Stable Inheritance with Rapidly Evolving DNA. Science (80-) 293:1098–1102. doi: 10.1126/science.1062939

Heslop-Harrison JSP, Schwarzacher T (2011) Organisation of the plant genome in chromosomes. Plant J 66:18–33. doi: 10.1111/j.1365-313X.2011.04544.x

Hloušková P, Mandáková T, Pouch M, et al (2019) The large genome size variation in the *Hesperis* clade was shaped by the prevalent proliferation of DNA repeats and rarer genome downsizing. Ann Bot 1–17. doi: 10.1093/aob/mcz036

Hoscheid J, Bersani-Amado CA, Da Rocha BA, et al (2013) Inhibitory effect of the hexane fraction of the ethanolic extract of the fruits of *Pterodon pubescens* benth in acute and chronic inflammation. Evidence-based Complement Altern Med 2013.

Hřibová E, Neumann P, Matsumoto T, et al (2010) Repetitive part of the banana (*Musa acuminata*) genome investigated by low-depth 454 sequencing. BMC Plant Biol 10:204. doi: 10.1186/1471-2229-10-204

Iganci JRV, Soares M V., Guerra E, Morim MP (2016) A preliminary molecular phylogeny of the *Abarema* Alliance (Leg uminosae) and implications for taxonomic rearrangement. Int J Plant Sci 177:34–43. doi: 10.1086/684078

Leite VG, Mansano VF, Teixeira SP (2014a) Floral ontogeny in Dipterygeae (Fabaceae) reveals new insights into one of the earliest branching tribes in papilionoid legumes. Bot J Linn Soc 174:529–550. doi: 10.1111/boj.12158

Leite VG, Teixeira SP, Mansano VF, Prenner G (2014b) Floral development of the early-branching papilionoid legume *Amburana cearensis* (Leguminosae) reveals rare and novel characters. Int J Plant Sci 176:94–106. doi: 10.1086/678468

Lewis, G., Schrire B., Mackinder, B. & Lock M (eds. (2005) Legumes of the World. Kew: Royal Botanic Gardens, Richmond, U.K.

Lewis GP, Schrire BD, Mackinder BA, et al (2013) A 2013 linear sequence of legume genera set in a phylogenetic context - A tool for collections management and taxon sampling. South African J Bot 89:76–84. doi: 10.1016/j.sajb.2013.06.005

Lima S (2019) Filogeografia de *Pterodon emarginatus* e *Pterodon pubescens* (LEGUMINOSAE). Thesis, Universidade Federal de Goiás

Li W, Zhang P, Fellers JP, et al (2004) Sequence composition, organization, and evolution of the core Triticeae genome. Plant J 40:500–511. doi: 10.1111/j.1365-313X.2004.02228.x

Liu Q, Li X, Zhou X, et al (2019) The repetitive DNA landscape in *Avena* (Poaceae): Chromosome and genome evolution defined by major repeat classes in whole-genome sequence reads. BMC Plant Biol 19:1–17. doi: 10.1186/s12870-019-1769-z

LPWG (2017) A new subfamily classification of the leguminosae based on a taxonomically comprehensive phylogeny. Taxon 66:44–77. doi: 10.12705/661.3

Macas J, Novak P, Pellicer J, et al (2015) In depth characterization of repetitive DNA in 23 plant genomes reveals sources of genome size variation in the legume tribe Fabeae. PLoS One 10:1–23. doi: 10.1371/journal.pone.0143424

Malik CP, Thomas PT (1966) Karyotypic Studies in some *Lolium* and *Festuca* species. Caryologia 19(2), 167–196. doi:10.1080/00087114.1966.10796216

Martel E, De Nay D, Siljak-Yakovlev S, et al (1997) Genome size variation and basic chromosome number in pearl millet and fourteen related *Pennisetum* species. J Hered 88:139–143. doi: 10.1093/oxfordjournals.jhered.a023072

McKey D (1994) Legumes and nitrogen: The evolutionary ecology of a nitrogen-demanding lifestyle. Adv Legum Syst 5 Nitrogen Factor 5:211–228. doi: 10.1007/s13398-014-0173-7.2

Mehrotra S, Goyal V (2014) Repetitive Sequences in Plant Nuclear DNA: Types, Distribution, Evolution and Function. Genomics Proteomics Bioinformatics 12:164–171. doi: 10.1016/j.gpb.2014.07.003

Melo ZP (2018) Tamanho genômico, desenvolvimento de marcadores microssatélites para a espécie *Pterodon pubescens* e transferibilidade para *Pterodon emarginatus* (leguminosae). Dissertation, Universidade Federal de Goiás

Melters DP, Bradnam KR, et al (2013) Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. Genome Biol 14(1), R10. doi:10.1186/gb-2013-14-1-r10

Mendonça de Almeida E, Wanderley AM, Nollet F, et al (2016) A New Species of *Ameroglossum* (Scrophulariaceae) Growing on Inselbergs in Northeastern Brazil. Syst Bot 41:423–429. doi:10.1600/036364416x691740

Menna-Barreto RFS, Laranja GAT, Silva MCC, et al (2008) Anti-Trypanosoma cruzi activity of *Pterodon pubescens* seed oil: geranylgeraniol as the major bioactive component. Parasitol Res 103:111–117. doi: 10.1007/s00436-008-0937-0

Neumann P, Novák P, Hoštáková N, MacAs J (2019) Systematic survey of plant LTR-retrotransposons elucidates phylogenetic relationships of their polyprotein domains and provides a reference for element classification. Mob DNA 10:1–17. doi: 10.1186/s13100-018-0144-1

Novák P, Neumann P, Pech J, et al (2013) RepeatExplorer: A Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. Bioinformatics 29:792–793. doi: 10.1093/bioinformatics/btt054

Oliveira AEMFM, Duarte JL, Amado JRR, et al (2016) Development of α larvicidal nanoemulsion with *Pterodon emarginatus* Vogel oil. PLoS One 11:1–16. doi: 10.1371/journal.pone.0145835

Ortiz AM, Robledo G, Seijo G, et al (2017) Cytogenetic evidences on the evolutionary relationships between the tetraploids of the section Rhizomatosae and related diploid species (*Arachis*, Leguminosae). J Plant Res 130:791–807. doi: 10.1007/s10265-017-0949-x

Pamponét VCC, Cayres P, Souza MM, et al (2019) Low coverage sequencing for repetitive DNA analysis in *Passiflora edulis* Sims: citogenomic characterization of transposable elements and satellite DNA. BMC Genomics 20:262. doi:10.1186/s12864-019-5576-6

Peška V, Mandáková T, Ihradská V, Fajkus J (2019) Comparative dissection of three giant genomes: Allium cepa, Allium sativum, and Allium ursinum. Int J Mol Sci 20:733. doi: 10.3390/ijms20030733

Pich U, Fritsch R, Schubert I (1996) Closely related *Allium* species (Alliaceae) share a very similar satellite sequence. Plant Syst Evol 202:255–264. doi: 10.1007/BF00983386

Pinto MVP (2017) Diversidade e estrutura genética em populações naturais de *Pterodon emarginatus* vogel (Leguminosae). Dissertation, Universidade Federal de Goiás

Plohl M, Meštrović N, Mravinac B (2014) Centromere identity from the DNA point of view. Chromosoma 123:313–325. doi: 10.1007/s00412-014-0462-0

Rando JG, Zuntini AR, Conceição AS, et al (2016) Phylogeny of *Chamaecrista* ser. Coriaceae (Leguminosae) unveils a lineage recently diversified in Brazilian campo Rupestre vegetation. Int J Plant Sci 177:3–17. doi: 10.1086/683846

Rice A, Glick L, Abadi S, et al (2015) The Chromosome Counts Database (CCDB) - a community resource of plant chromosome numbers. New Phytol 206:19–26. doi: 10.1111/nph.13191

Rocha DMS (2006) Aspectos taxômicos, genéticos e reprodutivos de *Pterodon pubescens* (Benth.) Benth. e *P. emarginatus* Vog. (Leguminosae, Dipteryxeae). Dissertation, Universidade Estadual de Campinas

Rosin LF, Mellone BG (2017) Centromeres Drive a Hard Bargain. Trends Genet 33:101–117. doi: 10.1016/j.tig.2016.12.001

Said M, Hřibová E, Danilova T V., et al (2018) The *Agropyron cristatum* karyotype, chromosome structure and cross-genome homoeology as revealed by fluorescence *in situ* hybridization with tandem repeats and wheat single-gene probes. Theor Appl Genet 131:2213–2227. doi: 10.1007/s00122-018-3148-9

Samoluk SS, Chalup LMI, Chavarro C, et al (2019) Heterochromatin evolution in *Arachis* investigated through genome-wide analysis of repetitive DNA. Planta 249:1405–1415. doi: 10.1007/s00425-019-03096-4

Scherson RA, Vidal R, Sanderson MJ (2008) Phylogeny, biogeography, and rates of diversification of New World *Astragalus* (Leguminosae) with an emphasis on South American radiations. Am J Bot 95:1030–1039. doi: 10.3732/ajb.0800017

Siljak-Yakovlev S, Peccenini S, Muratovic E, et al (2003) Chromosomal differentiation and genome size in three European mountain *Lilium* species. Plant Syst Evol 236:165–173. doi: 10.1007/s00606-002-0240-y

Siljak-Yakovlev S, Pustahija F, Vicic V, Robin O (2014) Molecular Cytogenetics (FISH and Fluorochrome Banding): Resolving Species Relationships and Genome Organization. In: Besse P. (eds) Molecular Plant Taxonomy. Methods in Molecular Biology (Methods and Protocols). pp 309–323

Snak C, Vatanparast M, Silva C, et al (2016) A dated phylogeny of the papilionoid legume genus *Canavalia* reveals recent diversification by a pantropical liana lineage. Mol Phylogenet Evol 98:133–146. doi: 10.1016/j.ympev.2016.02.001

Stace H (1978) Cytoevolution in the genus *Calotis* R. Br. (Compositae: Astereae). Aust J Bot 26:287–307.

Talukdar D, Sinjushin A (2015) Cytogenomics and Mutagenomics in Plant Functional Biology and Breeding. In: PlantOmics: The Omics of Plant Science. Springer India, New Delhi, pp 113–156

Taquary AMA (2017) Tamanho, montagem de novo e anotação do genoma de *Dipteryx alata* (Leguminosae). Thesis, Universidade Federal de Goiás

Van-Lume B, Esposito T, Diniz-Filho JAF, et al (2017) Heterochromatic and cytomolecular diversification in the Caesalpinia group (Leguminosae):

Relationships between phylogenetic and cytogeographical data. Perspect Plant Ecol Evol Syst 29:51–63. doi: 10.1016/j.ppees.2017.11.004

Van-Lume B, Mata Y, Mariana S, et al (2019) Evolutionary convergence or homology ? Comparative cytogenomics of Caesalpinia group species (Leguminosae) reveals diversification in the pericentromeric heterochromatic composition. Planta 250:2173–2186. doi: 10.1007/s00425-019-03287-z

Viana AJC, Souza MM (2012) Comparative cytogenetics between the species *Passiflora edulis* and *Passiflora cacaoensis*. Plant Biol 14:820–827. doi: 10.1111/j.1438-8677.2011.00557.x

Volkov RA, Panchuk II, Borisjuk N V., et al (2017) Evolutional dynamics of 45S and 5S ribosomal DNA in ancient allohexaploid *Atropa belladonna*. BMC Plant Biol 17:1–15. doi: 10.1186/s12870-017-0978-6

Weiss H, Stuessy TF, Grau J, et al (2003) Chromosome Reports from South American *Hypochaeris* (Asteraceae). Ann Missouri Bot Gard 90:56. doi: 10.2307/3298525

Wessler SR (2006) Teaching Old Genomes New Tricks. In: Implicit Genome. pp 138–165

Won SY, Hwang YJ, Jung JA, et al (2018) Identification of repetitive DNA sequences in the *Chrysanthemum boreale* genome. Sci Hortic (Amsterdam) 236:238–243. doi: 10.1016/j.scienta.2018.03.055

Yahara T, Javadi F, Onoda Y, et al (2013) Global legume diversity assessment: Concepts, key indicators, and strategies. Taxon 62:249–266. doi: 10.12705/622.12

Zuccolo A, Sebastian A, Talag J, et al (2007) Transposable element distribution, abundance and role in genome size variation in the genus *Oryza*. BMC Evol Biol 7:1–15. doi: 10.1186/1471-2148-7-152