



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

**VARIABILIDADE GENÉTICA EM *Dipteryx alata* Vogel  
(Leguminosae): NOVOS MARCADORES MICROSSATÉLITES,  
DISPERSÃO DE PÓLEN *IN SITU* E *EX SITU* E GENÉTICA DA  
PAISAGEM**

**REJANE ARAÚJO GUIMARÃES**

Orientadora:  
**Prof.<sup>a</sup> Dr.<sup>a</sup> Thannya Nascimento Soares**

Fevereiro – 2018

**TERMO DE CIÊNCIA E DE AUTORIZAÇÃO 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 nº 9610/98, 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.

1. Identificação do material bibliográfico:     Dissertação     Tese

2. Identificação da Tese ou Dissertação:

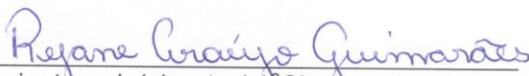
Nome completo do autor: Rejane Araújo Guimarães

Título do trabalho: Variabilidade genética em *Dipteryx alata* Vogel (Leguminosae):  
Novos marcadores microssatélites, Dispersão de pólen *in situ* e *ex situ* e Genética da paisagem

3. Informações de acesso ao documento:

Concorda com a liberação total do documento  SIM     NÃO<sup>1</sup>

Havendo concordância com a disponibilização eletrônica, torna-se imprescindível o envio do(s) arquivo(s) em formato digital PDF da tese ou dissertação.

  
Assinatura do(a) autor(a)<sup>2</sup> Ciente e de acordo:

  
Assinatura do(a) orientador(a)<sup>2</sup>

Data: 23 / 02 / 2018

<sup>1</sup> Neste caso o documento será embargado por até um ano a partir da data de defesa. A extensão deste prazo suscita justificativa junto à coordenação do curso. Os dados do documento não serão disponibilizados 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.

<sup>2</sup> A assinatura deve ser escaneada.

---

**TERMO DE CIÊNCIA E DE AUTORIZAÇÃO 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 nº 9610/98, 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.

- 1. Identificação do material bibliográfico:**     Dissertação     Tese  
**2. Identificação da Tese ou Dissertação:**

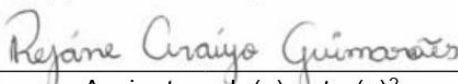
Nome completo do autor: Rejane Araújo Guimarães

Título do trabalho: Variabilidade genética de *Dipteryx alata* vogel (Leguminosae): novos marcadores microssatélites, fluxo de pólen in situ e ex situ e genética da paisagem em escala local

**3. Informações de acesso ao documento:**

Concorda com a liberação total do documento  SIM     NÃO<sup>1</sup>

Havendo concordância com a disponibilização eletrônica, torna-se imprescindível o envio do(s) arquivo(s) em formato digital PDF da tese ou dissertação.

  
Assinatura do(a) autor(a)<sup>2</sup>

Ciente e de acordo:

  
Assinatura do(a) orientador(a)<sup>2</sup>

Data: 30 / 10 / 2020

---

<sup>1</sup> Neste caso o documento será embargado por até um ano a partir da data de defesa. A extensão deste prazo suscita justificativa junto à coordenação do curso. Os dados do documento não serão disponibilizados 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.

<sup>2</sup> A assinatura deve ser escaneada.

**VARIABILIDADE GENÉTICA EM *Dipteryx alata* Vogel  
(Leguminosae): NOVOS MARCADORES MICROSSATÉLITES,  
DISPERSÃO DE PÓLEN *IN SITU* E *EX SITU* E GENÉTICA DA  
PAISAGEM**

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

Orientadora:

**Prof.<sup>a</sup> Dr.<sup>a</sup> Thannya Nascimento Soares**

Coorientadora:

**Prof.<sup>a</sup> Dr.<sup>a</sup> Mariana Pires de Campos Telles**

Goiânia, GO – Brasil

2018

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

Araújo Guimarães, Rejane

Variabilidade genética em *Dipteryx alata* vogel (Leguminosae):  
Novos marcadores microssatélites, Dispersão de pólen in situ e ex situ  
e Genética da paisagem [manuscrito] / Rejane Araújo Guimarães.  
2018.

XCVI, 96 f.

Orientador: Prof. Thannya Nascimento Soares; co-orientador Prof.  
Mariana Pires de Campos Telles.

Tese (Doutorado) - Universidade Federal de Goiás, Escola de  
Agronomia e Engenharia de Alimentos (EAEA), Programa de Pós-  
Graduação em Genética & Melhoramentos de Plantas, Goiânia, 2018.

1. Análise de parentesco. 2. Fluxo gênico. 3. Illumina MiSeq. 4.  
Microssatélites. 5. Spatial Genetic Structure - SGS. I. Nascimento  
Soares, Thannya , orient. II. Título.

CDU 575



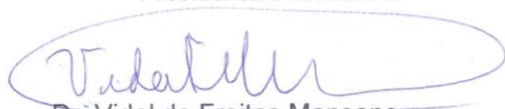
SERVIÇO PÚBLICO FEDERAL  
MINISTÉRIO DA EDUCAÇÃO  
UNIVERSIDADE FEDERAL DE GOIÁS  
ESCOLA DE AGRONOMIA  
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E  
MELHORAMENTO DE PLANTAS



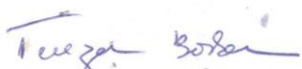
**ATA DE DEFESA PÚBLICA DE TESE DE REJANE ARAÚJO GUIMARÃES.** Aos vinte e três dias do mês de Fevereiro do ano de dois mil e dezoito (23.02.2018), às 13:30h, no Auditório do Setor de Melhoramento de Plantas, da Escola de Agronomia da Universidade Federal de Goiás, reuniram-se os membros da Banca Examinadora: Dra. Thannya Nascimento Soares – Orientadora/Presidente, Dr. Vidal de Freitas Mansano, Dra. Tereza Cristina de Oliveira Borba, Dr. Lázaro José Chaves e Dr. José Alexandre Felizola Diniz Filho. Sob a presidência da orientadora, em sessão pública, procedeu-se à avaliação da defesa da tese intitulada: **“VARIABILIDADE GENÉTICA EM *Dipteryx alata* Vogel (Leguminosae): NOVOS MARCADORES MICROSSATÉLITES, DISPERSÃO DE PÓLEN *IN SITU* E *EX SITU* E GENÉTICA DA PAISAGEM”**, de autoria de **Rejane Araújo Guimarães**, discente do Programa de Pós-Graduação em Genética e Melhoramento de Plantas, no nível de Doutorado, da Universidade Federal de Goiás. A sessão foi aberta pela presidente da Banca Examinadora, Profa. Dra. Thannya Nascimento Soares, que fez a apresentação formal dos membros da Banca. A palavra, a seguir, foi concedida ao autor da Tese que, em 40 minutos, apresentou o seu trabalho. Terminada a apresentação, cada membro da Banca arguiu a doutoranda, tendo-se adotado o sistema de diálogo sequencial. Ao final, a banca reunida em separado procedeu à avaliação da defesa. A tese foi considerada aprovada pela Banca Examinadora, cumprindo integralmente este requisito para fins de obtenção do título de DOUTOR EM GENÉTICA E MELHORAMENTO DE PLANTAS, pela Universidade Federal de Goiás, em conformidade com o estabelecido pela resolução nº 1403/2016, do Conselho de Ensino, Pesquisa, Extensão e Cultura da UFG (CEPEC/UFG), que regulamenta o Programa de Pós-Graduação em Genética e Melhoramento de Plantas. Para fins de publicação eletrônica, a doutoranda poderá efetuar as modificações eventualmente sugeridas pela Banca Examinadora e encaminhar à Secretaria do PGMP, respeitando-se o prazo máximo de 30 dias após a data da Defesa. A conclusão do curso e a emissão do diploma dar-se-ão em conformidade com o estabelecido pela Resolução CEPEC nº 1403/2016. Cumpridas as formalidades de pauta, às 17:40 a presidência da mesa encerrou esta sessão de defesa de Tese e, para constar eu, Jéssica Almeida Silva, Secretária do PGMP, lavrei a presente Ata que depois de lida e aprovada, segue assinada pelos membros da Banca Examinadora, em duas vias de igual teor.

  
Drª. Thannya Nascimento Soares  
Presidente/Orientadora

  
Dr. Lázaro José Chaves  
Membro Interno

  
Dr. Vidal de Freitas Mansano  
Membro Externo

  
Dr. José Alexandre Felizola Diniz Filho  
Membro Interno

  
Drª. Tereza Cristina de Oliveira Borba  
Membro Externo

*“A persistência é o melhor caminho para o êxito”*

**Charles Chaplin**

Aos meus Pais, Venerando e Isabel e a minha  
querida Avó, Derciolina, pelo amor incondicional.

*Ofereço*

Ao meu esposo, Giuliano Carvalho Frugeri, pelo  
amor, companheirismo e incentivo sem medida.

*Dedico*

## Agradecimentos

À todas as agências que viabilizaram o desenvolvimento deste trabalho, pelo apoio financeiro. Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico, através do GENPAC (Genética Geográfica e Planejamento Regional de Recursos Naturais no Cerrado Brasileiro) do CNPq / FAPEG (projetos nº 563839 / 2010-4 e # 201110267000125), pelo CERGEN (Núcleo de Excelência em Recursos Genéticos Vegetais do Cerrado - PRONEX / FAPEG / CNPq 07-2012), e pela Chamada pública 08/2014 Capes/FAPEG. Este estudo foi desenvolvido no contexto do Instituto Nacional de Ciência e Tecnologia (INCT) em Ecologia, Evolução e Conservação da Biodiversidade, apoiado pelo MCTIC/CNPq (proc. 465610 / 2014-5) e FAPEG.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e à Fundação de Amparo a pesquisa do estado de Goiás (FAPEG) pela concessão da bolsa de estudo concedida durante o período de realização do doutorado.

Ao Programa de Pós-Graduação *Stricto Sensu* em Genética e Melhoramento de Plantas, da UFG, pelo suporte na realização deste trabalho; À Jéssica Almeida Silva, secretária do PGMP, pela prontidão e suporte administrativo.

À minha orientadora Dra. Thannya Nascimento Soares pela disponibilidade, dedicação, atenção, paciência, apoio e companheirismo. Pela oportunidade de realizar este trabalho como parceiras, e não apenas como aluno/professor; pelos ensinamentos e sábios conselhos não só profissionais, mas também pessoais; pela oportunidade e por acreditar no meu potencial me instigando sempre a crescer cada dia mais ao longo dessa jornada.

À minha coorientadora Dr. Mariana de Campos Telles, pelo apoio, pelas correções e disponibilidade em contribuir com o aperfeiçoamento deste trabalho.

Ao Carlos Vinicius, pelo apoio, companheirismo, meu parceiro de trabalho, quanto ensinamentos trocados.

À Kelly da Silva e Souza, pela paciência, apoio e por todos os ensinamentos, que proporcionaram a construção do capítulo III.

Ao Laboratório de Genética & Biodiversidade da UFG pela possibilidade de utilização do espaço físico, equipamentos e material de trabalho.

Ao Prof. Lázaro José Chaves, Prof. Ronaldo Veloso Naves, Prof<sup>a</sup>. Mariana de Campos Telles, Prof<sup>a</sup>. Thannya Nascimento Soares e por todos aqueles que contribuíram nas coletas cada folha e fruto de *Dipteryx alata* que viabilizaram esse trabalho. Obrigado pela dedicação e esforço.

Aos professores, Dr. Alexandre S. G. Coelho, Dr. José Alexandre Felizola Diniz-Filho, Dr.<sup>a</sup> Rosane Garcia Collevatti, Dr. Lázaro José Chaves, Dr. Ronaldo Veloso Naves, pelo apoio e ajuda no esclarecimento de dúvidas;

Aos membros da banca examinadora, por terem aceitado o convite, pela disponibilidade em participar do aperfeiçoamento deste trabalho;

Aos companheiros de caminhada Ueric, Vanessa, Thais, Rhewter, Lays, Ariane, Edivaldo, Priscila, Fernanda, Tatiane, Mariana, Elias, Ramilla, Nicole, Patrícia, Josimar, pelos momentos de descontração, pelas risadas e lágrimas, pelo carinho; pelo incentivo e oportunidade de convívio e em especial pela força!! A vocês deixo essa mensagem. “*Aqueles que passam por nós não vão sós. Deixam um pouco de si, levam um pouco de nós.*” **Antoine de Saint-Exupery**

À minha amiga irmã, Kássia Marques pelo apoio incondicional em todos os momentos, principalmente nos de incerteza, muito comuns para quem tenta trilhar novos caminhos; À minha amiga Warita Alves, pelo apoio, por fazer parte desse momento tão importante, me mostrando que a presença física nem sempre é tudo, que o importante é saber que em algum lugar terá sempre alguém torcendo por mim.

Ao meu esposo Giuliano, pelo amor e por compreender minha ausência em tantos momentos e compartilhar da minha ansiedade, pelo carinho, pelo incentivo e pela torcida.

Aos meus pais, que dignamente me apresentaram à importância da família e o caminho da honestidade e da persistência; À minha querida Avó Derciolina, por fazer de cada momento juntas uma inspiração para continuar lutando pela conquista desse grande objetivo! Aos demais familiares, pelo carinho e apoio.

À Deus por guiar meu caminho e colocar nele oportunidades tão grandiosas! Por me dar forças para seguir sempre em frente com os meus objetivos e não desanimar com as dificuldades.

## SUMÁRIO

<b>RESUMO</b> .....	11
<b>ABSTRACT</b> .....	12
<b>1 INTRODUÇÃO</b> .....	14
<b>2 DISCOVERY AND CHARACTERIZATION OF NEW MICROSATELLITE LOCI IN <i>Dipteryx alata</i> VOGEL (FABACEAE) USING NEXT-GENERATION SEQUENCING DATA</b> .....	17
<b>3 MATING SYSTEM AND POLLEN DISPERSAL IN <i>Dipteryx alata</i> VOGEL (FABACEAE): COMPARING IN SITU AND EX SITU CONDITIONS</b> .....	24
<b>4 EFEITO DA PAISAGEM NA ESTRUTURA GENÉTICA ESPACIAL LOCAL EM INDIVÍDUOS JUVENIS E ADULTOS DE <i>Dipteryx alata</i> Vogel (FABACEAE)</b> .....	61
<b>5 CONSIDERAÇÕES FINAIS</b> .....	91
<b>6 REFERÊNCIAS</b> .....	92

## RESUMO

GUIMARÃES, R. A. **Variabilidade genética em *Dipteryx alata* Vogel (Leguminosae): Novos marcadores microssatélites, Dispersão de pólen *in situ* e *ex situ* e Genética da paisagem.** 2018. 96 f. Tese (Doutorado em Genética e Melhoramento de Plantas) – Escola de Agronomia, Universidade Federal de Goiás, Goiânia, 2018<sup>1</sup>.

*Dipteryx alata* é uma árvore neotropical amplamente distribuída em todo o Cerrado brasileiro. Popularmente conhecida como baru, é uma das espécies do Cerrado mais promissoras para a domesticação e cultivo, em função do seu amplo potencial de uso. O objetivo geral do trabalho foi conhecer o sistema reprodutivo e a dinâmica do fluxo gênico via pólen em condições *in situ* e *ex situ* e verificar a influência da paisagem na estrutura genética intrapopulacional. Para alcançar estes objetivos foram desenvolvidos novos marcadores microssatélites para a espécie, utilizando tecnologias de sequenciamento de alto desempenho. Dentre os marcadores desenvolvidos para espécie foram selecionados os dez mais polimórficos para serem utilizados neste trabalho. A avaliação do sistema reprodutivo e da dispersão de pólen foi realizada em uma coleção de germoplasma (*ex situ*) localizada na Universidade Federal de Goiás e em uma população natural (*in situ*) no município de Orizona-GO, Brasil. A população natural também foi utilizada para avaliar a influência da paisagem na estrutura genética intrapopulacional em árvores adultas e juvenis. As condições *in situ* e *ex situ* apresentaram alta taxa de fecundação cruzada ( $t_m = 0,815$  e  $t_m = 0,934$ ), respectivamente, confirmando que as espécies de *D. alata* apresentaram um sistema misto de reprodução predominantemente alógamo. O número de sementes que compartilhavam os mesmos doadores de pólen foi alto, indicando um baixo número de doadores de pólen por árvore em ambas as condições. Isso pode estar relacionado à proximidade entre grupos de árvores, considerando que há menos polinização cruzada entre grupos de árvores mais distantes. Outro fator que pode ter influenciado a polinização cruzada a longas distâncias é a paisagem heterogênea devido à fragmentação do habitat. Os resultados indicam a presença de estrutura genética espacial (SGS) positiva e significativa em ambos os estágios de vida (adultos e juvenis) sendo o valor de  $S_p$  maior nos juvenis. As análises de agrupamento bayesiano mostraram a formação de dois grupos, tanto em adultos quanto em juvenis. Os indivíduos juvenis apresentaram maior interferência das composições da paisagem. Assim a inclusão das características da paisagem trouxe um ganho de explicação para o padrão de estrutura genética intrapopulacional na espécie *D. alata*. Estes resultados são importantes para subsidiar estratégias mais eficientes de conservação da espécie.

*Palavras-chave:* Análise de parentesco, Fluxo gênico, *Illumina MiSeq*, Microssatélites, *Spatial Genetic Structure* - SGS.

---

<sup>1</sup> Orientadora: Dr<sup>a</sup> Thannya Nascimento Soares. ICB-UFG  
Co-orientadora: Dr<sup>a</sup> Mariana Pires de Campos Telles. ICB-UFG

## ABSTRACT

GUIMARÃES, R. A. **Genetic variability in *Dipteryx alata* Vogel (Leguminosae): New microsatellite markers, Dispersion of pollen *in situ* and *ex situ* and Landscape genetics.** 2018. 96 f. Tese (Doctor of Science in Genetics and Plant Breeding) – Escola de Agronomia, Universidade Federal de Goiás, Goiânia, 2018<sup>1</sup>.

*Dipteryx alata* is a Neotropical tree widely distributed throughout the Brazilian Cerrado. Popularly known as baru, it is one of the most promising Cerrado species for domestication and cultivation due to its wide potential use. The general goal of this work was to assess reproductive system and pollen-mediated gene flow patterns of *in situ* and *ex situ* conditions and to evaluate how landscape may influence intrapopulation genetic structure. New microsatellite markers were developed using high performance sequencing technologies, then the ten most polymorphic were selected and used in this work. The evaluation of the reproductive system and pollen dispersal was carried out in a germplasm collection (*ex situ*) located at the Universidade Federal de Goiás and in one natural population (*in situ*) in Orizona-GO, central Brazil. The natural population was also used to evaluate landscape influence on intrapopulation genetic structure in adult and juvenile trees. *In situ* and *ex situ* conditions showed a high cross fertilization rate ( $t_m = 0.815$  and  $t_m = 0.934$ ), respectively, confirming that the *D. alata* presents a mixed-mating system, predominantly allogamous. The number of seeds sharing the same pollen donors was high, indicating a low number of pollen donors per tree under both conditions. This may be related to the proximity between groups of trees, considering that there is less cross-pollination between groups of more distant trees. Another factor that may have influenced cross-pollination over long distances is the heterogeneous landscape due to habitat fragmentation. The results indicate the presence of a positive and significant *spatial genetic structure* (SGS) in both stages of life (adults and juveniles) with the highest  $S_p$  value in juveniles. Bayesian cluster analysis showed the formation of two groups in both adults and juveniles. Juvenile individuals presented greater interference from landscape compositions. Thus the inclusion of the landscape features brought an explanation gain to the pattern of intrapopulation genetic structure in the *D. alata* species. These results are important to support more efficient conservation strategies for this species.

**Keywords:** Parentage analysis, Gene flow, Illumina MiSeq, Microsatellites, *Spatial Genetic Structure* - SGS.

---

<sup>1</sup> Advisor: Dr<sup>a</sup> Thannya Nascimento Soares. ICB-UFG

Co-Advisor: Dr<sup>a</sup> Mariana Pires de Campos Telles. ICB-UFG

# 1 INTRODUÇÃO

*Dipteryx alata* Vogel é uma espécie arbórea neotropical nativa, endêmica e com ampla distribuição no Cerrado brasileiro. As populações naturais de *D. alata* formam grandes agrupamentos homogêneos de indivíduos nas áreas de solos mais férteis do bioma e ausência quase total em outras áreas menos favoráveis à ocorrência da espécie (Almeida et al., 1998; Sano et al., 2004; Nabout et al., 2010).

Popularmente, a espécie *D. alata* é conhecida como baru e é uma das espécies frutíferas nativas mais utilizadas pela população da região Centro-oeste do Brasil. O principal uso do baru é o alimentício, no qual a sua amêndoa é consumida torrada ou para compor diferentes tipos de receitas doces ou salgadas (Rocha & Santiago, 2009; Pineli et al., 2015). Sano et al. (2016) discutem que a exploração comercial da amêndoa do baru teve início em 1990 e que desde então, o seu uso se dá de forma extrativista, o que pode levar ao declínio das populações por falta de um manejo sustentável. Assim, o cultivo do baru passa a ser uma forma mais eficiente de exploração desta espécie, que se apresenta com um importante recurso vegetal do Cerrado.

Para a domesticação e cultivo de uma espécie nativa é necessário que haja amplo conhecimento sobre suas características biológicas e agronômicas. Diante de tal fato, vários grupos de pesquisa do Brasil já realizou diversos estudos com *D. alata*, como: aspectos de uso alimentício, nutricional e medicinal (Takemoto et al., 2001; Rocha & Santiago, 2009; Batista et al., 2012; Czeder et al., 2012; Lemos et al., 2012; Siqueira et al., 2012; Fraguas et al., 2014; Pineli et al., 2015; Silva et al., 2016), aspectos agronômicos e silviculturais (Monteiro et al., 2013; Mota, 2013; Pereira et al., 2013; Santos et al., 2014; Sano, 2016), ecológicos (Wolf, 2015; Ferreira et al., 2017), a variabilidade genética populacional (Soares et al., 2008; Collevatti et al., 2010; Telles et al., 2014; Soares et al. 2015) e conservação *in situ* e *ex situ* (Melo et al., 2011; Diniz-Filho et al., 2012; Canuto et al., 2015).

Visando a conservação *ex situ*, na qual a diversidade genética é conservada fora do hábitat natural (Havens et al., 2006; Hoban & Schlarbaum, 2014), foram criadas duas coleções de germoplasma de *D. alata* na Escola de Agronomia da Universidade Federal de Goiás. A primeira coleção foi montada em 1998 a partir de 150 árvores-mães, com no máximo duas sementes da mesma árvore. A amostragem das sementes foi realizada em áreas de Cerrado do estado de Goiás (Corrêa et al., 2000). A segunda coleção foi criada em 2011,

e é composta por 600 indivíduos oriundos de 25 populações coletadas em áreas de ocorrência da espécie no Cerrado brasileiro (Mota 2013).

A determinação dos padrões de distribuição da variabilidade genética dentro de uma espécie em condições *in situ* é importante para o conhecimento da atuação dos processos microevolutivos. Dentre estes processos, o fluxo gênico é responsável pela homogeneização da variabilidade genética entre as populações o que, conseqüentemente, diminui os efeitos da deriva genética (Epperson, 2003; Ellstrand, 2014). Análises anteriores revelaram baixa diversidade genética e estrutura genética com padrões moderados a altos em populações de *D. alata* (Melo et al., 2011; Collevatti et al., 2013; Tambarussi et al., 2017). Além disso, a espécie exibe endogamia, com valores significativos para o índice de fixação intrapopulacional (Tarazi et al. 2010). A variabilidade genética das populações da espécie é espacialmente estruturada e pode estar relacionada às condições climática e à ocupação humana e fragmentação do habitat no Cerrado (Soares et al., 2008; Collevatti et al., 2010; Telles et al., 2014; Soares et al., 2015).

Devido à baixa diversidade genética exibida pela espécie *D. alata*, esta não dispunha de marcadores microssatélites suficientes (em número e polimorfismo) para a condução de análises de identificação individual, que são necessárias, por exemplo, para o estudo de fluxo gênico via pólen, com a determinação de paternidade das sementes. O advento das tecnologias de sequenciamento de nova geração (NGS) tornou possível a identificação de grande número de marcadores microssatélites com reduzidos custo e esforço, quando comparados às abordagens tradicionais (Zalapa et al., 2012; Unamba et al., 2015).

Outra lacuna de conhecimento sobre a diversidade genética da espécie *D. alata* é o de se saber qual o impacto da paisagem na estrutura genética espacial em escala local. Considerando que a estrutura genética espacial também pode ser afetada pela fragmentação do habitat (Barluenga et al., 2011; Collevatti et al., 2014; Costa et al., 2017). Entretanto, os estudos de genética da paisagem em plantas têm sido realizados em escala regional, utilizando-se populações como as unidades amostrais (Manel et al., 2003, 2010; Balkenhol et al., 2015). São raros os estudos com plantas que exploram a influência da paisagem em escala local, nos quais os indivíduos são as unidades amostrais (Cushman & Landguth, 2010a, 2010b; Segelbacher et al., 2010; Kierepka & Latch, 2015).

Diante do status atual de conhecimento sobre a variabilidade genética *D. alata*, os objetivos deste trabalho foram conhecer o sistema reprodutivo e a dinâmica do fluxo

gênico via pólen em condições *in situ* e *ex situ* e verificar a influência da paisagem na estrutura genética em escala local nesta espécie. Para atingir aos objetivos propostos foram desenvolvidos novos marcadores microssatélites para *D. alata*, a partir de dados de NGS, com o intuito de aumentar o conjunto de marcadores de microssatélites polimórficos disponíveis para a espécie (capítulo 1). O capítulo 2 aborda o sistema reprodutivo e a análise de dispersão de pólen, em condições "*in situ*" e "*ex situ*" na espécie *D. alata*; e no terceiro capítulo é apresentado o estudo que testa se a estrutura genética em escala local, em indivíduos adultos e juvenis, é explicada somente pela distância geográfica ou se existe interferência da composição da paisagem neste processo, utilizando uma população de *D. alata* nativa do Cerrado.

## CAPÍTULO 1

# DISCOVERY AND CHARACTERIZATION OF NEW MICROSATELLITE LOCI IN *Dipteryx alata* Vogel (Fabaceae) USING NEXT-GENERATION SEQUENCING DATA<sup>1</sup>

Rejane Araújo Guimarães<sup>2</sup>, Mariana Pires de Campos Telles<sup>2,3</sup>, Adriana Maria Antunes<sup>2</sup>, Kássia Marques Corrêa<sup>2</sup>, Carlos Vinicius Gonçalves Ribeiro<sup>2</sup>, Alexandre Siqueira Guedes Coelho<sup>4</sup>; Thannya Nascimento Soares<sup>2</sup>

---

<sup>1</sup> Capítulo publicado como artigo no periódico científico *Genetics and Molecular Research*;

<sup>2</sup>Laboratório de Genética & Biodiversidade, Instituto de Ciências Biológicas, Universidade Federal de Goiás, Goiânia, GO, Brasil.

<sup>3</sup>Escola de Ciências Agrárias e Biológicas, Pontifícia Universidade Católica de Goiás, Goiânia, GO, Brasil. <sup>4</sup>Escola de Agronomia, Universidade Federal de Goiás, Goiânia, GO, Brasil.

GUIMARÃES, R. A.; TELLES, M. P. C.; ANTUNES, A. M.; CORRÊA, K. M.; RIBEIRO, C. V. G.; COELHO, A. S. G.; SOARES, T. N. Discovery and characterization of new microsatellite loci in *Dipteryx alata* Vogel (Fabaceae) using next-generation sequencing data. **Genetics and Molecular Research**, v. 16, p. 1–6, 2017.

## Discovery and characterization of new microsatellite loci in *Dipteryx alata* Vogel (Fabaceae) using next-generation sequencing data

R.A. Guimarães<sup>1</sup>, M.P.C. Telles<sup>1,2</sup>, A.M. Antunes<sup>1</sup>, K.M. Corrêa<sup>1</sup>,  
C.V.G. Ribeiro<sup>1</sup>, A.S.G. Coelho<sup>3</sup> and T.N. Soares<sup>1</sup>

<sup>1</sup>Laboratório de Genética & Biodiversidade, Instituto de Ciências Biológicas,  
Universidade Federal de Goiás, Goiânia, GO, Brasil

<sup>2</sup>Escola de Ciências Agrárias e Biológicas,  
Pontifícia Universidade Católica de Goiás, Goiânia, GO, Brasil

<sup>3</sup>Escola de Agronomia, Universidade Federal de Goiás, Goiânia, GO, Brasil

Corresponding authors: R.A. Guimarães / T.N. Soares

E-mail: [rejanearaujog@hotmail.com](mailto:rejanearaujog@hotmail.com) / [tsoares@gmail.com](mailto:tsoares@gmail.com)

Genet. Mol. Res. 16 (2): gmr16029639

Received February 2, 2017

Accepted March 7, 2017

Published April 28, 2017

DOI <http://dx.doi.org/10.4238/gmr16029639>

Copyright © 2017 The Authors. This is an open-access article distributed under the terms of the Creative Commons Attribution ShareAlike (CC BY-SA) 4.0 License.

**ABSTRACT.** The use of next-generation sequencing (NGS) technologies provides a great volume of genome sequence data even for non-model species. The development of microsatellite markers using these data is a relatively quick and easy process. *Dipteryx alata* Vogel (Fabaceae) is an arboreal species from the Cerrado biome and is considered an important plant genetic resource. Here, we report the development of microsatellite markers for *D. alata* using NGS data. DNA samples from four individuals were sequenced using the Illumina MiSeq platform and high-quality reads were assembled into contigs of the *D. alata* genome sequence. Microsatellite regions were identified using the IMEX webserver and primer pairs were designed using the Primer3 software. The amplification settings for each locus were optimized. Fluorescent-labeled primers were developed and used to genotype individuals derived from three natural populations of *D.*

*alata*. Fifty-four microsatellite regions were identified, from which 27 were elected to primer design. Among the amplified loci, 11 were polymorphic, with the number of alleles ranging from 2 to 10. The expected heterozygosity under Hardy-Weinberg Equilibrium (HWE) per locus varied from 0.191 to 0.807. Genotype and allele frequencies for all loci agreed with those expected under HWE and linkage disequilibrium was not significant for all pairs of loci. The probabilities of exclusion of paternity and of combined identity were equal to 0.993 and  $5.65 \times 10^{-8}$ , respectively. The markers developed in this study are useful to several types of population genetic studies with *D. alata* and, eventually, for closely related species.

**Key words:** Baru; Cerrado; Illumina MiSeq; SSR

## INTRODUCTION

The species *Dipteryx alata* Vogel belongs to the Fabaceae family and is popularly known as Baru. Baru is widely used in the food and pharmaceutical industry, thus presenting good economic potential (Pineli et al., 2015). This economic value makes *D. alata* a very important genetic resource (Sano et al., 2004).

The availability of microsatellite markers for *D. alata* enables the development of population genetic studies with the species and contributes to increase our knowledge about microevolutionary processes in the Cerrado biome. To date, only a small number of microsatellite markers are available for *D. alata*, each presenting a low level of polymorphism, restricting the type of studies that can be performed (Tarazi et al., 2010; Melo et al., 2011; Soares et al., 2012).

The development of microsatellite markers has traditionally been a difficult and expensive process. Next-generation sequencing technologies (NGS) enable efficient identification of large numbers of microsatellites at a fraction of the cost and effort compared to traditional approaches (Zalapa et al., 2012). In the present study, we report the development of new microsatellite markers for *D. alata* from NGS data, increasing the set of polymorphic microsatellite markers available for the species.

## MATERIAL AND METHODS

DNA was extracted from leaf tissues from four *D. alata* individuals, using the CTAB (2%) protocol described by Doyle and Doyle (1987). The four DNA samples were used in the preparation of four barcoded libraries following the Nextera Illumina protocol (Illumina, Inc.). Quantification and validation procedures for each library were performed in Bioanalyzer 2100 (Agilent Technologies) using the High Sensitivity DNA kit, and by real-time PCR using the Kappa kit. The libraries were pooled and sequenced on the Illumina MiSeq platform using the MiSeq Reagent Kit v3 (PE 2 x 300 bp).

Quality control of sequencing data was performed using FastQC (Andrews, 2010) and Trimmomatic (Bolger et al., 2014). Contigs were assembled using MaSuRCA (Zimin et al., 2013). Contigs were submitted to the IMEX webserver (Imperfect Microsatellite Extractor Webserver) (Mudunuri and Nagarajaram, 2007) to identify perfect and imperfect

(<10%) microsatellite regions in the *D. alata* genome. Critical size parameters considered in the search for microsatellites regions were: for dinucleotides, a minimum of 20 tandem repeats; for tetranucleotides, pentanucleotides, and hexanucleotides, a minimum of 8 tandem repeats. Mononucleotide and trinucleotide microsatellites were removed from the analysis. Mononucleotide microsatellite loci present high genotyping errors and trinucleotides are more abundant in coding regions of the genome (Tóth et al., 2000).

Primer pairs were designed using the Primer3 software (Rozen and Skaletsky, 2000) with the following parameters: amplicon size between 100 and 400 bp, primer length between 20 and 24 bp, melting temperature between 57° and 63°C, and CG content between 40 and 80%.

Designed primer pairs were tested for amplification in PCRs with different annealing temperatures using *D. alata* DNA samples. The reactions were assembled to a final volume of 10 µL using 7.5 ng DNA and 0.22 µM primers (forward + reverse), 0.23 µM dNTPs, 3.25 mg BSA, 1X buffer (10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM MgCl<sub>2</sub>), 0.75 U Taq DNA polymerase. The amplification process followed the steps of initial denaturation at 94°C for 5 min; 30 cycles of 94°C for 1 min, 58°-62°C (depending on primer) for 1 min, and 72°C for 1 min; and a final extension of 72°C for 45 min.

Forward sequences of selected primer pairs were labeled with one of the four fluorescent dyes (VIC, NED, 6-FAM, or PET). Lengths of the amplification products were determined using the internal marker GeneScan 600 LIZ (Applied Biosystems), in the ABI PRISM® 3500 DNA Genetic Analyzer (Applied Biosystems). Microsatellite loci that presented a good pattern in capillary electrophoresis were arranged in multiplex panels for routine analysis.

The levels of polymorphism of the developed markers were evaluated in 72 individuals, derived from three natural populations of *D. alata*. The sampled natural populations are from three geographically distant municipalities in Brazil: Pirenópolis, GO (15°59'82.9"S, 49°02'06.9"W), Sonora, MS (17°51'18.1"S, 54°42'21.1"W), and Alvorada, TO (12°26'92.8"S, 49°06'88.0"W).

After PCR amplification and capillary electrophoresis, genotypes were called using GeneMapper (Applied Biosystems). Genetic diversity of each locus in each population was evaluated using the following parameters: average number of alleles per locus and per population ( $N_A$ ), observed heterozygosity ( $H_O$ ), expected heterozygosity ( $H_E$ ) under Hardy-Weinberg equilibrium (HWE). Intrapopulation fixation index ( $F_{IS}$ ) was also estimated for each locus in each population. The significance of genotypic linkage disequilibrium estimates for all pairs of loci was also evaluated. These analyses were performed in the FSTAT 2.9.3 program (Goudet, 2002). The power of individual discrimination for each locus and for the total set of loci was evaluated by estimating the probability of genetic identity (I) and the probability of exclusion of paternity (Q), with the aid of the IDENTITY 4.0 program (Wagner and Sefc, 1999).

## RESULTS

The assembled contigs comprised 27 Mb of the genomic sequence of *D. alata*. Using the IMEx webserver, 46 imperfect and 8 perfect microsatellite regions were identified. The most frequent identified class was imperfect dinucleotide microsatellites (56.52%), followed by microsatellites with tetranucleotide (28.26%), pentanucleotide (4.34%), and hexanucleotide (10.86%) repeat motifs. Twenty-seven microsatellite regions with favorable characteristics for designing primers were obtained. Among the 27 designed pairs of primers, 12 presented a good

amplification pattern, although only 11 were polymorphic (Table 1). The overall probability of exclusion and probability of identity was 0.993 and  $5.65 \times 10^{-8}$ , respectively. The mean number of alleles per locus was 4.64, ranging from 2 to 10 alleles. The average estimates of  $H_O$  and  $H_E$  under HWE were 0.348 and 0.515, respectively. The deviations from the expected genotype proportions under HWE given the allele frequencies were not significant for any loci. Using the Bonferroni criteria, no linkage disequilibrium were detected for any pair of loci in the three populations.

The 11 polymorphic microsatellite markers were arranged in three multiplex panels (Table 2) for capillary electrophoresis for the population analysis. The mean  $N_A$  per population ranged from 2.9 to 3.8. Estimates of expected heterozygosity under HWE conditions for each population ranged from 0.363 to 0.469. A significant value (0.301) for the  $F_{IS}$  was found only for the population of Alvorada, TO (Table 3).

**Table 1.** Genetic parameters estimates for the 11 microsatellite markers, developed for *Dipteryx alata*.

Marker	GenBank accession No.	Primer sequence (5'→3')	Motif	Ta (°C)	$N_A$	$H_E$	$H_O$	$Q$	$I$
Dal11	KX427119	F: CATTGCCCTTCTGTCTTT	(TC) <sub>22</sub>	62	2	0.407	0.186	0.332	0.220
		R: AATGCCGATAATTTGTGTGTC							
Dal12	KX427121	F: TGCTGCGTTTCATTTTATAGTTTT	(TC) <sub>22</sub>	58	4	0.445	0.127	0.201	0.376
		R: CTTTCTCTTTGGGAGTTTGCT							
Dal14	KX427125	F: CATCTCACAAAGCCATACAGA	(AG) <sub>21</sub>	58	5	0.473	0.429	0.279	0.295
		R: TTGTTGCTTCCCGTTTTCTC							
Dal15	KX427115	F: CAACTCCATCAACCAATACACC	(CAGGCA) <sub>9</sub>	58	6	0.463	0.361	0.267	0.321
		R: AAATGAACCCCTCCAACACTT							
Dal16	KX427122	F: AATTGCGAGGCACAAAACT	(TC) <sub>22</sub>	58	4	0.721	0.322	0.560	0.086
		R: TGAATGATAATGGGGGCAAA							
Dal18	KX427124	F: CGATAAGCATTTACTATTTCCCTTT	(AG) <sub>22</sub>	58	6	0.609	0.592	0.376	0.196
		R: GTGATTGACATCTAACCTCCTCT							
Dal20	KX427118	F: CACGACTAGGAACCCCTTATTTT	(TC) <sub>27</sub>	58	5	0.191	0.174	0.141	0.563
		R: TCTTGTCTATCCTTTCCCTTTG							
Dal21	KX427120	F: TCTAGCCTCAACACACTGCTTC	(TC) <sub>21</sub>	62	4	0.530	0.250	0.339	0.227
		R: CAAGAAAGATGATATGGGAAAAGG							
Dal23	KX427116	F: TAGACAAAGTGCTTGGGGAAA	(AG) <sub>23</sub>	58	10	0.807	0.772	0.672	0.047
		R: TTCTTGATTTTGGATCTCTATCG							
Dal24	KX427123	F: ATTTGAGGAAGTCTCTTTGGT	(AC) <sub>22</sub>	58	2	0.396	0.358	0.230	0.344
		R: TGGAGTTCATATCCTTATCTTTG							
Dal25	KX427117	F: CCAGGTGGTGGGATGAGATA	(ATTGAG) <sub>5</sub>	58	3	0.595	0.261	0.364	0.199
		R: ATGATGGACACGAAAAGTGAA							

Ta: Annealing temperature;  $N_A$ : number of alleles;  $H_E$ : expected heterozygosity under HWE;  $H_O$ : observed heterozygosity;  $Q$ : probability of exclusion paternity;  $I$ : probability of genetic identity.

**Table 2.** Panels of microsatellite markers for *Dipteryx alata*, arranged in multiplex for genotyping in capillary electrophoresis.

Multiplex	Markers	Dye	Allele size range (bp)
1	Dal15	6FAM	190-214
	Dal23	NED	134-172
	Dal25	VIC	152-164
	Dal20	PET	184-208
2	Dal11	NED	162-170
	Dal12	6FAM	188-194
	Dal16	VIC	174-190
	Dal21	6FAM	158-162
3	Dal14	6FAM	164-172
	Dal18	VIC	186-198
	Dal24	VIC	124-128

**Table 3.** Genetic characterization of 11 microsatellite loci in three natural populations of *Dipteryx alata* (n = 3 x 24).

Loci	Pirenópolis-GO				Sonora-MS				Alvorada-TO			
	$N_A$	$H_E$	$H_O$	$F_{IS}$	$N_A$	$H_E$	$H_O$	$F_{IS}$	$N_A$	$H_E$	$H_O$	$F_{IS}$
<i>Dal15</i>	3	0.228	0.167	0.273	6	0.665	0.542	0.189	3	0.414	0.375	0.096
<i>Dal23</i>	4	0.683	0.739	-0.084	8	0.810	0.938	-0.163	3	0.646	0.667	-0.033
<i>Dal25</i>	2	0.496	0.250	0.502	3	0.318	0.208	0.350	3	0.528	0.333	0.375
<i>Dal20</i>	3	0.228	0.250	-0.100	1	0.000	0.000	0.000	5	0.327	0.273	0.168
<i>Dal11</i>	2	0.413	0.222	0.469	2	0.480	0.067	0.865*	2	0.268	0.300	-0.125
<i>Dal12</i>	2	0.043	0.043	0.000	4	0.264	0.292	-0.107	2	0.042	0.042	0.000
<i>Dal16</i>	3	0.412	0.300	0.276	3	0.513	0.455	0.116	4	0.619	0.176	0.721*
<i>Dal21</i>	4	0.241	0.087	0.644	3	0.616	0.500	0.192	3	0.528	0.143	0.735*
<i>Dal14</i>	3	0.377	0.375	0.005	5	0.673	0.864	-0.291	3	0.082	0.083	-0.011
<i>Dal18</i>	5	0.382	0.391	-0.026	5	0.649	0.792	-0.226	4	0.665	0.583	0.125
<i>Dal24</i>	2	0.488	0.625	-0.287	2	0.169	0.182	-0.077	2	0.438	0.238	0.462
Average	2.9	0.363	0.314	0.138	3.8	0.469	0.440	0.063	3.0	0.414	0.292	0.301*

$N_A$ : number of alleles;  $H_E$ : expected heterozygosity under HWE;  $H_O$ : observed heterozygosity;  $F_{IS}$ : intrapopulation fixation index. \*Statistically different from 0.000.

## DISCUSSION

Several studies have demonstrated success in the development of microsatellite markers for native Cerrado species (Croft and Schaal, 2012; Soares et al., 2012; Bernardes et al., 2014). However, these studies use methodologies that limit the quantity and quality of microsatellite markers obtained. In the present study, using NGS data, 54 microsatellites were detected in a relatively quick and easy way. From the 54 pre-selected microsatellites, further selection was made based on favorable characteristics for design and amplification, resulting in 11 polymorphic microsatellite markers. This short set of markers showed a diversity of repeat motifs, dinucleotides being the most frequent, but also tetra-, penta-, and hexanucleotides. The combined probability of exclusion and of identity showed that the set of developed microsatellite loci has a high power of exclusion of false paternities (Evet and Weir, 1998) and is suitable for individual discrimination (Paetkau et al., 1995).

The polymorphism found in this study for *D. alata* was similar to the observed by Tarazi et al. (2010), Melo et al. (2011), and Soares et al. (2012), who found an average  $N_A$  per locus of 4.7, 2.7 and 4.3, respectively. The genetic diversity estimates observed in this study is also similar to other studies with *D. alata* populations. The increase in the number of available markers will allow the use of a wider set of informative markers in more powerful genetic studies. Of the 27 pairs of primers designed in the present study, 11 were polymorphic in *D. alata*. It is suggested that the set of 27 primers be tested for transferability to other Cerrado species of the Fabaceae family, further reducing the costs in obtaining this kind of markers.

## Conflicts of interest

The authors declare no conflict of interest.

## ACKNOWLEDGMENTS

Research has been continuously supported by several grants and fellowships to the research network GENPAC (Geographical Genetics and Regional Planning for Natural Resources in Brazilian Cerrado) from CNPq/FAPEG (projects #563839/2010-4 and #201110267000125), by the CERGEN (Núcleo de Excelência em Recursos Genéticos Vegetais

do Cerrado - PRONEX/FAPEG/CNPq 07-2012) and our study has also been developed in the context of the National Institutes for Science and Technology (INCT) in Ecology, Evolution and Biodiversity Conservation, supported by MCTIC/CNPq (process #465610/2014-5) and FAPEG, besides support of PPGS CAPES/FAPEG (Public Call #08/2014). R.A. Guimarães, A.M. Antunes, and K.M. Corrêa were supported by fellowships from CAPES. M.P.C. Telles and T.N. Soares are supported by productivity grants from CNPq.

## REFERENCES

- Andrews S (2010). FastQC a Quality Control Tool for High Throughput Sequence Data. Available at [<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>].
- Bernardes V, Dos Anjos DE, Gondim SG, Murakami DM, et al. (2014). Isolation and characterization of microsatellite loci in *Byrsonima cydoniifolia* (Malpighiaceae) and cross-amplification in *B. crassifolia*. *Appl. Plant Sci.* 2: 1400016. <http://dx.doi.org/10.3732/apps.1400016>
- Bolger AM, Lohse M and Usadel B (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114-2120. <http://dx.doi.org/10.1093/bioinformatics/btu170>
- Croft GK and Schaal BA (2012). Development of microsatellite markers in *Byrsonima crassifolia* (Malpighiaceae). *Am. J. Bot.* 99: e111-e113. <http://dx.doi.org/10.3732/ajb.1100457>
- Doyle JJ and Doyle JJ (1987). Isolation of plant DNA from fresh tissue. *Focus* 12: 13-15.
- Evetts IW and Weir BS (1998). Interpreting DNA evidence: Statistical for forensic scientists. Sinauer Associates, Inc. Publishers, Sunderland.
- Goudet J (2002). Fstat (Version 2.9.3.2.): A computer program to calculate F-statistics. *J. Hered.* 86: 485-486. <http://dx.doi.org/10.1093/oxfordjournals.jhered.a111627>
- Melo DB, Diniz-Filho JAF, Oliveira G, Santana LL, et al. (2011). Optimizing sampling efforts for *ex situ* conservation of genetic variability of *Dipteryx alata* Vogel. *BMC Proc.* 5: 18. <http://dx.doi.org/10.1186/1753-6561-5-S7-P18>
- Mudunuri SB and Nagarajaram HA (2007). IMEX: Imperfect Microsatellite Extractor. *Bioinformatics* 23: 1181-1187. <http://dx.doi.org/10.1093/bioinformatics/btm097>
- Paetkau D, Calvert W, Stirling I and Strobeck C (1995). Microsatellite analysis of population structure in Canadian polar bears. *Mol. Ecol.* 4: 347-354. <http://dx.doi.org/10.1111/j.1365-294X.1995.tb00227.x>
- Pineli LLO, Carvalho MV, Aguiar LA, Oliveira GT, et al. (2015). Use of baru (Brazilian almond) waste from physical extraction of oil to produce flour and cookies. *LWT - Food Sci. Technol. (Campinas)* 60: 50-55.
- Rozen S and Skaletsky HJ (2000). Primer3: Bioinformatics Methods and Protocols. In: *Methods in Molecular Biology* (Krawetz S and Misener S, eds.). Humana Press, New Jersey, 365-386. Available at [[http://frodo.wi.mit.edu/cgi-bin/primer3/primer3\\_www.cgi](http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi)].
- Sano SM, Ribeiro JF and Brito MA (2004). Baru: Biologia e Uso. Embrapa Cerrados, Brasília, DF, Brazil.
- Soares TN, Melo DB, Resende LV, Vianello RP, et al. (2012). Development of microsatellite markers for the neotropical tree species *Dipteryx alata* (Fabaceae). *Am. J. Bot.* 99: e72-e73. <http://dx.doi.org/10.3732/ajb.1100377>
- Tarazi R, Moreno MA, Gandara FB, Ferraz EM, et al. (2010). High levels of genetic differentiation and selfing in the Brazilian cerrado fruit tree *Dipteryx alata* Vog. (Fabaceae). *Genet. Mol. Biol.* 33: 78-85. <http://dx.doi.org/10.1590/S1415-47572010005000007>
- Tóth G, Gáspári Z and Jurka J (2000). Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res.* 10: 967-981. <http://dx.doi.org/10.1101/gr.10.7.967>
- Wagner HW and Sefc KM (1999). IDENTITY 1.0. Centre for Applied Genetics, University of Agricultural Sciences, Vienna. Available at [<http://www.uni-graz.at/~sefck>].
- Zalapa JE, Cuevas H, Zhu H, Steffan S, et al. (2012). Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. *Am. J. Bot.* 99: 193-208. <http://dx.doi.org/10.3732/ajb.1100394>
- Zimin AV, Marçais G, Puiu D, Roberts M, et al. (2013). The MaSuRCA genome assembler. *Bioinformatics* 29: 2669-2677. <http://dx.doi.org/10.1093/bioinformatics/btt476>

## CAPÍTULO 2

### **MATING SYSTEM AND POLLEN DISPERSAL IN *Dipteryx alata* VOGEL (FABACEAE): COMPARING IN SITU AND EX SITU CONDITIONS<sup>1</sup>**

Rejane Araújo Guimarães<sup>2,3</sup>, Kássia Marques Corrêa<sup>2,3</sup>, Lázaro José Chaves<sup>5</sup>, Mariana Pires de Campos Telles<sup>2,4</sup>; Thannya Nascimento Soares<sup>2</sup>

---

<sup>1</sup>Capítulo submetido como artigo ao periódico científico *Tree Genetics & Genomes*;

<sup>2</sup>Laboratório de Genética & Biodiversidade, Instituto de Ciências Biológicas, Universidade Federal de Goiás, Goiânia, Goiás, Brasil.

<sup>3</sup>Programa de Pós-graduação em Genética e Melhoramento de Plantas - Escola de Agronomia – UFG, Goiás, Brasil.

<sup>4</sup>Escola de Ciências Agrárias e Biológicas, Pontifícia Universidade Católica de Goiás, Goiânia, Goiás, Brasil

<sup>5</sup>Escola de Agronomia, Universidade Federal de Goiás, Goiânia, Goiás, Brasil.

1 **Mating system and pollen dispersal in *Dipteryx alata* Vogel (Fabaceae): comparing *in***  
2 ***situ* and *ex situ* conditions**

3

4 Rejane Araújo Guimarães<sup>1,2</sup>, Kássia Marques Corrêa<sup>1,2</sup>, Lázaro José Chaves<sup>4</sup>, Ronaldo  
5 Veloso Naves<sup>4</sup>, Mariana Pires de Campos Telles<sup>1,3</sup>, Thannya Nascimento Soares<sup>1,2</sup>.

6

7 <sup>1</sup>Laboratório de Genética & Biodiversidade, Instituto de Ciências Biológicas, Universidade  
8 Federal de Goiás, Goiânia, Goiás, Brasil.

9 <sup>2</sup>Programa de Pós-graduação em Genética e Melhoramento de Plantas - Escola de  
10 Agronomia – UFG, Goiás, Brasil.

11 <sup>3</sup>Escola de Ciências Agrárias e Biológicas, Pontifícia Universidade Católica de Goiás,  
12 Goiânia, Goiás, Brasil

13 <sup>4</sup>Escola de Agronomia, Universidade Federal de Goiás, Goiânia, Goiás, Brasil.

14

15 ORCID: [orcid.org/0000-0003-0186-6589](https://orcid.org/0000-0003-0186-6589)

16 Corresponding author: R. A. Guimarães/ T. N. Soares

17 E-mail: [rejanearaujog@gmail.com](mailto:rejanearaujog@gmail.com) / [tsoares@gmail.com](mailto:tsoares@gmail.com)

18

19 **Abstract**

20

21 *Dipteryx alata* is a Neotropical tree widely that is distributed throughout the Brazilian  
22 Cerrado biome and is popularly known as baru. In this work, we evaluated the mating  
23 system of *D. alata* and compared pollen dispersal patterns between *in situ* and *ex situ*  
24 conditions. For this, we used 515 genotypes of adults, juveniles and progeny from a natural  
25 population (*in situ*) in Orizona-GO, Brazil. In addition, we used 488 genotypes of adults

26 and progeny from a germplasm collection (*ex situ*) located at the Federal University of  
27 Goiás. Both locations are situated in central Brazil. The genetic diversity, cross-pollination  
28 rates and pollen dispersal distance under both conditions were estimated. Genetic diversity  
29 and polymorphism differed between the *in situ* and *ex situ* conditions. The average number  
30 of alleles found *in situ* (5.2) and *ex situ* (6.20) showed that the germplasm collection stores  
31 greater genetic diversity than do *in situ* conditions. Cross-pollination detected among  
32 mother trees under *in situ* and *ex situ* conditions was high ( $t_m=0.815$  and  $t_m=0.934$ ,  
33 respectively), indicating that the species presented a mixed reproduction system that was  
34 predominantly allogamous. The difference between  $t_m$  and  $t_s$  indicated that greater  
35 biparental inbreeding occurred under *in situ* conditions. Our results showed that the *ex situ*  
36 conditions presented greater efficiency in preserving a greater number of alleles,  
37 confirming that the *D. alata* germplasm collection plays a role in conserving genetic  
38 diversity. The presence of self-fertilization suggests self-compatibility. Paternity  
39 correlation and the dispersion distance of the pollen donors were higher under *in situ*  
40 conditions, which can be explained by the number of pollinators and the tree spatial  
41 distribution.

42 **Keywords:** Cerrado, Gene flow, Microsatellites, Parentage analysis.

43

## 44 **Introduction**

45

46 Gene flow is related to all mechanisms that result in the movement of alleles within  
47 populations. Therefore, the role of gene flow in genetic variability within populations  
48 depends both on how populations are distributed geographically and on the evolutionary  
49 factors acting on these populations (Slatkin 1987; Ellstrand 2014). The genetic variation  
50 observed in populations is the result of balancing evolutionary forces: forces that act

51 together to produce differentiation (genetic drift and diversifying selection) and forces that  
52 act to homogenize populations (gene flow) (Ellstrand 2014). Estimates of these  
53 evolutionary forces are influenced not only by the dispersal process but also by the  
54 reproductive success of the individual migrant and the size of the population receiving the  
55 migrant (Ennos 1994; Richardson et al. 2014). Estimating genetic variability patterns  
56 within populations is important for understanding evolutionary processes influencing  
57 populations. This knowledge allows us to formulate effective strategies for the  
58 conservation and maintenance of genetic resources (Slatkin 1987; Richardson et al. 2014).

59         In recent decades, strategies were developed to increase the conservation of plant  
60 genetic resources (Rao and Hodgkin 2002; Gepts 2006; Rajpurohit and Jhang 2015). *In situ*  
61 genetic resource conservation implies the possibility of continuous evolution and the  
62 development of new adaptive strategies in the natural environment of the species (Scariot  
63 and Servilha 2007). In this way, the occurrence of gene flow within and among populations  
64 affects the amount and structure of the genetic variability of these populations, a factor that  
65 must be considered for the conservation and management of plant genetic resources  
66 (Martins 1987; Williams 2017). However, the fragmentation of ecosystems has been a  
67 crucial problem in the conservation of plant genetic resources in Brazil (Santos-Filho  
68 1995), leading to the necessity of *ex situ* conservation. In *ex situ* conservation, the genetic  
69 resource can be stored in seed banks, pollen banks, and field genebanks and per *in vitro*  
70 storage methods (Engels and Visser 2003; Santonieri and Bustamante 2016). These forms  
71 of *ex situ* conservation are also known as germplasm collections or germplasm banks  
72 (Engels and Visser 2003), and they are complementary to *in situ* conservation for wild  
73 species.

74         Germplasm collections play an important role in the conservation of genetic  
75 diversity and the domestication of species because they are maintained to represent the

76 inherent genetic diversity of a species and of their wild relatives (Rao and Hodgkin 2002;  
77 Engels and Visser 2003). Rich genetic diversity is a prerequisite for a species  
78 domestication and breeding programme, but domestication sometimes decreases genetic  
79 diversity (Rao and Hodgkin 2002).

80         A small number of native fruit tree species have been domesticated, and fruit  
81 collection from natural populations is a frequent activity, especially in the Brazilian  
82 Cerrado (Felfili et al. 2004). The main challenges in the domestication and breeding of the  
83 fruit tree species in Cerrado are the long generation period (first fruiting takes  
84 approximately 6 years), the seasonality and the variation in its production of flowers and  
85 fruits each year (Felfili et al. 2004; Vieira et al. 2006). In addition, information on the  
86 mating system and the dispersal of pollen for most native trees is essential for the choice of  
87 breeding methods and the establishment of seed orchards (Cruz et al. 2011; Hoban and  
88 Strand 2015).

89         Studies that evaluate the mating system and the dispersion of pollen have  
90 demonstrated that microsatellite markers are tools of great importance for evaluating the  
91 genetic resources of Cerrado plants (Collevatti et al. 2010; Braga and Collevatti 2011;  
92 Moraes and Sebbenn 2011; Manoel et al. 2012; Antiqueira and Kageyama 2015;  
93 Rodrigues et al. 2016). *Dipteryx alata* is an economically important genetic resource  
94 that is widely distributed throughout the Cerrado biome and used for a wide variety of  
95 purposes, such as its edible cookies and oils (Pineli et al. 2015), timber, fodder, and  
96 industrial utility (Corrêa 1984; Almeida et al. 1998; Sano et al. 2004). However, the  
97 main product is the “baru” nut, with high added value. While *D. alata* has been the  
98 object of many studies (e.g., Soares et al. 2008; Collevatti et al. 2013; Santos et al.  
99 2014; Soares et al. 2015), its pollen dispersal and the mechanism of determining its  
100 mating system under "*in situ*" and "*ex situ*" conditions are unknown, and studies

101 involving different temporal scales (adults, juveniles and progeny) are lacking. Such  
102 studies can provide important information for breeding and conservation programmes  
103 for *D. alata*.

104 In this work, we intend to evaluate whether there is difference in the genetic  
105 diversity parameters between "*in situ*" and "*ex situ*" conditions. For this, we hypothesize  
106 that the germplasm collection presents a higher level of genetic diversity since individuals  
107 come from populations distributed throughout the state of Goiás (Corrêa et al. 2000), while  
108 *in situ* conditions are restrictive to a local population. We also compare the genetic  
109 diversity in adults in with the genetic diversity in progeny because it is important for the  
110 conservation of genetic diversity that the progeny recover the genetic diversity found in  
111 adults. Lastly, we compare pollen dispersal and the mating system between "*in situ*" and  
112 "*ex situ*" conditions to better understand the dynamics of gene flow and to generate useful  
113 information to support future breeding and conservation programmes for *D. alata*. For this,  
114 we analyse adults, juveniles and progeny individuals for paternity assignment. Due to the  
115 dispersion modes of pollen by animals and seed dispersal by barochory by *D. alata*, we  
116 hypothesize that "*in situ*" and "*ex situ*" conditions present a broad pollen flow, a restricted  
117 seed flow and a high cross-pollination rate.

118

## 119 **Materials and methods**

120

### 121 ***In situ* study site and sampling**

122

123 The "*in situ*" condition selected for this study was a natural population of *D. alata*  
124 located in Orizona, Goiás, Central Brazil (centred approximately 16°45'52.0"S and  
125 48°12'17.7"W, at 881 m altitude). This population is situated on private property with an

126 area of approximately 148 ha in a heterogeneous landscape composed of pasture,  
127 agricultural fields and forest remnants. This area is located in a region of tropical climate  
128 ( $A_w$ ) with a dry winter and hot summer, an average annual temperature of 22.9°C and high  
129 annual precipitation (1,360 mm) that is irregularly distributed throughout the year.

130 We sampled leaves of 151 *D. alata* trees that were either discontinuously  
131 distributed or isolated. The distance between sampled trees ranged from approximately 5 m  
132 to approximately 3 km. Thus, all the potential donors were sampled in the area screened.  
133 However, there may have been donor trees outside the selected area. Of the 151 trees, 69  
134 adult trees (height >2 m and presence of flowers or fruits) and 62 juvenile trees (ranging  
135 from seedlings to a height of approximately 2 m with no signs of flowering) were sampled.  
136 Additionally, twenty mother trees were randomly sampled to study pollen flow (Fig. 1).  
137 All individuals were mapped using a GPS (global positioning system) receptor and  
138 genotyped.

139 Seed sampling was carried out during the fruiting season (August and September)  
140 of 2014. We sampled 450 seeds, ranging from 20 to 30 per mother tree. These seeds were  
141 grown in a nursery in a completely randomized experimental design, and the leaf tissue of  
142 364 seedlings was collected for analysis in the present study. The number of seedlings  
143 analysed per mother tree differed due to variation in germination (see support information  
144 Table S5).

145

#### 146 ***Ex situ* study site and sampling**

147

148 The *ex situ in vivo* conditions were the germplasm collection maintained in an  
149 experimental area of the School of Agronomy of the Federal University of Goiás, in the  
150 city of Goiânia-Goiás, Central Brazil (EA/UFG, 16°35'38.64"S and 49°7'23.14"W, 750 m

151 altitude). This germplasm collection of *D. alata* encompasses an area of 4,518 m<sup>2</sup> in a  
152 region with a tropical climate (Aw) with a dry winter and hot summer, average annual  
153 temperature of 23.1°C and precipitation of 1,414 mm per year. This germplasm collection  
154 was created in January 1998 from 150 mother trees with at most two seeds from the same  
155 tree. The seeds were sampled in the Cerrado Biome of Goiás in 38 localities (see Corrêa et  
156 al. 2000 for more details).

157         The seeds were grown in a nursery, and 186 seedlings were planted in 6 × 6 m area.  
158 However, due to the death of 13 trees, the germplasm collection currently presents 173  
159 reproductive *D. alata* trees that have been collected and genotyped. However, there were  
160 other trees that were planted on the campus as well as isolated trees within the urban area  
161 at distances of up to 5 km that were not collected. Thus, there may have been donor trees  
162 outside the selected area.

163         In total, twenty mother trees were sampled during the fruiting season (August and  
164 September) of 2014 that were genotyped. Four hundred and fifty seeds were sampled,  
165 ranging from 15 to 30 per mother tree. These seeds were cultivated in a nursery in a  
166 completely randomized design, and the leaf tissue of 316 seedlings was collected and  
167 genotyped for analysis in the present study. The number of seedlings analysed per mother  
168 tree differed due to variation in germination (see supporting information Table S5).

169         The records containing the identifications and the places of collection of each  
170 individual in the germplasm collection were lost. There are only records of the collection  
171 points within the populations (Corrêa et al. 2000). All individuals were mapped using a  
172 handmade sketch (Fig. 2).

173

174

175

176 **Microsatellite analyses**

177

178 Genomic DNA extraction from leaf tissue was performed using a standard protocol of 2%  
179 CTAB (Doyle and Doyle 1987). All plants were genotyped using ten microsatellite loci  
180 previously developed for *D. alata*: DaE34, DaE41, and DaE63 (Soares et al. 2012) and  
181 Dal15, Dal20, Dal23, Dal25, Dal21, Dal18, and Da80 (Guimarães et al. 2017) (see  
182 supporting information Table S1). PCR amplifications were performed in a final volume of  
183 10  $\mu$ L using 7.5 ng of template DNA, 0.22  $\mu$ M *primers* (*forward + reverse*), 0.23  $\mu$ M dNTP,  
184 3.25 mg of bovine serum albumin (BSA), 1 $\times$  reaction buffer (10 mM Tris-HCl [pH 8.3], 50  
185 mM KCl, 1.5 mM MgCl<sub>2</sub>) and 0.75 unit of *Taq* DNA polymerase (5U; Phoneutria, Belo  
186 Horizonte, Brazil) under the following conditions: 94°C for 5 min; 30 cycles of 94°C for 1  
187 min, 58-62 ° C for 1 min, and 72 ° C for 1 min; and one cycle of 72°C for 45 min. PCR  
188 fragments were determined using an ABI PRISM® 3500 automated DNA sequencer and a  
189 GeneScan LIZ600™ size Standard (Applied Biosystems). Allele calling was performed  
190 using GeneMapper 5.0 software (Applied Biosystems). Genotypes were confirmed using the  
191 "allelic ladder" with new capillary electrophoresis and the minimum number of individuals  
192 (in duplicate) containing all alleles of a given locus. Then, Micro-Checker software  
193 (Oosterhout et al. 2004) was used to detect errors due to stutter bands, *drop-out* alleles and  
194 null alleles using only adult individuals from both *in situ* and *ex situ* conditions. In addition,  
195 the genotypes were visually reviewed to check the number of mismatches between mother  
196 plants and their progeny.

197 To distinguish inbreeding from the presence of null alleles, the INEST program v2.0  
198 (Chybicki and Burczyk 2009) was used, which uses an individual inbreeding approach  
199 model (IIM) to simultaneously estimate null allele frequencies at each locus and the average  
200 level of inbreeding. In the IIM, 500 000 Markov chain Monte Carlo (MCMC) iterations, 5

201 000 thinnings and a burn-in period of 50 000 were used. Then, Bayesian model comparison  
202 was performed. The INEST program computes the deviance information criterion (DIC) for  
203 each model according to Chybicki & Burczyk (2009); the lowest DIC corresponds to the  
204 model best fit to the data. Among the calculated models, we have  $n$ =null alleles,  
205  $f$ =inbreeding,  $b$ =genotyping error, and a combination of these.

206

## 207 **Genetic diversity**

208

209 The ten microsatellite markers were characterized using all individuals at different  
210 life stages from both *in situ* and *ex situ* conditions. We estimated the genetic parameters  
211 number of alleles per locus ( $A$ ), observed heterozygosity ( $H_o$ ) and expected heterozygosity  
212 under Hardy–Weinberg equilibrium ( $H_e$ ) using the software Genetic Data Analysis version  
213 1.0 (GDA) (Lewis and Zaykin 2001). We also estimated the power of individual  
214 discrimination over the total set of loci and over each locus, assessing probabilities of  
215 genetic identity (I) and paternity exclusion, using the software Identity 1.0 (Wagner and  
216 Sefc 1999).

217 To compare genetic diversity and polymorphism at the different life stages  
218 under both *in situ* and *ex situ* conditions, we estimated  $A$ ,  $H_o$ ,  $H_e$  and the inbreeding  
219 coefficient ( $F_{IS}$ ) using the software GDA (Lewis and Zaykin 2001). To test whether genetic  
220 parameters were significantly different in adults and progeny arrays, a Wilcoxon's paired-  
221 test was performed.

222 Bayesian clustering approaches implemented in STRUCTURE ver. 2.3.1  
223 (Pritchard et al. 2000) were used to check whether the *ex situ* collection was genetically  
224 similar to the natural population through a test of attribution and population differentiation.  
225 The number of populations ( $K$ ) was estimated with ten replicates each for  $K=1$  to  $K=10$

226 using 100,000 MCMC iterations and 100,000 burn-in periods using the admixture model.  
227 K was used to detect the most likely number of clusters (Evanno et al. 2005), and the  
228 computation was performed using the STRUCTURE HARVESTER program (Earl and  
229 VonHold 2012). Due to the presence of progeny structure, only adults were used in this  
230 analysis.

231

### 232 **Mating system**

233

234 We investigated the mating system based on mixed-mating models  
235 implemented in the software MLTR (Ritland 2002). The estimated parameters were single-  
236 locus ( $t_s$ ) and multilocus ( $t_m$ ) outcrossing rates; biparental inbreeding rates ( $t_m - t_s$ );  
237 correlation of outcrossed paternity within progeny arrays ( $r_p$ ), and correlation of self-  
238 fertilization between progeny arrays ( $r_s$ ). The confidence intervals (95%) of the parameter  
239 estimates were obtained with 10,000 randomizations.

240 The mean number of pollen donor trees ( $N_{ep}$ ) was calculated from the  $r_p$  value  
241 by the expression  $N_{ep} = 1/r_p$  (Ritland 2002). Posteriorly, the genetic structure within  
242 maternal families was evaluated according to Sebbenn (2002), combining the estimates of  
243  $t_m$  and  $r_p$  to determine the proportions of selfing sibs ( $P_{IA} = 1 - t_m$  is the selfing rate), half-sibs  
244 ( $P_{MI} = t_m(1 - r_p)$ ), and full-sibs ( $P_{IC} = r_p t_m$ ).

245

### 246 **Parentage analysis of paternity and dispersion of pollen**

247

248 Seed parentage was determined using an assignment test implemented in the  
249 software CERVUS 3.0 (Kalinowski et al. 2007). We considered all adults as candidate  
250 parents to determine the putative pollen donor of the 364 (*in situ*) and 316 (*ex situ*) seeds.

251 This number of seeds analysed differs from that of seeds collected due to variation in  
252 germination (see support information Table S5). Paternity was determined based on the  
253 critical values of delta ( $\Delta$ ) with simulation (10,000 replications); we considered a genotype  
254 error rate of 1%; 100% of the loci were genotyped, and 95% of the candidate parents were  
255 sampled. This percentage of candidate parents was attributed considering that in the area  
256 screened, all potential donors were sampled. However, there may have been donor trees  
257 outside the selected area. Critical values of  $\Delta$  were obtained from 10 000 parentage  
258 simulations at 95% (restricted) and 90% (relaxed) confidence. The dispersion distance of  
259 the pollen was obtained by checking the position of each mother tree and pollen donor.

260

## 261 **Results**

262

### 263 **Genetic diversity**

264

265 Analysis of raw data using the software Micro-Checker showed the presence of  
266 stuttering in loci Dal18, Dal15, Dal 23 and Dal21 and null alleles in loci Dal15, DaE63,  
267 DaE25, Dal20. This may have resulted from the excess homozygous individuals found.  
268 However, the overall values of probability of *I* and exclusion paternity (*Q*) under the *in situ*  
269 ( $4.69 \times 10^{-7}$  and 0.983) and *ex situ* ( $5.74 \times 10^{-8}$  and 0.994) conditions showed that the ten loci  
270 are suitable for parentage analysis (Table 1). Of the ten microsatellite loci analysed from  
271 the total 515 *in situ* and 488 *ex situ* samples, 52 and 62 alleles, respectively, were observed.  
272 The mean *A* was equal to 5.2 (*in situ*) and 6.2 (*ex situ*), ranging from 2 to 8 and 2 to 13,  
273 respectively, and showed greater polymorphism under *ex situ* conditions (Table 1). Under  
274 *in situ* conditions, the mean *A* in adults (3.6) and juveniles (3.7) were similar at the  
275 different life stages (Table S2). However, that in progeny arrays (4.8) differed

276 significantly from that in adults (Table 2). Different results were observed under *ex situ*  
277 conditions; the mean  $A$  in adults (6.8) and progeny arrays (6.3) did not significantly differ  
278 (Table 2). Almost all alleles were represented in the progeny but at different frequencies  
279 (see supporting information Table S2 and S3).

280         The  $H_e$  combining all loci was equal to 0.532 (*in situ*) and 0.566 (*ex situ*) (Table 1)  
281 and  $H_e$  did not significantly differ at the different life stages under either condition (Table  
282 2), showing that genetic diversity is being preserved in each generation (see supporting  
283 information Tables S2 and S3). The average inbreeding coefficient ( $F_{IS}$ ) was positive and  
284 significantly different from zero at all life stages. According to the result of INEST  
285 program, these  $F_{IS}$  values relate the presence of inbreeding, null alleles and some  
286 genotyping errors (Table 2) (see supporting information Table S4). Under *ex situ*  
287 conditions, the  $F_{IS}$  in adults (0.375) and progeny arrays (0.207) significantly differed.  
288 However, under *in situ* conditions, there was no difference, showing that the mean  $F_{IS}$  is  
289 similar in adults and offspring (Table 2).

290         When analysing both *in situ* and *ex situ* conditions, the Bayesian clustering  
291 approaches implemented in STRUCTURE estimated the number of clusters with the  
292 highest  $\Delta K$  at  $K=2$ . This result suggests that *ex situ* collection is genetically different from  
293 the natural population, since a partition of individuals from each location into 2 clusters  
294 was observed (Fig. 3).

295

## 296         **Mating system**

297

298         Both  $t_m$  and  $t_s$  were statistically nonzero under both conditions (Table 3) in all *in*  
299 *situ* and *ex situ* seedling families (see supporting information Table S5). The  $t_m-t_s$  was

300 significantly greater than zero under both *in situ* (0.171) and *ex situ* (0.070) conditions  
301 (Table 3).

302 The  $t_m-t_s$  at the family level was significantly greater than zero in eleven of the *in*  
303 *situ* seedling families, ranging from 0.036 to 0.106, and fifteen of the *ex situ* seedling  
304 families, ranging from 0.008 to 0.110 (see supporting information Table S5). Under *in situ*  
305 conditions, the M18 family had fewer cross-pollination events ( $t_m=0.436$ , a value that may  
306 be due to the isolation of the mother tree). The estimated  $t_s$  was also high, except for in  
307 M18. These results suggest that the species has a mixed-mating system (see supporting  
308 information Table S5).

309 The  $r_s$  under *in situ* and *ex situ* conditions was positive and significant, equal to  
310 0.215 and 0.119, respectively (Table 3). However, the values were low but nonzero,  
311 indicating the existence of individuals generated from self-fertilization. The multilocus  $r_p$   
312 obtained under *in situ* conditions was high ( $r_p=0.479$ ), while that for the germplasm  
313 collection was low ( $r_p=0.099$ ), suggesting that for each mother tree, the chance that the  
314 same pollen donor sired two random sibs was 47.9% and 9.9%, respectively (Table 3). The  
315  $N_{ep}$ , i.e., the average number of individuals who effectively pollinated the matrices, under  
316 *in situ* and the *ex situ* conditions was equal to 2.087 and 10.101, respectively. The  
317 proportion of full sibs was high under *in situ* conditions, equal to 0.390, and low under *ex*  
318 *situ* conditions, equal to 0.092, whereas the proportions of half sibs were 0.424 and 0.841,  
319 respectively (Table 3).

320

### 321 **Analysis of paternity and pollen dispersal**

322

323 Of the 364 seeds evaluated under *in situ* conditions, it was possible to determine the  
324 pollen donor for 117 seeds (32%). Under *ex situ* conditions, from 316 seeds, the probable

325 pollen donor was determined in 126 seeds (39%). Under both conditions, a 95%  
326 confidence level was considered. The maximum pollinator distance detected by paternity  
327 analysis under *in situ* conditions was 2.9 km (Fig. 1). Under *ex situ* conditions, a maximum  
328 pollination distance of 135.37 m was observed (Fig. 2).

329

## 330 **Discussion**

331

332 The values of genetic diversity obtained for *D. alata* support previous studies that  
333 report low levels of genetic diversity in *D. alata* (Tarazi et al. 2010; Melo et al. 2011;  
334 Collevatti et al. 2013; Guimarães et al. 2017; Tambarussi et al. 2017). Thus, our study and  
335 previous studies suggest that *D. alata* presents lower polymorphism than other native  
336 Cerrado species also evaluated using microsatellite markers (*Lychnophora ericoides*  
337 (Asteraceae), Rabelo et al. 2011; *Tabebuia roseo-alba* (Bignoniaceae), Feres et al. 2012;  
338 *Qualea grandiflora* (Vochysiaceae), Antiqueira and Kageyama 2014; *Himatanthus*  
339 *drasticus* (Apocynaceae), Baldauf et al. 2014; *Byrsonima cydoniifolia* (Malpighiaceae),  
340 Bernardes et al. 2014; *Eugenia dysenterica* (Myrtaceae), Barbosa et al. 2015; *Hancornia*  
341 *speciosa* (Apocynaceae), Costa et al. 2017).

342 The *ex situ* conditions presented greater efficiency in preserving a greater number  
343 of alleles. The difference in genetic diversity between *in situ* and *ex situ* conditions can be  
344 determine by analysing the way that genetic diversity is distributed under the different  
345 conditions. *Ex situ* conditions include alleles from different populations located in the state  
346 of Goiás (Corrêa et al. 2000), while *in situ* conditions are restricted to a local population  
347 sharing common alleles. These results confirm that the *D. alata* germplasm collection  
348 plays a role in conserving genetic diversity, as does the natural population, each in

349 proportion. This difference in genetic diversity between *in situ* and *ex situ* conditions was  
350 observed by the Bayesian assignment of individuals, at  $K=2$ , clearly demonstrating genetic  
351 differentiation between *in situ* and *ex situ* conditions.

352 The progenies analysed under *in situ* and *ex situ* conditions recovered the genetic  
353 diversity found in adults, with a significant difference between adult and juvenile  
354 individuals under *ex situ* conditions. This result shows that the mating structure most  
355 likely guarantees the conservation of genetic diversity and polymorphism between the  
356 stages of life, suggesting that genetic diversity is being maintained over generations. This  
357 factor was also observed by Rodrigues et al. (2016) when evaluating a neotropical tree  
358 species. Thus, it becomes necessary to establish conservation strategies for genetic  
359 diversity for future generations.

360 *D. alata* evaluated in the present study presented a mixed-mating system under *ex*  
361 *situ* and *in situ* conditions and was predominantly allogamous, with a  $t_m$  of 0.934 and  
362 0.815, respectively. The rate of crossing between matrices varied slightly, and the majority  
363 of matrices showed a high cross-pollination rate. The  $t_m$  and the attribution of paternity  
364 showed a prevalence of crossover with a low proportion of self-fertilization. However,  
365 some individual trees may withstand high levels of self-fertilization (e.g., M18 in the  
366 natural population), while the majority showed low self-fertilization with a  $t_m$  close to 0.99.

367 The high  $t_m$  observed for the matrices under *in situ* and *ex situ* conditions agrees  
368 with other work performed with *D. alata* (Tarazi et al. 2010; Tambarussi et al. 2017) and  
369 the results found for other fruit species of the Brazilian Cerrado, which present similar  
370 mating systems (Collevatti et al. 2010; Moraes and Sebbenn 2011; Antiqueira and  
371 Kageyama 2014; Collevatti et al. 2016; Rodrigues et al. 2016). However, the frequency of  
372 crosses between relatives under *ex situ* conditions ( $t_m-t_s=0.070$ ) was lower than that under

373 *in situ* conditions ( $t_m-t_s=0.171$ ) since the average kinship of the individuals under *ex situ*  
374 conditions was minimal. The *ex situ* conditions included plants that supposedly did not  
375 present biparental inbreeding. Thus, the kinship found under *ex situ* conditions may be  
376 related to crossing with some of their nearby daughter trees. This is because in the vicinity  
377 of the germplasm collection, there were other *D. alata* plants, including some plants from  
378 seeds collected in the germplasm collection itself. These plants were located at distances  
379 ranging from a few tens of metres up to 5 km from the germplasm collection.

380         The *ex situ* conditions showed a high number of seeds that shared pollen donors  
381 ( $r_p=0.099$ ), but this number under *in situ* conditions was higher ( $r_p=0.479$ ), indicating a  
382 low number of pollen donors per tree. These factor was confirmed by average effective  
383 number of pollen donors observed under *in situ* conditions (2.08 individuals) and *ex situ*  
384 conditions (10.10 individuals), demonstrating restricted variation in the number of pollen  
385 donors per tree. This may be related to a low movement of pollinators and pollen transport  
386 among the trees due to spatial distribution of individuals. However, under *ex situ*  
387 conditions, the number of pollen donors per tree was higher. The trees were distant from  
388 each other by approximately 6 m; thus, the proximity among trees may have influenced  
389 this result. under *in situ* conditions, the proximity among trees was also a relevant factor  
390 because although the area was larger, it presents groups of trees that are very close to each  
391 other. As a result, less cross-pollination is occurring among the more distant tree groups,  
392 contributing to a low number of pollen donors per tree under *in situ* conditions.

393         Among the factors that may have influenced the formation of these groups along  
394 the landscape under *in situ* conditions is the heterogeneous landscape due to the  
395 fragmentation and life history of *D. alata*, characterized by the formation of groups of  
396 individuals with family structure (Soares et al. 2008; Collevatti et al. 2010). Thus, these

397 factors may have influenced the low effective number of pollen donors that fertilized the  
398 trees. Our results are similar to those found in *D. alata* by Tambarussi et al. (2017), in  
399 which the effective number of pollen donors that fertilized the trees ranged from 1.5 to  
400 10.3, showing that the higher the population, the lower is the number of pollen donors per  
401 tree. In view of this, it is possible to observe that larger populations are being more  
402 affected than smaller populations by habitat fragmentation.

403         In the paternity analysis, it was possible to assign the pollen donor for less than half  
404 of the seeds under both conditions. This result may be due to missing data, since both  
405 conditions included a potential source of pollen contamination outside of the collection  
406 area. Under *ex situ* conditions, the pollen donor attribution percentage was higher (39%),  
407 suggesting that pollinators carried pollen short distances due to the high density of  
408 individuals. This hypothesis could be confirmed under *ex situ* conditions, where the most  
409 pollination events occurred at distances of less than 70 m (75%). The shortest distance for  
410 cross-pollination was 6 m between nearest neighbours, while the maximum pollen  
411 dispersal distance was 135.37 m. Other works in germplasm collections of neotropical fruit  
412 trees show that most pollination events occur at short distances, as in *Eugenia dysenterica*  
413 (Myrtaceae) (Rodrigues et al. 2016) and *Hancornia speciosa* (Apocynaceae) (Collevatti et  
414 al. 2016).

415         Under *in situ* conditions, a maximum distance of 2.9 km for gene flow by pollen  
416 was detected, which is exactly the maximum distance between two trees, demonstrating  
417 that the pollinators are reaching great distances of pollen dispersion. This result may be  
418 due to the low density of individuals and habitat fragmentation given presence of areas of  
419 pasture and agricultural fields between the natural area fragments. Tambarussi et al. (2017)  
420 evaluated two *in situ* conditions and obtained a mean pollen dispersal distance estimated

421 between 1.3 km and 6.5 km, confirming that *D. alata in situ* conditions can present great  
422 distances of dispersion of pollen. This factor was observed in a study by Balfour et al.  
423 (2015) with bees that also identified a foraging ‘hotspot’ in an agricultural field at a  
424 distance of 2–3 km. This may explain the dispersion distance observed under *in situ*  
425 conditions (2.9 km), considering that the landscape of the *in situ* conditions consisted of  
426 agricultural fields, pasture and natural reserve.

427         It is concluded that the *ex situ* germplasm collection of UFG plays a role in  
428 conserving genetic diversity, as does the natural population, each in its proportion. This is  
429 important considering that genetic diversity is a prerequisite for a species domestication  
430 and breeding (Rao and Hodgkin 2002). The *in situ* and *ex situ* conditions presented a high  
431 cross fertilization rate, confirming that *D. alata* species presented a mixed-mating system  
432 that was predominantly allogamous. However, the number of seeds that shared the same  
433 pollen donors was high, indicating a low number of pollen donors per tree. This may be  
434 related to the proximity between groups of trees, considering that there is less cross-  
435 pollination between groups of more distant trees. Another factor that may have influenced  
436 cross-pollination at long distances is the heterogeneous landscape due to habitat  
437 fragmentation. Thus, these results should be considered when defining strategies for the  
438 use and conservation of this species.

439

#### 440 **Acknowledgements**

441

442 Our research has been continuously supported by several grants and fellowships to the  
443 research network Núcleo de excelência em recursos genéticos vegetais do Cerrado  
444 (CERGEN) (PRONEX/FAPEG/CNPq, Proc. 201210267000802) and “Chamada pública

445 08/2014 Capes/FAPEG. Work by R. A. Guimarães and K. M. Corrêa was supported by a  
446 fellowship from CAPES, and work by M. P. C. Telles, L. J. Chaves and T. N. Soares was  
447 also supported by productivity grants from CNPq. Current research is developed in the  
448 context of National Institutes for Science and Technology (INCT) in Ecology, Evolution and  
449 Biodiversity Conservation, supported by MCTIC/CNPq (proc. 465610/2014-5) and FAPEG.

450

#### 451 **Conflicts of interest**

452 The authors declare no conflicts of interest.

453

#### 454 **Data archiving statement**

455 All data provided as supplementary material are available in the online version of this  
456 article.

457

#### 458 **References**

459 Almeida SP de, Proença CEB, Sano SM, Ribeiro JF (1998) Cerrado: espécies vegetais  
460 úteis.

461 Antiqueira LMOR, Kageyama PY (2015) Reproductive system and pollen flow in  
462 progenies of *Qualea grandiflora* Mart., a typical species of the Brazilian Cerrado. Rev  
463 *Árvore* 39:337–344. doi: 10.1590/0100-67622015000200013

464 Antiqueira LMOR, Kageyama PY (2014) Genetic diversity of four populations of *Qualea*  
465 *grandiflora* Mart. in fragments of the Brazilian Cerrado. *Genetica* 142:11–21. doi:  
466 10.1007/s10709-013-9750-5

467 Baldauf C, Ciampi-Guillardi M, Aguirra TJ, et al (2014) Genetic diversity, spatial genetic

468 structure and realised seed and pollen dispersal of *Himatanthus drasticus*  
469 (Apocynaceae) in the Brazilian savanna. *Conserv Genet* 15:1073–1083. doi:  
470 10.1007/s10592-014-0600-5

471 Balfour NJ, Fensome KA, Samuelson EEW, Ratnieks FLW (2015) Following the dance:  
472 Ground survey of flowers and flower-visiting insects in a summer foraging hotspot  
473 identified via honey bee waggle dance decoding. *Agric Ecosyst Environ* 213:265–  
474 271. doi: 10.1016/j.agee.2015.08.007

475 Barbosa AC de OF, Collevatti RG, Chaves LJ, et al (2015) Range-wide genetic  
476 differentiation of *Eugenia dysenterica* (Myrtaceae) populations in Brazilian Cerrado.  
477 *Biochem Syst Ecol* 59:288–296. doi: <http://dx.doi.org/10.1016/j.bse.2015.02.004>

478 Bernardes V, Anjos DE dos, Gondim SG de CA, et al (2014) Isolation and characterization  
479 of microsatellite loci in *Byrsonima cydoniifolia* (Malpighiaceae) and cross-  
480 amplification in *B. crassifolia*. *Appl Plant Sci* 2:1400016. doi: 10.3732/apps.1400016

481 Braga AC, Collevatti RG (2011) Temporal variation in pollen dispersal and breeding  
482 structure in a bee-pollinated Neotropical tree. *Heredity (Edinb)* 106:911–9. doi:  
483 10.1038/hdy.2010.134

484 Chybicki IJ, Burczyk J (2009) Simultaneous estimation of null alleles and inbreeding  
485 coefficients. *J Hered* 100:106–113. doi: 10.1093/jhered/esn088

486 Collevatti RG, Estolano R, Garcia SF, Hay JD (2010) Short-distance pollen dispersal and  
487 high self-pollination in a bat-pollinated neotropical tree. *Tree Genet Genomes* 6:555–  
488 564. doi: 10.1007/s11295-010-0271-4

489 Collevatti RG, Olivatti AM, Telles MPC, Chaves LJ (2016) Gene flow among *Hancornia*  
490 *speciosa* (Apocynaceae) varieties and hybrid fitness. *Tree Genet Genomes* 12:74. doi:  
491 10.1007/s11295-016-1031-x

492 Collevatti RG, Telles MPC, Nabout JC, et al (2013) Demographic history and the low

493 genetic diversity in *Dipteryx alata* (Fabaceae) from Brazilian Neotropical savannas.  
494 Heredity (Edinb) 111:97–105. doi: 10.1038/hdy.2013.23

495 Corrêa GDC, Rocha MR Da, Naves RV (2000) Germinação de sementes e emergência de  
496 plântulas de baru (*Dipteryx alata* Vog.) nos cerrados do Estado de Goiás. Pesqui  
497 Agropecuária Trop 30:17–23. doi: 10.5216/pat.v30i2.2580

498 Corrêa MP (1984) Dicionário de plantas úteis do Brasil e das exóticas cultivadas. In:  
499 Dicionário de plantas úteis do Brasil e das exóticas cultivadas. Imprensa Nacional  
500 Brasília.

501 Costa CF, Collevatti RG, Chaves LJ, et al (2017) Genetic diversity and fine-scale genetic  
502 structure in *Hancornia speciosa* Gomes ( Apocynaceae ). Biochem Syst Ecol J 1–5.  
503 doi: 10.1016/j.bse.2017.03.001

504 Cruz CD, Ferreira FM, Pessoni LA (2011) Biometria aplicada ao estudo da diversidade  
505 genética.

506 Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh  
507 leaf tissue. Phytochem Bull 19:11–15.

508 Ellstrand NC (2014) Is gene flow the most important evolutionary force in plants? Am J  
509 Bot 101:737–753. doi: 10.3732/ajb.1400024

510 Engels JMM, Visser L (2003) A Guide to Effective Management of Germplasm  
511 Collections. IPGRI Handbooks for Genebanks, Maccarese Rome, Italy.

512 Ennos RA (1994) Estimating the relative rates of pollen and seed migration among plant  
513 populations. Heredity (Edinb) 72:250–259. doi: 10.1038/hdy.1994.35

514 Felfili JM, Silva Júnior MC, Sevilha AC, et al (2004) Diversity, floristic and structural  
515 patterns of cerrado vegetation in Central Brazil. Plant Ecol Former 'Vegetation'  
516 175:37–46. doi: 10.1023/B:VEGE.0000048090.07022.02

517 Feres JM, Sebbenn AM, Guidugli MC, et al (2012) Mating system parameters at

518 hierarchical levels of fruits, individuals and populations in the Brazilian insect-  
519 pollinated tropical tree, *Tabebuia roseo-alba* (Bignoniaceae). *Conserv Genet* 13:393–  
520 405. doi: 10.1007/s10592-011-0292-z

521 Gepts P (2006) Plant genetic resources conservation and utilization: The accomplishments  
522 and future of a societal insurance policy. *Crop Sci* 46:2278–2292. doi:  
523 10.2135/cropsci2006.03.0169gas

524 Guimarães RA, Telles MPC, Antunes AM, et al (2017) Discovery and characterization of  
525 new microsatellite loci in *Dipteryx alata* Vogel (Fabaceae) using next-generation  
526 sequencing data. *Genet Mol Res* 16:1–6. doi: DOI  
527 <http://dx.doi.org/10.4238/gmr16029639>

528 Hoban S, Strand A (2015) *Ex situ* seed collections will benefit from considering spatial  
529 sampling design and species reproductive biology. *Biol Conserv* 187:182–191. doi:  
530 10.1016/j.biocon.2015.04.023

531 Jimenez HJ, Martins LSS, Montarroyos AV V, et al (2015) Genetic diversity of the  
532 Neotropical tree *Hancornia speciosa* Gomes in natural populations in Northeastern  
533 Brazil. *Genet Mol Res* 14:17749–17757. doi: 10.4238/2015. December.21.48

534 Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program  
535 CERVUS accommodates genotyping error increases success in paternity assignment.  
536 *Mol Ecol* 16:1099–1106. doi: 10.1111/j.1365-294X.2007.03089.x

537 Lewis PO, Zaykin D (2001) Genetic data analysis: computer program for the analysis of  
538 allelic data.

539 Manoel RO, Alves PF, Dourado CL, et al (2012) Contemporary pollen flow, mating  
540 patterns and effective population size inferred from paternity analysis in a small  
541 fragmented population of the Neotropical tree *Copaifera langsdorffii* Desf.  
542 (Leguminosae-Caesalpinioideae). *Conserv Genet* 13:613–623. doi: 10.1007/s10592-

543 011-0311-0

544 Martins PS (1987) Estrutura populacional, fluxo gênico e conservação “*in situ*” Ipef 71–78.

545 Melo DB, Diniz-Filho JAF, Oliveira G, Santana LL, Soares TN, Chaves LJ, Naves RV,  
546 Collevatti, RG, Telles MP de C (2011) Optimizing sampling efforts for *ex situ*  
547 conservation of genetic variability of *Dipteryx alata* Vogel. BMC Proc 5:P18. doi:  
548 10.1186/1753-6561-5-S7-P18

549 Moraes MLT de, Sebbenn AM (2011) Pollen dispersal between isolated trees in the  
550 Brazilian savannah: A case study of the neotropical tree *Hymenaea stigonocarpa*.  
551 Biotropica 43:192–199. doi: 10.1111/j.1744-7429.2010.00679.x

552 Oliveira MIB, Sigrist R (2008) Fenologia reprodutiva , polinização e reprodução de  
553 *Dipteryx alata* Vogel (Leguminosae-Papilionoideae) em Mato Grosso do Sul, Brasil  
554 1. Rev Bras Botânica 31:195–207.

555 Pineli LDL de O, Carvalho MV de, Aguiar L de A, Oliveira GT de, Celestino SMC,  
556 Botelho RBA, Chiarello MD (2015) Use of baru (Brazilian almond) waste from physical  
557 extraction of oil to produce flour and cookies. LWT - Food Sci Technol 60:50–55. doi:  
558 10.1016/j.lwt.2014.09.035

559 Oosterhout C Van, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO - CHECKER :  
560 software for identifying and correcting genotyping errors in microsatellite data. Mol  
561 Ecol Notes 4:535–538. doi: 10.1111/j.1471-8286.2004.00684.x

562 Rabelo SG, Teixeira CF, Telles MPC, Collevatti RG (2011) Development and  
563 characterization of microsatellite markers for *Lychnophora ericoides*, an endangered  
564 Cerrado shrub species. Conserv Genet Resour 3:741–743. doi: 10.1007/s12686-011-  
565 9447-y

566 Rajpurohit D, Jhang T (2015) *In situ* and *ex situ* conservation of plant genetic resources  
567 and traditional knowledge. In: Salgotra RK, Gupta BB (eds) Plant Genetic Resources

568 and Traditional Knowledge for Food Security. Springer Singapore, Singapore, pp  
569 137–162

570 Rao VR, Hodgkin T (2002) Genetic diversity and conservation and utilization of plant  
571 genetic resources. *Plant Cell Tissue Organ Cult* 68:1–19. doi:  
572 10.1023/A:1013359015812

573 Richardson JL, Urban MC, Bolnick DI, Skelly DK (2014) Microgeographic adaptation and  
574 the spatial scale of evolution. *Trends Ecol Evol* 29:165–176.  
575 doi:10.1016/j.tree.2014.01.002

576 Ritland K (2002) Extensions of models for the estimation of mating systems using n  
577 independent loci. *Heredity* 88:221–228. doi:10.1038/sj.hdy.6800029

578 Rodrigues EB, Collevatti RG, Chaves LJ, et al (2016) Mating system and pollen dispersal  
579 in *Eugenia dysenterica* (Myrtaceae) germplasm collection: tools for conservation and  
580 domestication. *Genetica* 144:139–146. doi: 10.1007/s10709-016-9884-3

581 Sano SM, Ribeiro JF, Brito MA (2004) Baru: biologia e uso. Embrapa cerrados -Brasilia  
582 DF

583 Santos-Filho PS (1995) Fragmentação de habitats: implicações para conservação *in situ*.  
584 *Oecologia Bras.* 1:365–393.

585 Santos AM, Rosado SC da S, Oliveira AN (2014) Estimation of genetic parameters and  
586 verification of early selection efficiency in baru (*Dipteryx alata*). *Crop Breed Appl*  
587 *Biotechnol* 14:238–243. doi: 10.1590/1984-70332014v14n4a37

588 Scariot AO, Servilha AC (2007) Conservação *in situ* de recursos genéticos vegetais. In:  
589 Nass LL (ed) Recursos genéticos vegetais, 1<sup>a</sup>. Nass, Luciano Lourenço, Brasilia, DF,  
590 pp 473–502

591 Sebbenn AM (2002) Número de árvores matrizes e conceitos genéticos na coleta de  
592 sementes para reflorestamentos com espécies nativas. *Rev. do Inst. Flor*, 14:115–132.

593 Slatkin M (1987) Gene flow and the geographic structure of natural populations. Sci  
594 236:787.

595 Soares TN, Chaves LJ, Telles M de CP, et al (2008) Distribuição espacial da variabilidade  
596 genética intrapopulacional de *Dipteryx alata*. *Pesqui Agropecuária Bras* 43:1151–  
597 1158.

598 Soares TN, Diniz-Filho JAF, Nabout JC, et al (2015) Patterns of genetic variability in  
599 central and peripheral populations of *Dipteryx alata* (Fabaceae) in the Brazilian  
600 Cerrado. *Plant Syst Evol* 301:1315–1324. doi: 10.1007/s00606-014-1155-0

601 Soares TN, Melo DB, Resende LV, et al (2012) Development of microsatellite markers for  
602 the Neotropical tree species *Dipteryx alata* (Fabaceae). *Am J Bot* 99:e72-3. doi:  
603 10.3732/ajb.1100377

604 Sokal RR, Rohlf FJ (1995) *Biometry: the principles and practice of statistics in biological*  
605 *research*, 3rd edn.

606 Tambarussi E V, Sebbenn AM, Vencovsky R, et al (2017) *Dipteryx alata* Vogel  
607 (Fabaceae), a Neotropical tree with high levels of selfing: implications for  
608 conservation and breeding programs. *Ann For Res*. doi: 10.15287/afr.2017.842

609 Tarazi R, Moreno MA, Gandara FB, et al (2010) High levels of genetic differentiation and  
610 selfing in the Brazilian cerrado fruit tree *Dipteryx alata* Vog. (Fabaceae). *Genet Mol*  
611 *Biol* 33:78–85. doi: 10.1590/S1415-47572010005000007

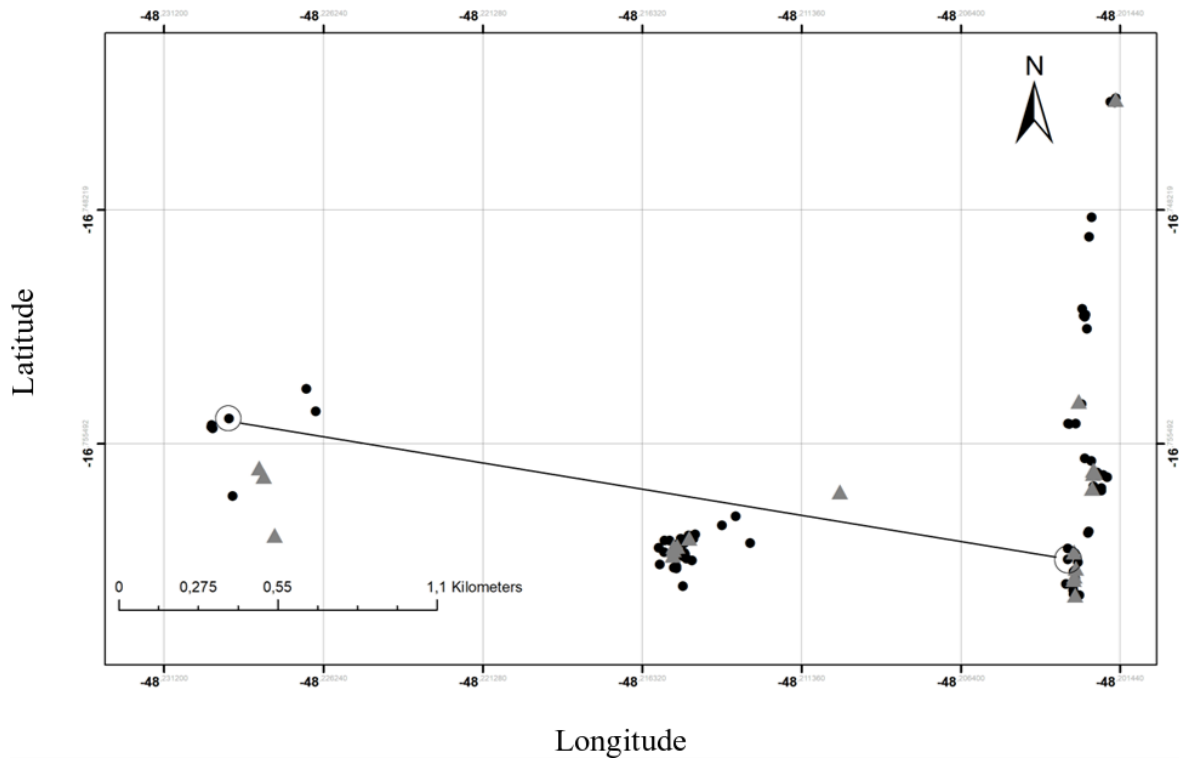
612 Vieira RF, Ferreira T da SAC, Silva DB, et al (2006) *Frutas nativas da região centro-oeste*  
613 *do Brasil*. Embrapa recursos genéticos e biotecnologia Brasília, DF

614 Wagner HW, Sefc KM (1999) *IDENTITY 1.0* Centre for Applied Genetics.

615 Williams CG (2017) How meso-scale pollen dispersal and its gene flow shape gene  
616 conservation decisions. *New For* 48:217–224. doi: 10.1007/s11056-017-9574-8  
617

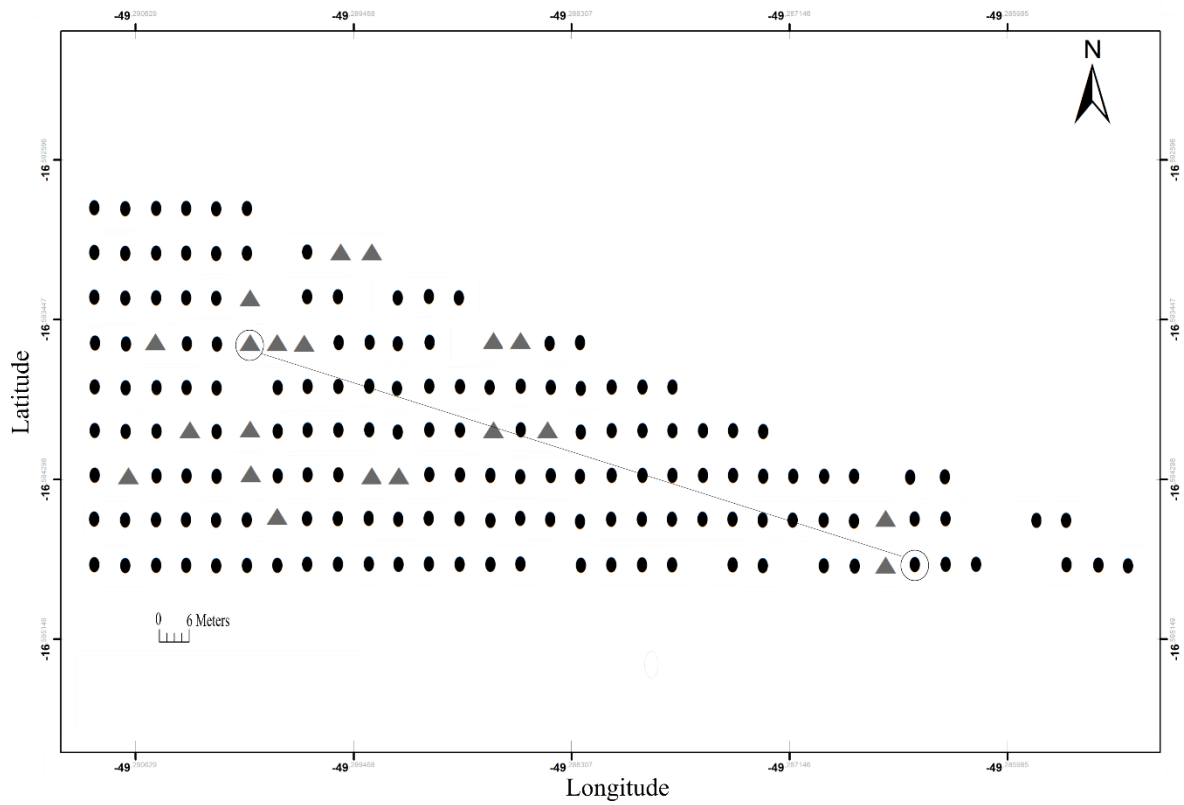
618 **Figure legends**

619



620

621 **Fig. 1** Figure representing sampling under *in situ* conditions. The triangles represent the  
622 matrices, while the circles represent the other adults. Maximum pollination distance  
623 detected through paternity analysis in *D. alata* from the natural population of Orizona,  
624 Goiás, Brazil.



625

626 **Fig. 2** Figure representing sampling under *ex situ* conditions. The triangles represent the  
 627 matrices, while the circles represent the other plants. Maximum distance of pollination  
 628 detected through paternity analysis in *D. alata* plants from the germplasm collection -  
 629 School of Agronomy UFG.

630

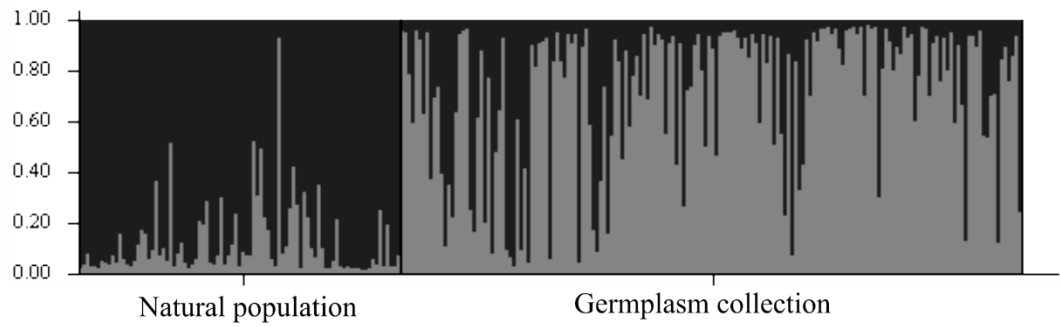
631

632

633

634

635



636

637 **Fig. 3** Result of genetic assignment of individuals to the natural population (A) and  
638 germplasm collection (B) in two groups (K=2) carried out in the STRUCTURE program.

639 Each column represents an individual, and the groups are different shades.

640

641

642

643

644

645

646

647

648

649

650

651

652

653

654

655

656 **Tables**

657 **Table 1** Characterization of ten microsatellite loci of *D. alata*, based on all individuals  
 658 sampled in two conditions, *in situ* and *ex situ*, of central Brazil.

Locus	Natural population (N=515)					Germplasm collection (N=48)				
	A	$H_e$	$H_o$	Q	I	A	$H_e$	$H_o$	Q	I
DaE34	8	0.364	0.290	0.284	0.290	8	0.695	0.472	0.531	0.101
DaE63	3	0.606	0.401	0.337	0.218	3	0.612	0.579	0.366	0.196
Dal15	3	0.653	0.345	0.371	0.187	3	0.357	0.329	0.203	0.416
Dal20	5	0.493	0.405	0.240	0.319	6	0.478	0.381	0.294	0.285
Dal23	8	0.679	0.499	0.448	0.145	10	0.783	0.666	0.615	0.066
Dal25	3	0.511	0.439	0.264	0.299	4	0.667	0.384	0.396	0.170
DaE41	8	0.732	0.667	0.514	0.110	9	0.655	0.545	0.442	0.153
Dal21	3	0.208	0.134	0.163	0.343	2	0.383	0.230	0.198	0.391
Dal18	3	0.462	0.399	0.222	0.155	4	0.328	0.274	0.240	0.363
Da80	8	0.620	0.492	0.429	0.290	13	0.698	0.553	0.548	0.093
Average	5.2	0.532	0.407	0.983	$4.69 \times 10^{-7}$	6.2	0.566	0.441	0.994	$5.74 \times 10^{-8}$

659 A number of alleles,  $H_e$  expected heterozygosity based on the Hardy-Weinberg  
 660 equilibrium,  $H_o$  observed heterozygosity, Q probability of exclusion paternity, I probability  
 661 of genetic identity.

662

663

664

665

666 **Table 2** Genetic diversity and inbreeding in adults and progeny arrays from *D. alata* in two  
 667 conditions, one *in situ* and the other *ex situ*, in central Brazil.

Natural population							
Group	<i>N</i>	<i>A</i>	<i>H<sub>e</sub></i>	<i>H<sub>o</sub></i>	<i>F<sub>IS</sub></i>	HPDI< <i>F<sub>IS</sub></i> (INEST)<HPDh	M
Adults	89	3.6 <sup>a</sup>	0.534 <sup>a</sup>	0.424 <sup>a</sup>	0.213 <sup>*a</sup>	0.1273<0.1778<0.2300	n/f/b
Progeny	364	4.8 <sup>b</sup>	0.519 <sup>a</sup>	0.519 <sup>b</sup>	0.330 <sup>*a</sup>	0.0193<0.0714<0.1234	n/f/b
Germplasm collection							
Group	<i>N</i>	<i>A</i>	<i>H<sub>e</sub></i>	<i>H<sub>o</sub></i>	<i>F<sub>IS</sub></i>	HPDI< <i>F<sub>IS</sub></i> (INEST)<HPDh	M
Adults	173	6.8 <sup>a</sup>	0.602 <sup>a</sup>	0.377 <sup>a</sup>	0.374 <sup>*a</sup>	0.0575<0.1153<0.1845	n/f/b
Progeny	316	6.3 <sup>a</sup>	0.572 <sup>a</sup>	0.454 <sup>b</sup>	0.198 <sup>*b</sup>	0.0149<0.0372<0.0589	n/f/b

668 A number of alleles, *H<sub>e</sub>* expected heterozygosity based on the Hardy-Weinberg  
 669 equilibrium, *H<sub>o</sub>* observed heterozygosity and *F<sub>IS</sub>* coefficient of inbreeding. Values followed  
 670 by the same letter in the same column did not differ,  $p > 0.05$ . \*Significant *F<sub>IS</sub>*-value  
 671 considering the Bonferroni correction. *F<sub>IS</sub>* (INEST) is the mean inbreeding coefficient taking  
 672 into account null alleles, and genotyping errors. M is the INEST model which best fit the  
 673 data: n=null alleles, f=inbreeding, b=genotyping error. HPDI / HPDh is the 95% highest  
 674 posterior density interval.

675

676

677

678

679

680

681

682

683

684 **Table 3** Estimates of parameters of the inbreeding reproductive system of *D. alata* in two  
 685 conditions, one *ex situ* and the other *in situ*, in central Brazil.

Parameters	Natural population	Germplasm collection
Total number of matrices ( $M$ )	20	20
Total number of progeny seedlings ( $n$ )	364	316
Fixation index of seed trees ( $F_m$ )	0.017	0.146
Multilocus outcrossing rate ( $t_m$ )	0.815	0.934
Single-locus outcrossing rate ( $t_s$ )	0.644	0.864
Biparental inbreeding rate ( $t_m - t_s$ )	0.171	0.070
Correlation of selfing ( $r_s$ )	0.215	0.119
Correlation of paternity ( $r_p$ )	0.479	0.099
Mean number of pollen donor trees ( $N_{ep}$ )	2.087	10.101
Proportion of full-sibs ( $P_{IC}$ )	0.390 (39%)	0.092 (9%)
Proportion of half-sibs ( $P_{MI}$ )	0.424 (42%)	0.841 (84%)
Proportion of selfing-sibs ( $P_{IA}$ )	0.185 (19%)	0.066 (7%)

686

687 **Mating system and pollen dispersal in *Dipteryx alata* Vogel (Fabaceae): comparing *in***  
688 ***situ* and *ex situ* conditions**

689

690 Rejane Araújo Guimarães<sup>1,2</sup>, Kássia Marques Corrêa<sup>1,2</sup>, Lázaro José Chaves<sup>4</sup>, Ronaldo  
691 Veloso Naves<sup>4</sup>, Mariana Pires de Campos Telles<sup>1,3</sup>, Thannya Nascimento Soares<sup>1,2</sup>.

692

693 Autor correspondente: R. A. Guimarães/ T. N. Soares

694 E-mail: [rejanearaujog@gmail.com](mailto:rejanearaujog@gmail.com) / [tnsoares@gmail.com](mailto:tnsoares@gmail.com)

695

696

697

698 **Supporting Information – Tables**

699

700 **Table S1** Multiplex set used in the genetic analyses of *D. alata* individuals in the present study.

701

<i>Primers</i>	Allele size range (bp) <sup>b</sup>	Fluorescence	Reference
Multiplex 1			
Dal15	198-212	6'FAM	Guimarães et al. 2017
Dal23	132-148	NED	Guimarães et al. 2017
DaE34	100-118	HEX	Soares et al. 2012
Dal25	152-162	VIC	Guimarães et al. 2017
DaE63	208-216	HEX	Soares et al. 2012
Dal20	184-210	PET	Guimarães et al. 2017
Multiplex 2			
DaE41	110-142	NED	Soares et al., 2012
Dal21	156-160	6'FAM	Guimarães et al. 2017
Multiplex 3			
Dal18	186-200	VIC	Guimarães et al. 2017
Da80	144-166	PET	Guimarães et al. 2017

702

703

704

705

706 **Table S2** Characterization of seven microsatellite loci of *D. alata* and comparison of genetic diversity and  
 707 polymorphism in three life stages of one natural population in central Brazil.

Loci	Adults (N = 89)				Juveniles (N = 62)				Seed (N = 364)			
	A	$H_E$	$H_O$	$F_{IS}$	A	$H_E$	$H_O$	$F_{IS}$	A	$H_E$	$H_O$	$F_{IS}$
DaE34	4	0.384	0.321	0.167	4	0.238	0.190	0.204	8	0.377	0.301	0.201
DaE63	3	0.591	0.483	0.183	3	0.618	0.290	0.532	3	0.571	0.400	0.300
Dal15	3	0.569	0.352	0.382	2	0.415	0.194	0.536	3	0.647	0.369	0.430
Dal20	2	0.500	0.466	0.068	3	0.522	0.484	0.074	4	0.484	0.376	0.224
Dal23	5	0.672	0.483	0.283	5	0.744	0.629	0.156	8	0.657	0.480	0.270
Dal25	3	0.541	0.303	0.440	3	0.366	0.194	0.473	3	0.517	0.515	0.004
DaE41	7	0.753	0.693	0.080	7	0.686	0.435	0.367	7	0.728	0.700	0.039
Dal21	2	0.245	0.148	0.399	2	0.338	0.230	0.323	3	0.170	0.112	0.337
Dal18	2	0.411	0.416	-0.011	2	0.217	0.115	0.474	3	0.490	0.445	0.091
Da80	5	0.683	0.584	0.146	6	0.694	0.583	0.161	6	0.554	0.454	0.181
Global	3,6	0.535	0.425	0.207*	3.7	0.484	0.334	0.311*	4.8	0.519	0.415	0.201*

708 \* Significant *f*-value considering the correction of Bonferroni,  $p \leq 0.00167$

709 A number of alleles,  $H_e$  expected heterozygosity based on the Hardy-Weinberg equilibrium,  $H_o$  observed  
 710 heterozygosity,  $F_{IS}$  coefficient of inbreeding.

711  
 712  
 713  
 714  
 715  
 716  
 717  
 718  
 719  
 720  
 721  
 722  
 723  
 724  
 725  
 726  
 727  
 728  
 729  
 730

731 **Table S3** Characterization of seven microsatellite loci of *D. alata* and comparison of genetic diversity and  
 732 polymorphism in two stages of the germplasm collection of central Brazil.

Locus	Adults ( <i>N</i> = 173)				Seed ( <i>N</i> = 316)			
	<i>A</i>	<i>H<sub>e</sub></i>	<i>H<sub>O</sub></i>	<i>F<sub>IS</sub></i>	<i>A</i>	<i>H<sub>e</sub></i>	<i>H<sub>O</sub></i>	<i>F<sub>IS</sub></i>
DaE34	9	0.765	0.360	0.530	8	0.724	0.462	0.363
DaE63	4	0.663	0.488	0.264	4	0.607	0.601	0.010
Dal15	4	0.371	0.285	0.232	4	0.382	0.345	0.097
Dal20	7	0.582	0.360	0.381	7	0.440	0.380	0.138
Dal23	10	0.780	0.483	0.382	10	0.811	0.703	0.134
Dal25	5	0.635	0.273	0.570	4	0.677	0.430	0.364
Dal21	3	0.346	0.145	0.581	3	0.466	0.263	0.437
DaE41	10	0.661	0.413	0.376	8	0.683	0.585	0.143
Dal18	5	0.433	0.342	0.212	4	0.262	0.238	0.091
Da80	11	0.789	0.621	0.214	11	0.674	0.536	0.205
Global	6.8	0.603	0.377	0.375*	6.3	0.573	0.454	0.207*

\* Significant *F<sub>IS</sub>*-value considering the Bonferroni correction,  $p \leq 0.00227$

733  
 734  
 735  
 736  
 737  
 738  
 739  
 740  
 741  
 742  
 743  
 744  
 745  
 746  
 747  
 748  
 749  
 750  
 751  
 752  
 753  
 754

755 Table S4 Analysis of INEST model their effects in inbreeding estimation using INEST program (Chybicki  
 756 and Burczyk 2009).  
 757

	Model <sub>(INEST)</sub>	FIS <sub>(INEST)</sub>	HPDI - HPDh	DIC
Natural population Adults	nf	0.1555	0.0861 – 0.231	2817
	fb	0.2151	0.1732 – 0.2300	2806
	<b>nfb</b>	<b>0.1778</b>	<b>0.1273 – 0.2300</b>	<b>2797</b>
	nb	0	0	2828
	n	0	0	2845
	b	0	0	29901
	f	-	-	-
Natural population Progeny	nf	0,0381	0.0138 – 0.0752	11780
	fb	0.2084	0.1732 – 0.2300	11766
	<b>nfb</b>	<b>0.0714</b>	<b>0.0193 – 0.1234</b>	<b>11700</b>
	nb	0	0	11711
	n	0	0	11790
	b	0	0	12010
	f	-	-	-
Germplasm collection Adults	nf	0.087	0.0385 – 0.1438	6418
	fb	0.3408	0.3102 – 0.3725	6364
	<b>nfb</b>	<b>0.1153</b>	<b>0.0575 – 0.1845</b>	<b>6318</b>
	nb	<b>0</b>	<b>0</b>	6369
	n	0	0	6454
	b	0	0	6723
	f	-	-	-
Germplasm collection Progeny	nf	0.0437	0.0202 – 0.0657	11879
	fb	0.1256	0.1064 – 0.1452	11818
	<b>nfb</b>	<b>0.0372</b>	<b>0.0149 – 0.0589</b>	<b>11740</b>
	nb	0	0	11759
	n	0	0	11967
	b	0	0	11937
	f	-	-	-

758 FIS <sub>(INEST)</sub> is the mean inbreeding coefficient taking into account null alleles, and genotyping errors. The  
 759 INEST model was computes for (n=null alleles, f=inbreeding, b=genotyping error) and for a combination of  
 760 these. HPDI / HPDh is the 95% highest posterior density interval. The DIC is deviance information criterion  
 761 for each model, the lowest DIC is corresponding with the model best fitting to data. Values with a dash (-)  
 762 were not computed by INEST due to missing data. Best INEST model is highlighted in bold.

763  
 764  
 765  
 766  
 767  
 768  
 769  
 770

771 **Table S5** Comparison of mating system parameters among the 20 mother trees from *D. alata* in two  
772 conditions: one *in situ* and the other *ex situ* in central Brazil.  
773

Matrices	Natural population				Germplasm collection			
	N	$T_m$	$T_s$	$T_m - T_s$	N	$T_m$	$T_s$	$T_m - T_s$
M1	21	0.831	0.861	-0.030	17	0.884	0.773	0.110
M2	15	0.989	0.883	0.106	13	0.990	0.898	0.092
M3	5	0.944	0.908	0.036	17	0.775	0.886	-0.111
M4	11	0.980	0.889	0.091	19	0.858	0.864	-0.006
M5	7	0.882	0.891	-0.009	15	0.905	0.897	0.008
M6	15	0.912	0.851	0.061	17	0.995	0.942	0.053
M7	26	0.866	0.887	-0.021	19	0.997	0.912	0.085
M8	25	0.996	0.898	0.098	7	0.971	0.885	0.086
M9	14	0.987	0.918	0.069	17	0.990	0.915	0.076
M10	6	0.958	0.891	0.067	23	0.987	0.914	0.072
M11	23	0.797	0.821	-0.024	11	0.986	0.926	0.059
M12	22	0.996	0.909	0.086	10	0.983	0.900	0.083
M13	21	0.594	0.776	-0.182	19	0.996	0.863	0.133
M14	21	0.697	0.813	-0.115	17	0.722	0.813	-0.091
M15	20	0.987	0.912	0.076	21	0.986	0.923	0.063
M16	23	0.740	0.847	-0.107	16	0.847	0.886	-0.038
M17	23	0.680	0.773	-0.094	16	0.992	0.887	0.105
M18	24	0.436	0.662	-0.227	16	0.993	0.884	0.109
M19	24	0.752	0.708	0.043	12	0.768	0.856	-0.088
M20	18	0.845	0.760	0.085	14	0.921	0.904	0.017

774  $N$  number of seeds analyzed,  $T_m$  multilocus outcrossing rate,  $T_s$  single-locus outcrossing rate,  $T_m - T_s$   
775 biparental inbreeding.

776

## CAPÍTULO 3

### **EFEITO DA PAISAGEM NA ESTRUTURA GENÉTICA ESPACIAL LOCAL EM INDIVÍDUOS JUVENIS E ADULTOS DE *Dipteryx alata* Vogel (FABACEAE)<sup>1</sup>**

Rejane Araújo Guimarães<sup>1,2</sup>, Kelly da Silva e Souza<sup>4,5</sup>, Mariana Pires de Campos  
Telles<sup>1,3</sup>, Thannya Nascimento Soares<sup>1,2</sup>

---

<sup>1</sup>Capítulo elaborado conforme as normas do periódico científico *Landscape Ecology*;

<sup>2</sup>Laboratório de Genética & Biodiversidade, Instituto de Ciências Biológicas, Universidade Federal de Goiás, Goiânia, Goiás, Brasil.

<sup>3</sup>Programa de pós-graduação e Genética e Melhoramento de Plantas - Escola de Agronomia – UFG, Goiás, Brasil.

<sup>4</sup>Escola de Ciências Agrárias e Biológicas, Pontifícia Universidade Católica de Goiás, Goiânia, Goiás, Brasil.

<sup>5</sup>Programa de Pós-graduação em Genética e Biologia Molecular, Instituto de Ciências Biológicas, Universidade Federal de Goiás – UFG, Goiânia, GO, Brasil.

1 **Efeito da paisagem na estrutura genética espacial local em indivíduos**  
2 **juvenis e adultos de *Dipteryx alata* Vogel (Fabaceae)**

3  
4 Rejane Araújo Guimarães<sup>1,2</sup>, Kelly da Silva e Souza<sup>4</sup>, Mariana Pires de Campos  
5 Telles<sup>1,3</sup>, Thannya Nascimento Soares<sup>1,2</sup>.

6  
7 <sup>1</sup>Laboratório de Genética & Biodiversidade, Instituto de Ciências Biológicas,  
8 Universidade Federal de Goiás, Goiânia, Goiás, Brasil.

9 <sup>2</sup>Programa de pós-graduação e Genética e Melhoramento de Plantas - Escola de  
10 Agronomia – UFG, Goiás, Brasil.

11 <sup>3</sup>Escola de Ciências Agrárias e Biológicas, Pontifícia Universidade Católica de Goiás,  
12 Goiânia, Goiás, Brasil

13 <sup>4</sup>Programa de Pós-graduação em Genética e Biologia Molecular, Instituto de Ciências  
14 Biológicas, Universidade Federal de Goiás – UFG, Goiânia, GO, Brasil.

15 ORCID: [orcid.org/0000-0003-0186-6589](https://orcid.org/0000-0003-0186-6589)

16 Autor correspondente: R. A. Guimarães/ T. N. Soares

17 E-mail: [rejanearaujog@hotmail.com](mailto:rejanearaujog@hotmail.com) / [tsoares@gmail.com](mailto:tsoares@gmail.com)

18  
19 **Resumo**

20  
21 *Contexto.* A estrutura genética espacial dentro de populações de plantas é influenciada  
22 principalmente pela dispersão de pólen e sementes. No entanto, a fragmentação do  
23 habitat pode afetar os processos de dispersão, impactando a estrutura genética espacial.

24 *Objetivos.* Testar se a estrutura genética espacial em escala local é explicada pela  
25 distância geográfica ou se existe interferência da composição da paisagem neste  
26 processo, em indivíduos adultos e juvenis de uma população de *D. alata*.

27 *Métodos.* Foram utilizados 10 locos microssatélites para avaliar 89 indivíduos adultos e  
28 62 juvenis de uma população de *D. alata* localizada em Orizona-GO, Brasil. O padrão  
29 da estrutura genética espacial foi avaliado a partir de coeficientes de parentesco de  
30 Nason ( $F_{ij}$ ) entre os pares de indivíduos, utilizando correlação de Mantel,  
31 autocorrelogramas e com base na estimativa da força da estrutura genética  $S_p$ . A análise  
32 de agrupamento entre indivíduos foi testada por métodos de agrupamento baseados em  
33 estatística bayesiana. A influência da paisagem na estrutura genética local foi avaliada  
34 com base na modelagem orientada por padrão (POM - *pattern-oriented modeling*) e na  
35 teoria de circuitos.

36 *Resultados.* Foi detectada moderada estrutura genética espacial em adultos e juvenis,  
37 que pode ser explicada pelos modos de dispersão de pólen e semente da espécie, que  
38 levam à formação de grupos de indivíduos com estruturação familiar. As análises de  
39 agrupamento bayesiano mostraram a formação de dois grupos, tanto em adultos quanto  
40 em juvenis. As composições da paisagem apresentaram maior influência na estrutura  
41 genética dos indivíduos juvenis.

42 *Conclusões.* Os fatores de dispersão restrita de sementes, que se dá principalmente por  
43 barocoria, e o histórico de perturbação antrópica dessa população de *D. alata*  
44 contribuem para o padrão de estrutura genética local. De modo geral, existe uma  
45 moderada estrutura genética espacial, porém a composição da paisagem explica melhor  
46 o padrão de variação genética, tanto para juvenis quanto para adultos. Além disso, foi  
47 detectada uma maior interferência da paisagem na variação genética dos indivíduos  
48 juvenis, o que tem implicações importantes para a conservação da espécie.

49

50 **Palavras-chave:** Autocorrelação espacial, *Dipteryx alata*, microssatélites, parentesco,  
51 resistência da paisagem.

52

## 53 **Introdução**

54

55 O conhecimento da estrutura genética espacial (*Spatial Genetic Structure* - SGS)  
56 é importante para avaliar os impactos da fragmentação nas populações naturais  
57 (Vekemans e Hardy 2004; Collevatti et al. 2010; Collevatti et al. 2014). A análise de  
58 SGS permite avaliar se há distribuição não aleatória de genótipos, resultante da ação de  
59 diferentes processos microevolutivos (Epperson 2003). A SGS dentro das populações  
60 pode ser consequência de vários fatores tais como: dispersão restrita de pólen e semente,  
61 da endogamia, das pressões de seleção, da deriva genética local e da história de  
62 colonização e sobrevivência diferencial de indivíduos, que podem favorecer um mesmo  
63 ou diferentes genótipos (Vekemans e Hardy 2004).

64 Vários estudos têm evidenciado SGS dentro de populações de espécies arbóreas  
65 (Collevatti et al. 2010; Barluenga et al. 2011; Yao et al. 2011; Costa et al. 2017),  
66 enquanto em outros estudos se observa a sua ausência de SGS (Doligez e Joly 1997; Ng  
67 et al. 2006; Alcalá et al. 2015). A existência ou não da SGS pode ser explicada por  
68 diferentes fatores ecológicos e genéticos atuando dentro das populações naturais  
69 (Vekemans e Hardy 2004). Por exemplo, espécies cuja dispersão de sementes ocorre por  
70 animais podem apresentar diferentes perfis de variação genética espacial, dependendo  
71 do comportamento do dispersor (Jordano et al. 2007; Barluenga et al. 2011; Medina-  
72 Macedo et al. 2015).

73 A SGS também pode ser afetada pela fragmentação das populações naturais  
74 dentro do seu habitat (Barluenga et al. 2011; Collevatti et al. 2014; Costa et al. 2017).  
75 Estudos recentes têm relacionado a SGS com características da paisagem, incluindo  
76 características naturais e modificações induzidas pelo homem (Telles et al. 2014),  
77 conforme proposto pela linha de pesquisa em genética da paisagem (Manel et al. 2003;  
78 Manel e Holderegger 2013).

79 Uma das formas de avaliar a influência da paisagem na estrutura genética  
80 espacial é através do conceito de “Isolamento por resistência”, baseado na teoria de  
81 circuitos elétricos (McRae 2006; McRae e Beier 2007). Nesta abordagem são  
82 desenvolvidos modelos de mapas de superfície de resistência paisagística com valores  
83 de “custos” que representam a resistência ao deslocamento do polinizador ou dispersor  
84 através das diferentes composições de paisagem (Milanesi et al. 2017).

85 Um dos principais desafios para os estudos de genética da paisagem é a  
86 atribuição dos valores de custos de resistência às características da paisagem (Spear et  
87 al. 2010). A modelagem orientada por padrão (POM - *pattern-oriented modeling*) pode  
88 ser uma alternativa para esta atribuição de custos de resistência (Grimm et al. 1996;  
89 Grimm et al. 2005; Diniz-filho et al. 2014). Na POM são utilizados procedimentos  
90 computacionais para se encontrar conjuntos de parâmetros que geram modelos que  
91 melhor explica o padrão observado. Esta estratégia melhora significativamente a  
92 qualidade das previsões do efeito das composições da paisagem sobre a SGS (Wiegand  
93 et al. 2003; Grimm et al. 2005; Diniz-filho et al. 2014).

94 A maioria dos estudos de genética da paisagem tem sido desenvolvida em escala  
95 local, utilizando populações como as unidades amostrais (Manel et al. 2003; Manel et  
96 al. 2010; Balkenhol et al. 2015). Porém nos últimos anos, a genética da paisagem tem  
97 explorado a influência da paisagem em escala local, em que os indivíduos são as

98 unidades amostrais (Cushman e Landguth 2010a; Segelbacher et al. 2010; Cushman e  
99 Landguth 2010b; Kierepka e Latch 2015). No entanto, esses estudos em escala local em  
100 plantas ainda são raros (Newman et al. 2013). Além disto, Anderson et al. (2010) e  
101 Cushman e Landguth (2010b) destacam a necessidade de estudos que envolvam  
102 múltiplas escalas espaciais e temporais para melhorar as contribuições práticas para a  
103 conservação das espécies.

104 *Dipteryx alata* Vogel (Leguminosae) é uma árvore Neotropical nativa do  
105 Cerrado. A espécie é hermafrodita e seus principais polinizadores são abelhas e suas  
106 sementes são dispersas por barocoria e zoocoria, sendo realizada principalmente por  
107 morcegos e macacos (Almeida et al. 1998; Sano et al. 2004; Oliveira e Sigrist 2008).  
108 Análises anteriores com *D. alata* revelaram que a variabilidade genética das populações  
109 da espécie é espacialmente estruturada e pode estar relacionada às condições climáticas  
110 e à fragmentação do habitat (Collevatti et al., 2010; Telles et al., 2014; Soares et al.,  
111 2015). Também foi detectada estrutura genética espacial em escala local, utilizando  
112 marcadores RAPD (Soares et al. 2008) e microsatélites (Collevatti et al. 2010). No  
113 entanto, estes estudos não avaliaram a influência da paisagem na SGS local. Desta  
114 forma, *D. alata* é considerada boa opção de espécie-modelo para o desenvolvimento do  
115 presente estudo, que possui o objetivo de verificar se o padrão de distribuição da  
116 variação genética intrapopulacional é explicado pela distância geográfica ou se a  
117 composição da paisagem pode estar interferindo neste padrão, tanto em indivíduos  
118 adultos e juvenis de uma população da espécie arbórea *D. alata*.

119

## 120 **Material e métodos**

121

### 122 **Área de estudo e amostragem**

123

124           Foram amostrados 151 indivíduos, sendo 89 adultos e 62 juvenis, em uma  
125 população de *D. alata* localizada em Orizona, Goiás, no Centro-Oeste do Brasil (16° 45'  
126 52,0"S e 48° 12'17,7" W, a 881 m de altitude) (**Fig. 1**). Esta população está situada em  
127 uma propriedade privada, com área de cerca de 148 ha e com paisagem heterogênea,  
128 composta de estradas, pastagens, agricultura e remanescentes florestais. Todos os  
129 indivíduos foram mapeados usando um receptor GPS (sistema de posicionamento  
130 global) portátil. As análises estatísticas foram conduzidas de forma separada para os  
131 indivíduos adultos e juvenis.

132

### 133 **Análises de Microsatélites**

134

135           O DNA genômico total foi extraído do tecido foliar dos indivíduos seguindo o  
136 procedimento padrão de CTAB 2% (Doyle e Doyle 1987). As reações em cadeia da  
137 polimerase (PCR) foram realizadas usando dez locos microsatélites (Soares et al. 2012;  
138 Guimarães et al. 2017) e seguiram as condições de amplificação propostas por  
139 Guimarães et al. (2017). A detecção dos fragmentos amplificados foi realizada por um  
140 sistema semi-automatizado de eletroforese capilar ABI PRISM® 3500 Genetic  
141 Analyzer (Applied Biosystems, CA) e um padrão de tamanho GeneScan LIZ600TM  
142 (AppliedBiosystems). Os indivíduos foram genotipados utilizando o software  
143 GeneMapper® v5.0 (Applied Biosystems, CA). A confirmação do perfil molecular foi  
144 realizada pela “escada alélica”, com a realização de uma nova eletroforese capilar, com  
145 um conjunto de indivíduos que continha todos os alelos de um dado loco.

146

147           Os locos foram testados quanto ao desequilíbrio de ligação e desvios das  
proporções de Hardy-Weinberg no software GDA versão 1.1 (Lewis e Zaykin 2002).

148 Além disso, foi testada a presença de alelos nulos na população através do software  
149 Microchecker (Oosterhout et al. 2004).

150

### 151 **Estrutura genética espacial**

152

153 As análises de estrutura genética espacial foram realizadas a partir de matrizes  
154 de coeficientes de parentesco de Nason ( $F_{ij}$ ) (Loiselle et al. 1995). Os coeficientes de  $F_{ij}$   
155 entre os pares de indivíduos foram gerados utilizando o software SPAGeDI (Hardy e  
156 Vekemans 2002). Esse estimador é adequado para avaliar grupos de indivíduos que  
157 possam apresentar endogamia biparental (Hardy e Vekemans 2015).

158 As matrizes de distâncias genéticas com base em coeficientes de parentesco  
159 foram obtidas a partir de [1- (matriz de  $F_{ij}$  par-a-par)]. As coordenadas geográficas dos  
160 indivíduos adultos e juvenis foram utilizadas para se obter as matrizes de distância  
161 geográfica par a par. O teste de Mantel foi realizado entre as matrizes de distâncias  
162 genética e geográfica dos adultos e juvenis.

163 A avaliação da autocorrelação espacial foi realizada seguindo o procedimento  
164 descrito por Vekemans e Hardy (2004) utilizando o software SPAGeDI (Hardy e  
165 Vekemans 2002). Foram obtidas as médias dos coeficientes de  $F_{ij}$  dentro de onze classes  
166 de distância geográfica com mesma dimensão.

167 Para se testar a estrutura do parentesco, foi realizada uma regressão linear dos  
168 coeficientes de  $F_{ij}$  médio de cada classe em função da distância geográfica entre os  
169 indivíduos (adultos e juvenis). A significância do parentesco médio em todas as classes  
170 de distância e da regressão foi verificada a partir de um teste com 10.000 permutações.  
171 Os intervalos de confiança dos coeficientes de  $F_{ij}$  foram estimados usando *jackknife*  
172 sobre os locos utilizando o software SPAGeDI (Hardy e Vekemans 2002). Na

173 construção do autocorrelogramas foram utilizadas as médias dos coeficientes de  $F_{ij}$ ,  
174 onze classes de distâncias geográficas e os intervalos de confiança.

175 A intensidade da estrutura genética espacial foi quantificada pelo parâmetro  
176 genético  $S_p$ , desenvolvido por Vekemans e Hardy (2004) e calculado de acordo com a  
177 fórmula:  $S_p = -B_{\log} / (1 - F_1)$ , onde  $B_{\log}$  é a inclinação da reta da regressão e  $F_1$  é o  
178 coeficiente  $F_{ij}$  entre indivíduos dentro da primeira classe de distância.

179

## 180 **Análise de agrupamento**

181

182 Dentre os vários métodos para detecção de agrupamentos genéticos, os métodos  
183 espaciais de agrupamento bayesiano estão entre os mais poderosos (Blair et al. 2012).  
184 Foram utilizados dois softwares que incorporam métodos de agrupamento Bayesiano, o  
185 STRUCTURE ver. 2.3.1 (Pritchard et al. 2000) e o GENELAND 4.8 (Guillot et al.,  
186 2005), implementado no R. A análise foi realizada para cada estágio de vida (1) 89  
187 adultos e (2) 62 juvenis.

188 O software STRUCTURE foi utilizado a fim de determinar o número de  
189 agrupamentos genéticos. Para isso, foram realizadas dez corridas independentes para  
190 cada K variando de 1 a 10 agrupamentos genéticos. Cada corrida foi programada para  
191 250.000 simulações de Monte Carlo via cadeias de Markov (MCMC) após 100.000  
192 repetições, assumindo um modelo de mistura com alelos correlacionados entre  
193 indivíduos. Os valores de K e  $\Delta K$  adequados para explicar os dados foram determinados  
194 conforme proposto por Evanno et al. (2005), utilizando software STRUCTURE  
195 HARVESTER (Earl e VonHold 2012).

196 O software GENELAND foi utilizado com intuito de gerar uma exibição gráfica  
197 da organização espacial dos indivíduos dentro de cada grupo determinado pelo

198 STRUCTURE. Para tanto, foi implementado um modelo assumindo: organismo  
199 diploide, certeza ligada às coordenadas espaciais, frequências alélicas com distribuições  
200 independentes, 1000.000 iterações de MCMC (*thinning* = 100), permitindo que o K  
201 variasse de acordo com o número de grupos identificado pelo software STRUCTURE.

202

### 203 **Genética da paisagem em escala local**

204

205 Os dados de uso da terra foram obtidos em dois períodos diferentes (**Fig. 6A e**  
206 **7A**) com intervalo de 30 anos. Para os adultos, a paisagem de 1986, foi extraída a partir  
207 de mapas de cobertura vegetal dos biomas brasileiros na escala espacial 1: 250.000, com  
208 na base no Landsat 7 ETM + do ano de 1986 (ver [www.mma.gov.br](http://www.mma.gov.br)). Para os  
209 indivíduos juvenis, foi utilizada uma paisagem de 2016, a partir de imagens de alta  
210 resolução (*Google Earth*). A configuração diferente do habitat em cada estágio de vida  
211 foi estabelecida com intuito de comparar a influência da paisagem há 30 anos com a  
212 paisagem mais recente, diante das mudanças ocorridas no Cerrado (Grecchi et al. 2015).

213 As imagens obtidas para cada estágio de vida foram vetorizadas em quatro  
214 composições de recursos paisagísticos que incluem estradas (Coffin 2007; Freitas et al.  
215 2013), pastagem, agricultura (Carvalho et al. 2009) e proporção de área coberta por  
216 remanescentes naturais (Telles et al. 2014). Os polígonos de cada composição da  
217 paisagem formados após a vetorização foram convertidos em uma imagem *raster* com  
218 uma dimensão de célula de 5m x 5m (0,00005°). Essas imagens *raster*, de cada  
219 composição da paisagem, foram reunidas em uma única camada usando o software  
220 ARCGIS versão 10.3 (ESRI, Redlands, CA).

221 Apesar de existirem 89 indivíduos adultos e 62 juvenis amostrados, foram  
222 utilizados nesta análise 88 e 47, respectivamente, pois os indivíduos muito próximos

223 geograficamente foram retirados, resultando-se em apenas um indivíduo por célula. Os  
224 pontos de ocorrência para esses indivíduos também foram convertidos em *raster*,  
225 utilizando-se a mesma metodologia. Os dados de composição da paisagem e os pontos  
226 de ocorrência dos indivíduos no formato *raster* foram utilizados como estrutura básica  
227 para a obtenção dos mapas de superfícies de resistência.

228         Para atribuir valores de resistência a diferentes composições da paisagem foram  
229 criados mapas de superfície de resistência utilizando-se a abordagem conhecida como  
230 “*pattern-oriented modeling – (POM)*”, de acordo com o protocolo proposto por  
231 (Wiegand et al. 2003) modificado por Souza (2016). A partir da paisagem estruturada  
232 em *raster*, foram gerados 1.000 mapas de superfícies de resistência, com valores de  
233 custos de resistência variando entre 0 (menor custo) e 100 (maior custo), em intervalos  
234 de 1. Esses valores de custos foram aleatoriamente designados para cada composição da  
235 paisagem, e indicaram a intensidade de resistência imposta por cada composição da  
236 paisagem (pastagem, estrada, agricultura e remanente natural). Os mapas de superfícies  
237 de resistência foram gerados a partir de um *script* no R proposto por Souza (2016).

238         Os 1000 mapas de superfícies de resistência gerados juntamente com os pontos  
239 de ocorrência foram utilizados para obter as matrizes de distância de resistência efetiva  
240 entre os pares de indivíduos. Essas matrizes de distâncias de resistência foram  
241 calculadas de acordo com a teoria do circuito elétrico em uma proposta de isolamento  
242 por resistência (McRae 2006), utilizando software *Circuitscape* (Shah e MacRae 2008)  
243 e o pacote *ResistanceGA* no R (Souza, 2016). O software *Circuitscape* calcula as  
244 matrizes de resistência considerando vários caminhos ao longo da paisagem, fazendo  
245 uma analogia entre a conectividade elétrica e os dados genéticos da espécie em estudo  
246 (Mcrae et al. 2008).

247 As 1000 matrizes de distância de resistências obtidas foram correlacionadas com  
248 a matriz de  $F_{ij}$  par-a-par, usando a função *mantel* no pacote *vegan* no R (Oksanen 2015).  
249 A partir desses resultados foram selecionadas as matrizes distâncias de resistências que  
250 melhor explicavam a distância genética entre os indivíduos, utilizando o critério de  
251 Akaike (AIC) (DeLeeuw 1992) menores que três. Com base nisso, foi gerada uma  
252 matriz de distância de resistência média, a partir da qual foi criado o mapa de “corrente”  
253 das possíveis rotas de dispersão da espécie na paisagem utilizando o *Circuitscape*  
254 (McRae, 2006).

255 Para avaliar o efeito exclusivo da paisagem na variação genética, foram  
256 realizados testes Mantel parcial entre as matrizes de distância genética e as matrizes de  
257 distância de resistência, descontando o efeito das matrizes de distância geográfica.

258

259

## 260 **Resultados**

261

### 262 **Análises de Microssatélites**

263

264 A análise das proporções genotípicas esperadas sobre o equilíbrio de Hardy-  
265 Weinberg (HWE), dadas as frequências dos alelos, identificou desvios significativos ao  
266 equilíbrio HWE em oito locos nos (adultos) e 10 locos nos (juvenis), atendendo aos  
267 pressupostos apenas o loco Dal20 e Dal18. Além disso, utilizando os critérios de  
268 *Bonferroni*, foi detectado desequilíbrio de ligação em um par de locos nos adultos  
269 (Dal23xDal15) e três pares nos juvenis (Dal23xDaE41, Da63xDal15, DaE41xDa80). A  
270 análise de dados brutos no software Micro-Checker não mostrou erros devido à  
271 genotipagem, mas os locos Dal15, DaE63, DaE25 e Dall18 apresentaram alelos nulos.

272

### 273 **Estrutura genética espacial**

274

275 A distância genética com base nos coeficientes de parentesco de Nason ( $F_{ij}$ ) foi  
276 positivamente correlacionada com a distância geográfica, segundo o teste de Mantel,  
277 mostrando que existe um padrão espacial para os indivíduos adultos ( $r = 0,153$ ;  $P <$   
278  $0,001$ ) e juvenis ( $r = 0,193$ ;  $P < 0,001$ ). Os baixos valores de correlação evidenciam que  
279 o parentesco entre os indivíduos é moderadamente explicado pela distância geográfica.

280 De forma similar ao observado pelo teste de Mantel, o autocorrelograma  
281 evidencia significativa autocorrelação espacial nos dois estágios de vida  
282 (adultos e juvenis). Além disso, o coeficiente de determinação em adultos ( $R^2 = 0,053$ ,  
283  $P < 0,001$ ) e juvenis ( $R^2 = 0,127$ ,  $P < 0,001$ ) indica que o parentesco está  
284 significativamente relacionado com a distância geográfica (**Tabela 1**). O coeficiente de  
285 parentesco diminuiu de maneira expressiva após a terceira classe de distância  
286 geográfica, tanto para os indivíduos adultos (**Fig. 2A**), quanto para os juvenis (**Fig. 2B**).  
287 Entretanto, ao observar os valores de coeficiente de parentesco entre os dois estágios de  
288 vida, verifica-se que o parentesco médio dos indivíduos juvenis na primeira classe de  
289 distância geográfica é quase o dobro do observado nos indivíduos adultos. De modo  
290 geral os adultos e juvenis, apresentam significativos valores de intensidade da  
291 estruturação espacial, com valores de  $S_p$  iguais a 0,032 e 0,051, respectivamente  
292 (**Tabela 1**).

293

### 294 **Análise de agrupamento**

295

296 A atribuição bayesiana de indivíduos com software STRUCTURE  
297 mostra a existência grupos genéticos, tanto em adultos ( $\Delta K = 86,66$ ) quanto em juvenis  
298 ( $\Delta K = 174,09$ ), com a formação de dois grupos (**Fig. 3A e B**).

299 A organização espacial dos dois grupos genéticos ao longo da paisagem para os  
300 dois estágios de vida (**Fig. 4A e D**), mostra que os grupos de indivíduos com maior  
301 parentesco dentro da paisagem é diferente entre os indivíduos adultos (**Fig. 4B e C**) e  
302 juvenis (**Fig. 4E e F**). Além disto, observa-se um complexo padrão espacial, ou seja, a  
303 formação destes grupos não é facilmente explicada pela distância geográfica,  
304 principalmente nos indivíduos adultos. Isso mostra que a correlação positiva observada  
305 entre a distância geográfica e o parentesco não está tão explícita.

306

### 307 **Genética da paisagem em escala local**

308

309 A composição da paisagem obtida para o ano de 1986 se diferenciou pouco da  
310 obtida em 2016 (**Fig. 6A e 7A**), com uma paisagem mais homogênea no passado, em  
311 relação à atual.

312 Os testes de Mantel entre as 1000 matrizes de superfície de resistência e as  
313 matrizes de distância genética par-a-par resultaram em valores de coeficiente de  
314 correlação variando entre 0,018 a 0,216 para os adultos e 0,045 a 0,388 para os juvenis.  
315 Foram selecionadas 25 superfícies de resistência para os juvenis e sete para os adultos,  
316 de acordo com os valores críticos de  $\Delta$  Akaike (AIC). As matrizes de distância de  
317 resistência média apresentaram coeficientes de correlação com a distância genética de  $r$   
318 = 0,216,  $P < 0,001$  para os adultos e  $r = 0,388$ ,  $P < 0,001$  para os juvenis (**Fig. 5A e B**).

319 Os mapas de corrente sugerem que as regiões com maior número de indivíduos  
320 aparentados se localizou nas áreas de remanescentes naturais (**Fig. 6B e 7B**). Para os

321 indivíduos adultos a agricultura é a composição da paisagem que impõe maior  
322 resistência entre os grupos (**Fig. 6B**). Para os indivíduos juvenis, tanto a agricultura  
323 quanto a pastagem estão impondo resistência entre os grupos (**Fig. 7B**).

324 Os testes de Mantel parcial mostraram que a paisagem tem uma relevante  
325 associação com a distância genética entre os indivíduos adultos ( $r = 0,165$ ,  $P < 0,001$ ) e  
326 juvenis ( $r = 0,355$ ,  $P < 0,001$ ). Esse resultado, no entanto, mostra que existe uma maior  
327 interferência da paisagem na distribuição espacial da variabilidade genética dos  
328 indivíduos juvenis.

329

### 330 **Discussão**

331

332 Os locos microssatélites apresentaram desvios aos pressupostos do equilíbrio  
333 HWE e desequilíbrio de ligação. O desvio ao equilíbrio HWE e a presença de alelos  
334 nulos pode ser resultado de endogamia biparental, considerando que foi observado alto  
335 grau de parentesco nesta população. Este fato também pode explicar a existência do  
336 desequilíbrio de ligação entre alguns pares de locos.

337 Os resultados mostram que existe moderada estrutura genética espacial (SGS)  
338 nos adultos e juvenis. No entanto, a SGS foi maior nos juvenis ( $r = 0,193$ ), com 3,6% da  
339 variação genética sendo explicada pela distância geográfica. Souza (2016) encontrou ( $r$   
340  $= 0,48$ ) avaliando *D. alata* a nível populacional, uma correlação também moderada,  
341 indicando que outros fatores além da distância geográfica podem influenciar a variação  
342 genética de *D. alata*. A SGS da espécie *D. alata* pode ser explicada pela história de  
343 vida, caracterizada pela formação de grupos de indivíduos com estruturação familiar  
344 conforme observado por Soares et al. (2008) e Collevatti et al. (2010). As sementes de  
345 *D. alata* são dispersas principalmente por barocoria e secundariamente por macacos,

346 morcegos e roedores (Sano et al. 2004), que podem restringir as distâncias de dispersão  
347 de sementes, causando SGS (Soares et al. 2008). No entanto, a dispersão de pólen é  
348 mediada por abelhas (Oliveira e Sigrist 2008), com dispersão do pólen por longa  
349 distância (Tambarussi et al. 2017). De modo que a estrutura genética observada é dada  
350 por um balanço entre a contribuição do pólen e da semente.

351 Os valores de  $S_p$  em adultos (0,0329) e juvenis (0,0514) são relevantes quando  
352 comparados aos observados para outras espécies de plantas por exemplo. Esses valores  
353 são semelhantes aos encontrados em espécies com sistema reprodutivo misto, que  
354 apresentam dispersão de pólen por animais e dispersão de sementes por barocoria  
355 (Vekemans e Hardy 2004). A diferença de  $S_p$  entre os estágios de vida também foi  
356 observada dentro de populações de espécies neotropicais dispersas por animais  
357 (Collevatti e Hay 2011; Costa et al. 2017). A intensidade da SGS observada para *D.*  
358 *alata* foi maior que para outras espécies arbóreas frutíferas do cerrado dispersas por  
359 animais como (*Caryocar brasiliense*, Collevatti et al. 2010; Collevatti e Hay 2011;  
360 *Annona crassiflora*, Collevatti et al. 2014; *Hancornia speciosa*, Costa et al. 2017) para  
361 as quais a fragmentação é apontada como a principal causa da alta endogamia. Assim  
362 mesmo estando em uma área com cerca de 60% de remanescente natural, as alterações  
363 antrópicas podem estar causando impactos na população estudada.

364 O parentesco médio entre os indivíduos em classes de distâncias diminuiu com  
365 aumento da distância geográfica, conforme foi observado por Soares et al. (2008),  
366 Tarazi et al. (2010) e Tambarussi et al. (2017). Para os adultos, o parentesco foi  
367 explicado pela distância até 500m, o dobro da distância observada nos juvenis. Esse fato  
368 pode ser explicado pela dispersão barocórica das sementes, na qual sementes mais  
369 aparentadas são dispersas mais próximas geograficamente. Por exemplo, indivíduos  
370 juvenis exibem maior estruturação genética dentro da população, em relação aos adultos

371 que podem ter outros fatores que explicam sua distribuição no espaço ao longo do  
372 desenvolvimento (Collevatti e Hay 2011; Costa et al. 2017).

373 A atribuição de indivíduos em grupos genéticos  $K = 2$  em ambos os estágios de  
374 vida produziu uma imagem clara da estrutura genética da população em dois grupos. No  
375 entanto, o maior parentesco (similaridade genética) dentro dos grupos não foi facilmente  
376 explicado pela distância geográfica entre os indivíduos, uma vez que indivíduos  
377 relativamente distantes geograficamente foram alocados em um mesmo grupo.

378 Os mapas de corrente mostram que a composição da paisagem que mais  
379 contribuiu para a existência de descontinuidade genética nos adultos foi a agricultura,  
380 enquanto nos juvenis tanto a área de pastagem como agricultura estão interferindo. Isso  
381 é confirmado pelo coeficiente de correlação da matriz de distância de resistência  
382 selecionada com a matriz distância genética que foi bem superior nos juvenis ( $r =$   
383  $0,388$ ), mostrando que as características da paisagem apresentam maior interferência na  
384 variação genética entre indivíduos juvenis do que nos adultos. Além disso, esse  
385 resultado é corroborado pelo teste Mantel parcial ( $r = 0,3553$ ) nos juvenis, que  
386 evidencia a importância da composição da paisagem da variabilidade genética destes  
387 indivíduos.

388 Os indivíduos juvenis apresentaram maior interferência das composições da  
389 paisagem. Isso pode estar relacionado com a paisagem mais heterogênea, fazendo com  
390 que as plantas recebam menos visitantes florais, devido ao declínio na abundância de  
391 polinizadores, e a limitação de movimento entre as manchas de remanescentes (Goverde  
392 et al. 2002). Além disso, a fragmentação também altera a abundância e o  
393 comportamento dos animais dispersores, o que afeta a dispersão de sementes (Ghazoul  
394 2005). Assim os efeitos de ocupação humana, relacionada à agricultura e pastagem  
395 podem está influenciando a distribuição da variabilidade genética dos indivíduos de *D*.

396 *alata*. Fator já identificado por Telles et al. (2014), que mostram que apesar da  
397 ocupação humana no Cerrado ser recente, já existem efeitos significativos da  
398 fragmentação do habitat nas populações de *D. alata*.

399 Este estudo mostra que as ações antrópicas, relacionadas à perda de habitat e  
400 fragmentação causam maior interferência na variação genética dos indivíduos juvenis.  
401 Este resultado é particularmente importante, pois estes indivíduos irão recompor a  
402 população de *D. alata* desta área. Assim, essas descobertas devem ser consideradas para  
403 subsidiar estratégias mais eficientes de conservação da espécie.

404

#### 405 **Agradecimentos**

406

407 O trabalho de R. A. Guimarães, K. da S. Souza foi apoiado por uma bolsa da  
408 CAPES e o trabalho de M. P. C. Telles e T. N. Soares também são apoiados por bolsas  
409 de produtividade do CNPq. Esta pesquisa foi desenvolvida no contexto do Instituto  
410 Nacional de Ciência e Tecnologia (INCT) em Ecologia, Evolução e Conservação da  
411 Biodiversidade, apoiado pelo MCTIC / CNPq (proc. 465610 / 2014-5) e pela FAPEG.

412

#### 413 **Conflito de interesses**

414 Os autores declaram não haver conflito de interesses.

415

#### 416 **Referências**

417

418 Alcalá RE, Cruz SD La, Gutiérrez-Granados G (2015) Genetic structure and genetic  
419 diversity of *Swietenia macrophylla* in areas subjected to selective logging in Quintana  
420 Roo, Mexico. Bot Sci 93:819–828. doi: 10.17129/botsci.256

421 Almeida SP de, Proença CEB, Sano SM, Ribeiro JF (1998) Cerrado: espécies vegetais  
422 úteis.

423 Anderson CD, Epperson BK, Fortin MJ, et al (2010). Considering spatial and temporal  
424 scale in landscape-genetic studies of gene flow. *Mol Ecol* 19:3565–3575. doi:  
425 10.1111/j.1365-294X.2010.04757.x

426 Balkenhol N, Cushman SA, Storfer A, Waits LP (2015) Landscape Genetics: concepts,  
427 methods, applications.

428 Barluenga M, Austerlitz F, Elzinga JA, et al (2011) Fine-scale spatial genetic structure  
429 and gene dispersal in *Silene latifolia*. *Heredity* (Edinb) 106:13–24. doi:  
430 10.1038/hdy.2010.38

431 Blair C, Weigel DE, Balazik M, et al (2012) A simulation-based evaluation of methods  
432 for inferring linear barriers to gene flow. *Mol Ecol Resour* 12:822–833. doi:  
433 10.1111/j.1755-0998.2012.03151.x

434 Carvalho FM V, Júnior PDM, Ferreira LG (2009) The Cerrado into-pieces : Habitat  
435 fragmentation as a function of landscape use in the savannas of central Brazil. *Biol*  
436 *Conserv* 142:1392–1403. doi: 10.1016/j.biocon.2009.01.031

437 Coffin AW (2007) From roadkill to road ecology : A review of the ecological effects of  
438 roads. *J Transp Geogr* 15:396–406. doi: 10.1016/j.jtrangeo.2006.11.006

439 Collevatti RG, Hay JD (2011) Kin structure and genotype-dependent mortality: A study  
440 using the Neotropical tree *Caryocar brasiliense*. *J Ecol* 99:757–763. doi:  
441 10.1111/j.1365-2745.2011.01796.x

442 Collevatti RG, Lima JS, Soares TN, Telles MP de C (2010) Spatial genetic structure and  
443 life history traits in Cerrado tree species: Inferences for Conservation. *Nat Conserv*  
444 8:54–59. doi: 10.4322/natcon.00801008

445 Collevatti RG, Telles MPC, Lima JS, et al (2014) Contrasting spatial genetic structure  
446 in *Annona crassiflora* populations from fragmented and pristine savannas. *Plant Syst*  
447 *Evol* 300:1719–1727. doi: 10.1007/s00606-014-0997-9

448 Costa CF, Collevatti RG, Chaves LJ, et al (2017) Genetic diversity and fine-scale  
449 genetic structure in *Hancornia* (Apocynaceae). *Biochem Syst Ecol J* 72:63–67. doi:  
450 10.1016/j.bse.2017.03.001

451 Cushman SA, Landguth EL (2010a) Spurious correlations and inference in landscape  
452 genetics. *Mol Ecol* 19:3592–3602. doi: 10.1111/j.1365-294X.2010.04656.x

453 Cushman SA, Landguth EL (2010b) Scale dependent inference in landscape genetics.  
454 *Landsc Ecol* 25:967–979. doi: 10.1007/s10980-010-9467-0

455 DeLeeuw J (1992) Introduction to Akaike (1973) Information Theory and an Extension  
456 of the Maximum Likelihood Principle. *Break Stat I*: 599–609.

457 Diniz-filho JAF, Soares TN, Telles MP de C (2014) Pattern-oriented modelling of  
458 population genetic structure. *Biol J Linn Soc* 113:1152–1161.

459 Doligez A, Joly HI (1997) Genetic diversity and spatial structure within a natural stand  
460 of a tropical forest tree species, *Carapa procera* (Meliaceae), in French Guiana.  
461 *Heredity (Edinb)* 79:72–82.

462 Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of  
463 fresh leaf tissue. *Phytochem Bull* 19:11–15.

464 Earl DA, VonHold BM (2012) STRUCTURE HARVESTER : a website and program  
465 for visualizing STRUCTURE output and implementing the Evanno method. *Conserv*  
466 *Genet Resour* 4:359–361. doi: 10.1007/s12686-011-9548-7

467 Epperson BK (2003) *Geographical genetics*. Princeton University Press

468 Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals  
469 using the software STRUCTURE: A simulation study. *Mol Ecol* 14:2611–2620. doi:  
470 10.1111/j.1365-294X.2005.02553.x

471 Freitas SR, Sousa COM, Boscolo D, Metzger JP (2013) How are native vegetation and  
472 reserves affected by different road types in a Southeastern Brazilian State ? *Oecologia*  
473 *Aust* 17:447–458.

474 Ghazoul J (2005) Pollen and seed dispersal among dispersed plants. *Biol Rev* 80:413–  
475 443. doi: 10.1017/S1464793105006731

476 Goverde M, Schweizer K, Baur B, Erhardt A (2002) Small-scale habitat fragmentation  
477 effects on pollinator behaviour: Experimental evidence from the bumblebee *Bombus*  
478 *veteranus* on *calcareous grasslands*.

479 Grecchi RC, Beuchle R, Shimabukuro YE, et al (2015) Assessing land cover changes in  
480 the Brazilian Cerrado between 1990 and 2010 using a remote sensing sampling

481 approach. In: Anais XVII Simpósio Brasileiro de Sensoriamento Remoto - SBSR. pp  
482 2860–2866

483 Grimm V, Franka K, Jeltsch F, et al (1996) Pattern-oriented modelling in population  
484 ecology. *Sci Total Environ* 183:151–166.

485 Grimm V, Revilla E, Berger U, et al (2005) Pattern-Oriented Modeling of Agent-Based  
486 Complex Systems : Lessons from Ecology. *Science* (80- ) 310:987. doi:  
487 10.1126/science.1116681

488 Guillot G, Mortier F, Estoup A (2005) GENELAND: A computer package for  
489 landscape genetics. *Mol Ecol Notes* 5:712–715. doi: 10.1111/j.1471-8286.2005.01031.x

490 Guimarães RA, Telles MPC, Antunes AM, et al (2017) Discovery and characterization  
491 of new microsatellite loci in *Dipteryx alata* vogel (Fabaceae) using next-generation  
492 sequencing data. *Genet Mol Res* 16:1–6. doi: <http://dx.doi.org/10.4238/gmr16029639>

493 Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyse  
494 spatial genetic structure at the individual or population levels OLIVIER. *Mol Ecol*  
495 *Notes* 2:618–620. doi:10.1046/j.1471-8278

496 Hardy OJ, Vekemans X V (2015) SPAGeDi 1.5 a program for Spatial Pattern Analysis  
497 of Genetic Diversity. *Evol Biol Ecol* 2002–2015.

498 Jordano P, García C, Godoy J a, García-Castaño JL (2007) Differential contribution of  
499 frugivores to complex seed dispersal patterns. *Proc Natl Acad Sci U S A* 104:3278–  
500 3282. doi: 10.1073/pnas.0606793104

501 Kierepka EM, Latch EK (2015) Performance of partial statistics in individual-based  
502 landscape genetics. *Mol Ecol Resour* 15:512–525. doi: 10.1111/1755-0998.12332

503 Lewis PO, Zaykin D (2002) GDA User’s Manual. 1–45.

504 Loiselle BA, Sork VL, Nason J, Graham C (1995) Spatial genetic structure of a tropical  
505 understory shrub, *Psychotria Officinalis* (Rubiaceae). *Am J* 82:1420–1425.

506 Manel S, Holderegger R (2013) Ten years of landscape genetics. *Trends Ecol Evol*  
507 28:614–621. doi: 10.1016/j.tree.2013.05.012

508 Manel S, Joost S, Epperson BK, et al (2010) Perspectives on the use of landscape  
509 genetics to detect genetic adaptive variation in the field. *Mol Ecol* 19:3760–3772. doi:  
510 10.1111/j.1365-294X.2010.04717.x

511 Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: Combining  
512 landscape ecology and population genetics. *Trends Ecol Evol* 18:189–197. doi:  
513 10.1016/S0169-5347(03)00008-9

514 McRae BH (2006) Isolation by resistance. *Evolution (N Y)* 60:1551–1561.

515 McRae BH, Beier P (2007) Circuit theory predicts gene flow in plant and animal  
516 populations. *Proc Natl Acad Sci* 104:19885–19890. doi: 10.1073/pnas.0706568104

517 Mcrae BH, Dickson BG, Keitt TH, Shah VB (2008) Using Circuit Theory to Model  
518 Connectivity in Ecology, Evolution, and Conservation. *Ecology* 89:2712–2724. doi: Doi  
519 10.1890/07-1861.1

520 Medina-Macedo L, Sebbenn AM, Lacerda AEB, et al (2015) High levels of genetic  
521 diversity through pollen flow of the coniferous *Araucaria angustifolia*: a landscape  
522 level study in Southern Brazil. *Tree Genet Genomes* 11:1–14. doi: 10.1007/s11295-014-  
523 0814-1

524 Milanesi P, Holderegger R, Caniglia R, et al (2017) Expert-based versus habitat-  
525 suitability models to develop resistance surfaces in landscape genetics. *Oecologia*  
526 183:67–79. doi: 10.1007/s00442-016-3751-x

527 Newman BJ, Ladd P, Brundrett M, Dixon KW (2013) Effects of habitat fragmentation  
528 on plant reproductive success and population viability at the landscape and habitat scale.  
529 *Biol Conserv* 159:16–23. doi: 10.1016/j.biocon.2012.10.009

530 Ng KKS, Lee SL, Saw LG, et al (2006) Spatial structure and genetic diversity of three  
531 tropical tree species with different habitat preferences within a natural forest. *Tree*  
532 *Genet Genomes* 2:121–131. doi: 10.1007/s11295-006-0035-3

533 Oksanen J (2015) *Multivariate Analysis of Ecological Communities in R : vegan*  
534 *tutorial*.

535 Oliveira MIB, Sigrist R (2008) Fenologia reprodutiva, polinização e reprodução de  
536 *Dipteryx alata* Vogel ( Leguminosae-Papilionoideae ) em Mato Grosso do Sul , Brasil  
537 1. *Rev Bras Botânica* 31:195–207.

538 Oosterhout C Van, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO -  
539 CHECKER : software for identifying and correcting genotyping errors in microsatellite  
540 data. *Mol Ecol Notes* 4:535–538. doi: 10.1111/j.1471-8286.2004.00684.x

541 Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using  
542 multilocus genotype data. *Genetics* 155:945–959. doi: 10.1111/j.1471-  
543 8286.2007.01758.x

544 Sano SM, Ribeiro JF, Brito MA (2004) *Baru: biologia e uso*. Embrapa cerrados -  
545 Brasília DF

546 Segelbacher G, Cushman SA, Epperson BK, et al (2010) Applications of landscape  
547 genetics in conservation biology: Concepts and challenges. *Conserv Genet* 11:375–385.  
548 doi: 10.1007/s10592-009-0044-5

549 Shah VB, MacRae B (2008) Circuitscape : A Tool for Landscape Ecology. *Proc 7th*  
550 *Python Sci Conf* 62–65.

551 Soares TN, Chaves LJ, Telles M de CP, et al (2008) Distribuição espacial da  
552 variabilidade genética intrapopulacional de *Dipteryx alata*. *Pesqui Agropecuária Bras*  
553 43:1151–1158.

554 Soares TN, Diniz-Filho JAF, Nabout JC, et al (2015) Patterns of genetic variability in  
555 central and peripheral populations of *Dipteryx alata* ( Fabaceae ) in the Brazilian  
556 Cerrado. *Plant Syst Evol* 301:1315–1324. doi: 10.1007/s00606-014-1155-0

557 Soares TN, Melo DB, Resende LV, et al (2012) Development of microsatellite markers  
558 for the neotropical tree species *Dipteryx alata* (Fabaceae). *Am J Bot* 99:5–6. doi:  
559 10.3732/ajb.1100377

560 Souza K da S e (2016) Desenvolvimento de superfícies de resistência explicando a  
561 variação genética por modelagem orientada por padrão: uma análise com espécies  
562 arbóreas de cerrado. *Dissertação*, Universidade Federal de Goiás.

563 Spear SF, Balkenhol N, Fortin M-J, et al (2010) Use of resistance surfaces for landscape  
564 genetic studies : considerations for parameterization and analysis. *Mol Ecol* 19:3576–  
565 3591. doi: 10.1111/j.1365-294X.2010.04657.x

566 Tambarussi E V, Sebbenn AM, Alves-Pereira A, et al (2017) *Dipteryx alata* Vogel (  
567 Fabaceae ), a Neotropical tree with high levels of selfing : implications for conservation  
568 and breeding programs. *Ann For Res*. doi: 10.15287/afr.2017.842

569 Tarazi R, Moreno MA, Gandara FB, et al (2010) High levels of genetic differentiation  
570 and selfing in the Brazilian cerrado fruit tree *Dipteryx alata* Vog. (Fabaceae). *Genet*  
571 *Mol Biol* 33:78–85. doi: 10.1590/S1415-47572010005000007

572 Telles MP de C, Dobrovolski R, Souza K da S e, et al (2014) Disentangling landscape  
 573 effects on population genetic structure of a Neotropical savanna tree. *Nat Conserv*  
 574 12:65–70. doi: 10.4322/natcon.2014.012

575 Vekemans X, Hardy OJ (2004) New insights from fine-scale spatial genetic structure  
 576 analyses in plant populations. *Mol Ecol* 13:921–935. doi: 10.1046/j.1365-  
 577 294X.2004.02076.x

578 Wiegand T, Jeltsch F, Hanski I, Grimm V (2003) Using pattern-oriented modeling for  
 579 revealing hidden information: a key for reconciling ecological theory and application.  
 580 *Oikos* 100:209–222. doi: 10.1034/j.1600-0706.2003.12027.x

581 Yao X, Zhang J, Ye Q, Huang H (2011) Fine-scale spatial genetic structure and gene  
 582 flow in a small, fragmented population of *Sinojackia rehderiana* (Styracaceae), an  
 583 endangered tree species endemic to China. *Plant Biol* 13:401–410. doi: 10.1111/j.1438-  
 584 8677.2010.00361.x

585  
 586  
 587

588 **Tabela 1** Parâmetros de estrutura genética espacial de indivíduos adultos e juvenis de  
 589 uma população de *D. alata*

Estágios de vida	R <sup>2</sup>	B <sub>log</sub>	F <sub>1</sub> (1 classe)	S <sub>p</sub>
Adultos	0,0535	-0,0297	0,0960	0,0329
Juvenis	0,1268	-0,0416	0,1910	0,0514

590 R<sup>2</sup> coeficiente de determinação, B<sub>log</sub> inclinação da regressão, F<sub>1</sub> parentesco da primeira  
 591 classe de distância, S<sub>p</sub> intensidade da estruturação espacial

592  
 593  
 594  
 595  
 596

597

598

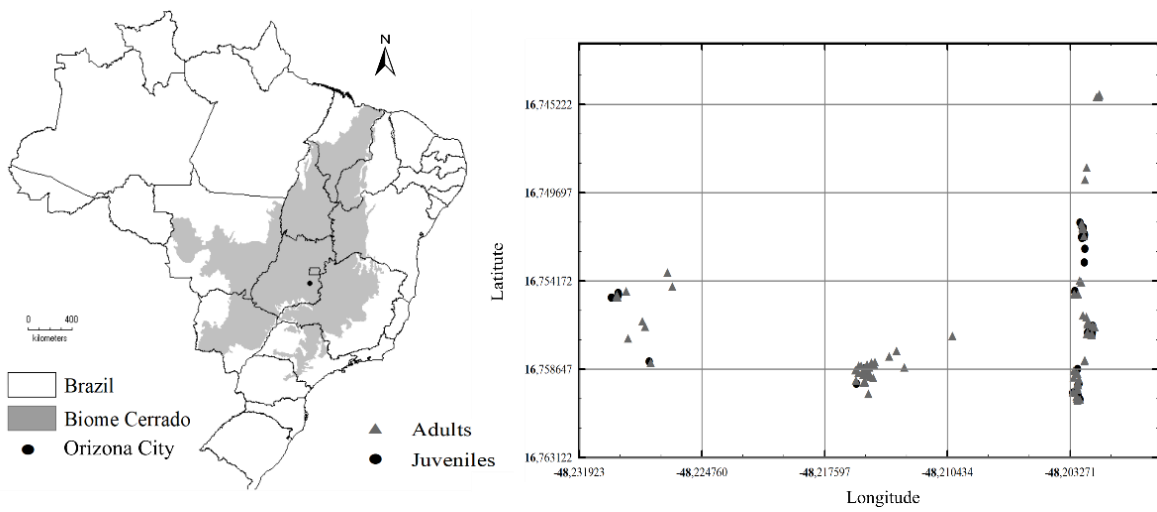
599

600

601 **Figuras**

602

603



604

605 **Fig. 1** Distribuição espacial de indivíduos adultos e juvenis amostrados em uma  
606 população da espécie *Dipteryx alata* localizada em Orizona-Goiás, no Centro-Oeste do  
607 Brasil.

608

609

610

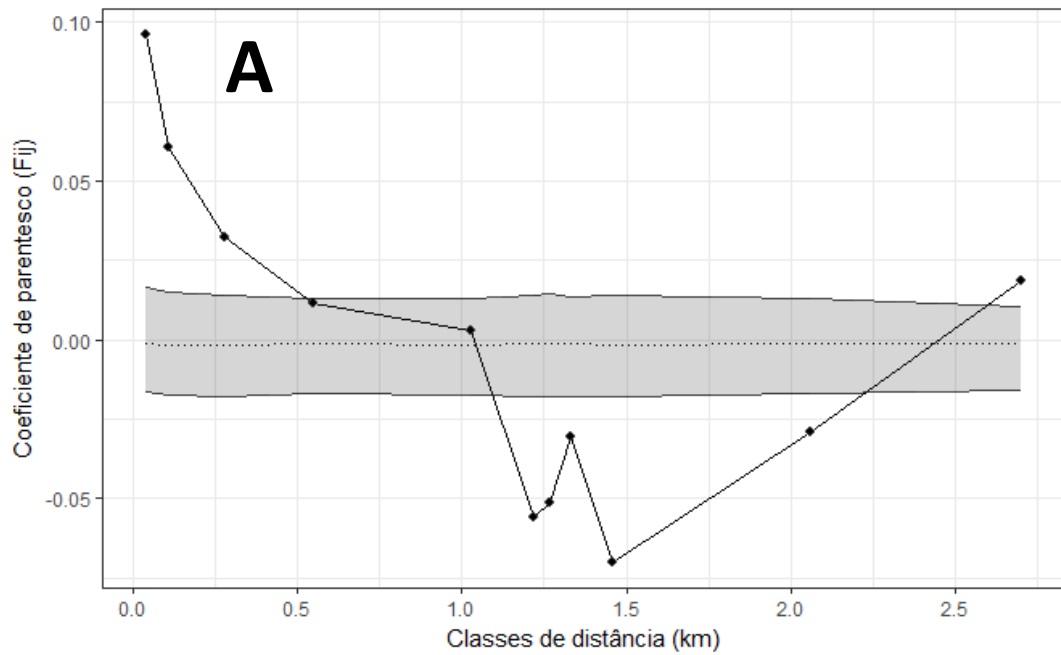
611

612

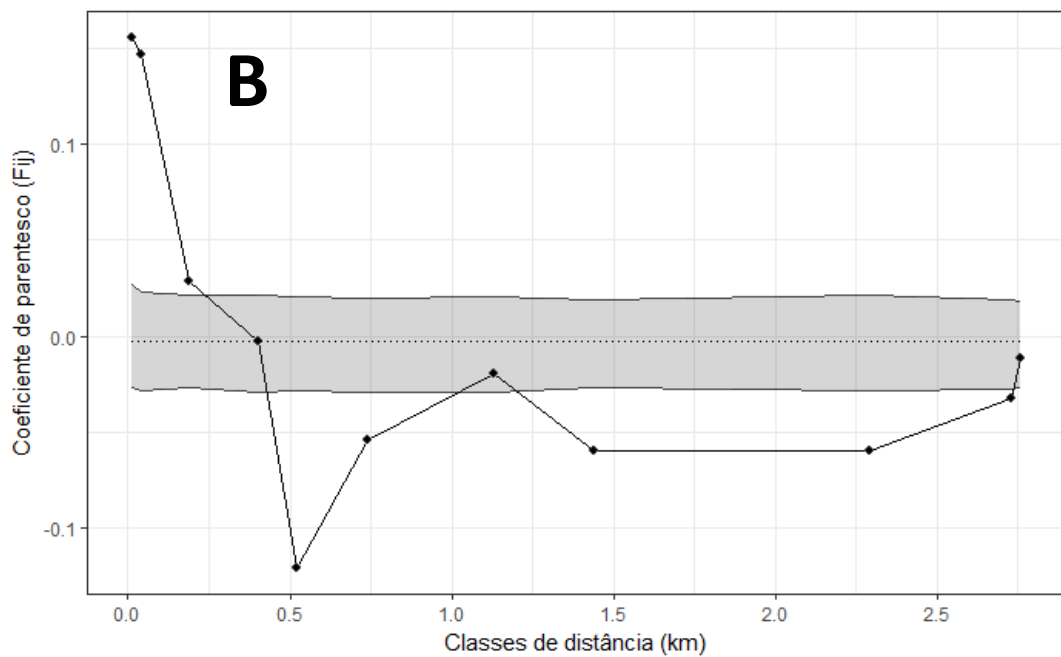
613

614

615



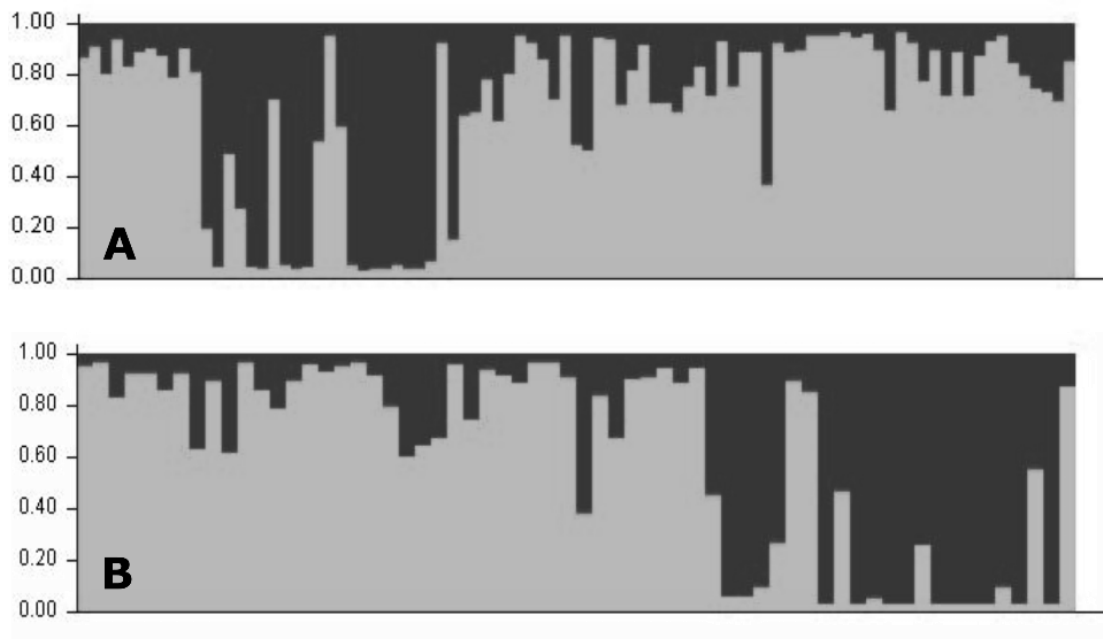
616



617

618 **Fig. 2** Autocorrelogramas espaciais que mostram a relação entre parentesco médio ( $F_{ij}$ )  
 619 e 11 classes de distância geográfica para adultos (A) e juvenis (B) de uma população da  
 620 espécie *Dipteryx alata* no Cerrado Brasileiro. A faixa cinza indica o intervalo de  
 621 confiança (95%, 10.000 permutações).

622



623

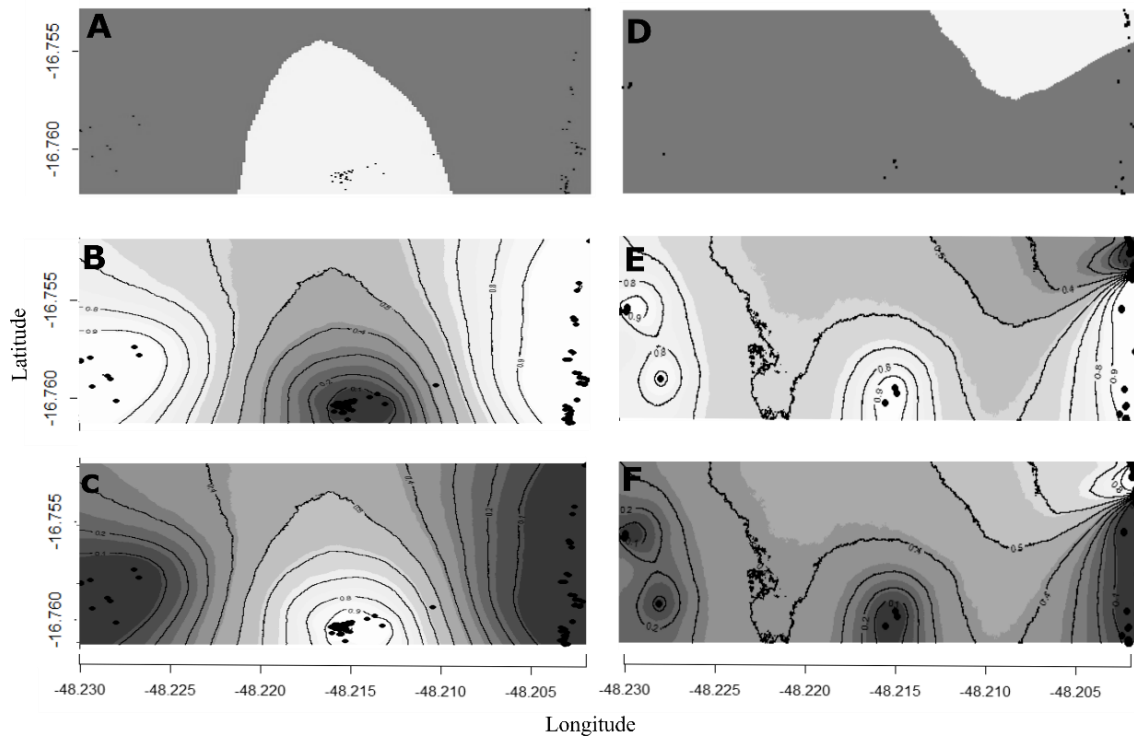
624 **Fig. 3** Resultado da atribuição genética de indivíduos adultos (A) e juvenis (B) em dois  
 625 grupos ( $K=2$ ) realizada no software STRUCTURE. Cada coluna representa um indivíduo  
 626 e os grupos são codificados com cores diferentes. A cor é atribuída a cada indivíduo a  
 627 partir da probabilidade de atribuição a cada grupo.

628

629

630

631



632

633 **Fig. 4** Resultados das análises no software GENELAND para os indivíduos adultos (A,  
 634 B e C), e juvenis (D, E e F) de uma população da espécie *Dipteryx alata* no Cerrado  
 635 Brasileiro. A figura mostra: (A e D) a organização espacial em dois grupos genéticos;  
 636 (B e C, E e F) são mapas de probabilidades posteriores de cada grupo. As  
 637 probabilidades posteriores mais altas são representadas em branco e as mais baixas  
 638 estão em cinza escuro. As linhas de contorno representam a mudança espacial na  
 639 probabilidade de atribuição de cada grupo.

640

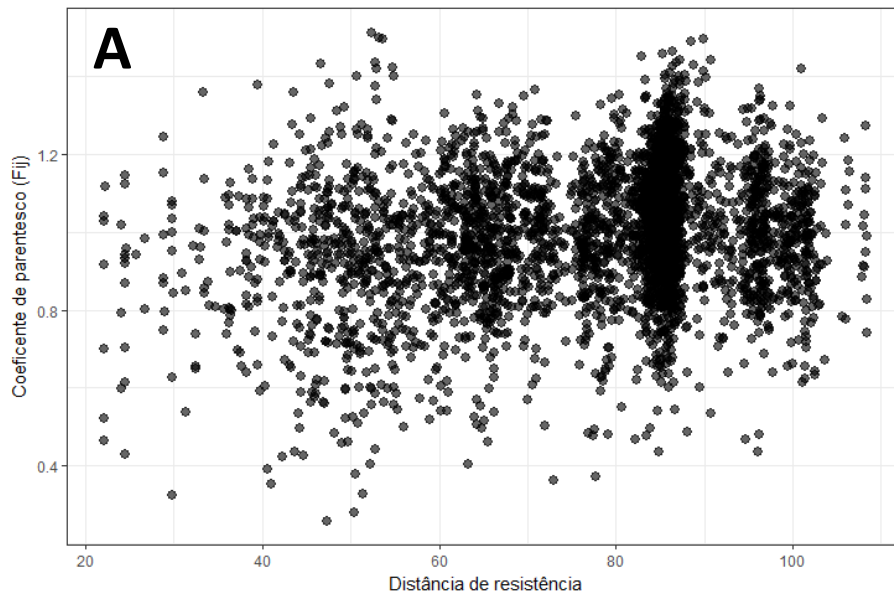
641

642

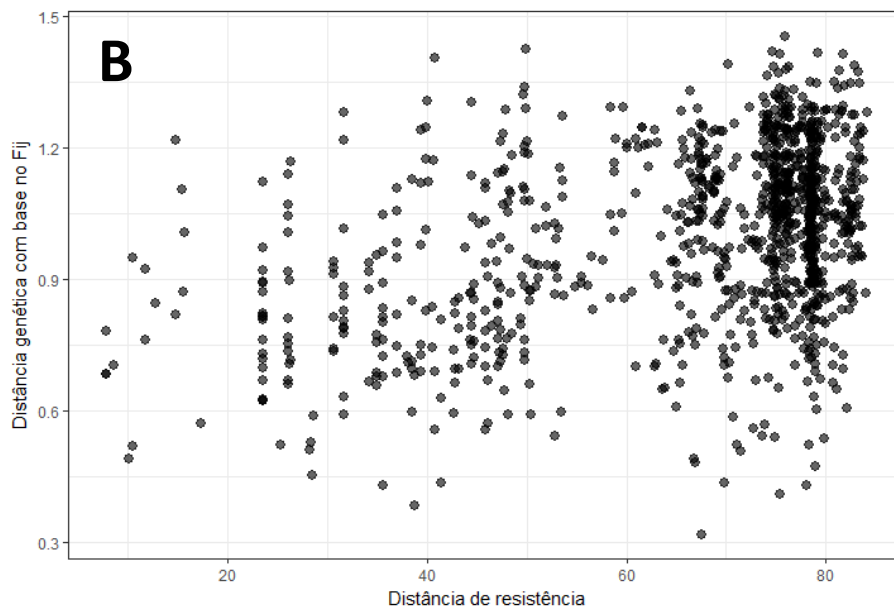
643

644

645



646



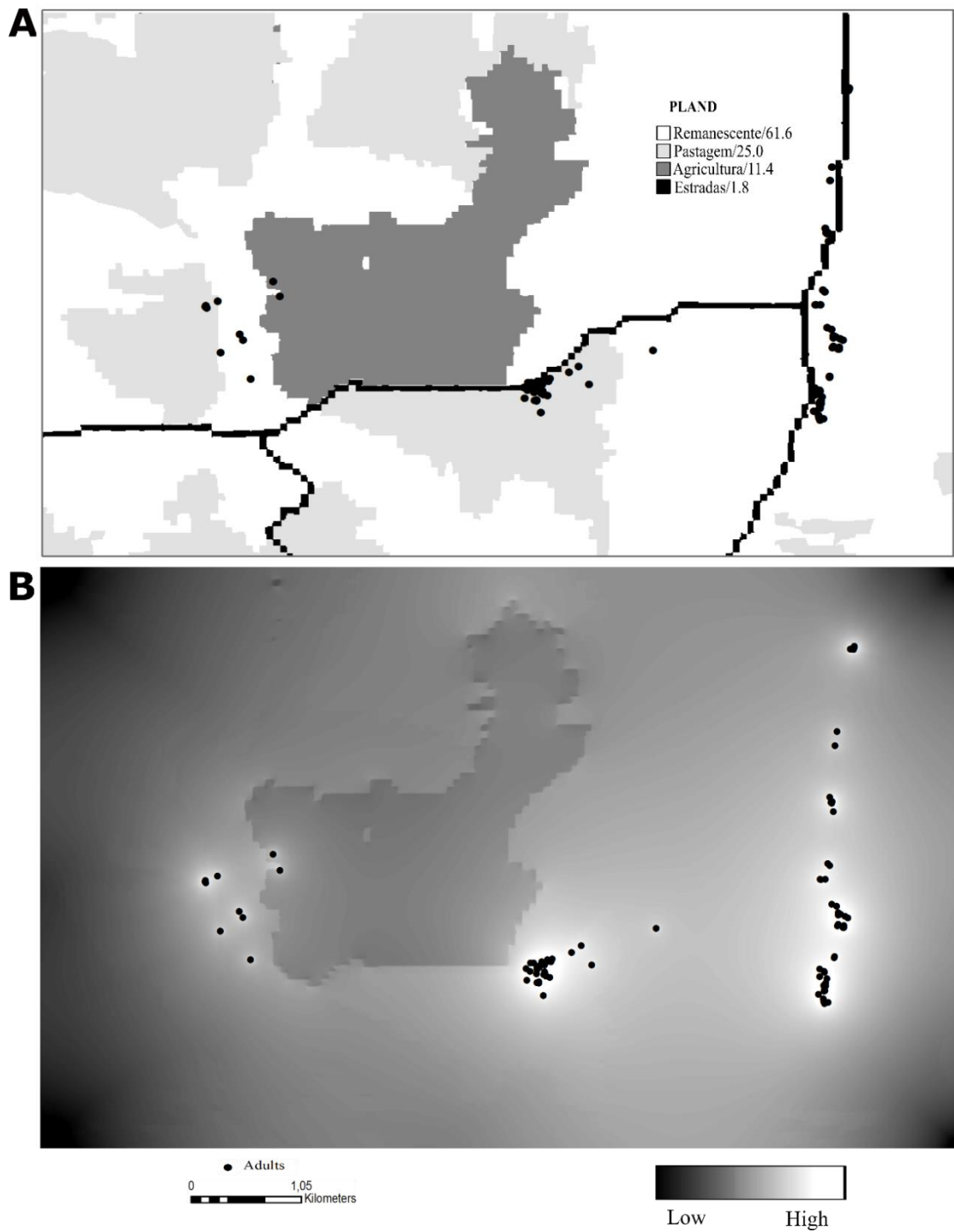
647

648 **Fig. 5** Relação entre o coeficiente de parentesco de ( $F_{ij}$ ) e a distância de resistência entre  
 649 pares de indivíduos (A) adultos e (B) juvenis em uma população da espécie *Dipteryx*  
 650 *alata* no Cerrado Brasileiro.

651

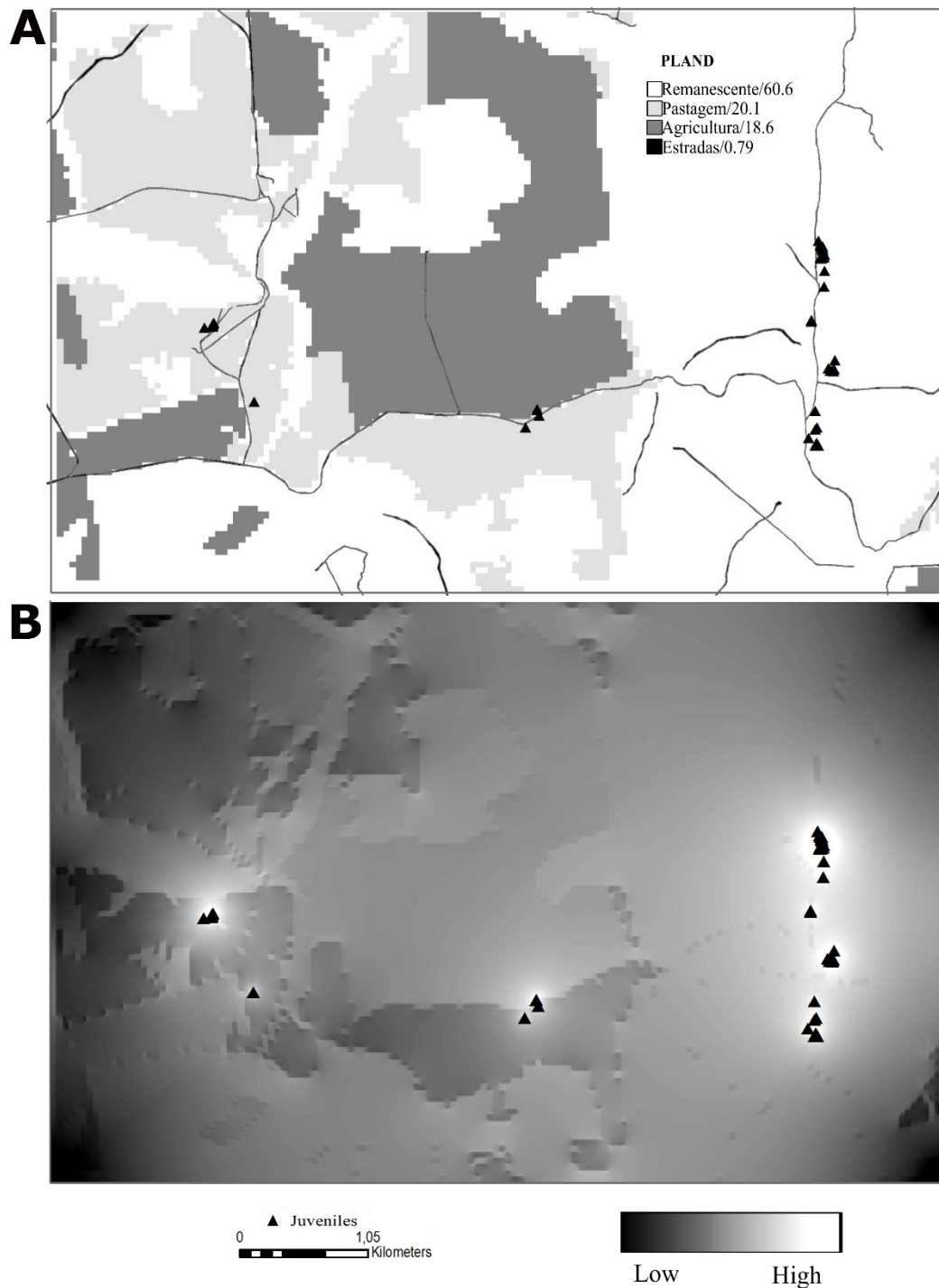
652

653



654

655 **Fig. 6** Paisagem de 1986 para a área de amostragem de indivíduos adultos de uma  
 656 população da espécie *Dipteryx alata* no Cerrado Brasileiro. (A) Caracterização das  
 657 composições da paisagem. PLAND: porcentagem de área na paisagem, (B) Mapa de  
 658 corrente com as principais vias de menor resistência para indivíduos adultos. As áreas  
 659 mais escuras possuem menor corrente e as mais claras tem corrente mais alta.



660

661 **Fig. 7** Paisagem de 2016 para a área de amostragem de indivíduos juvenis de uma  
 662 população da espécie *Dipteryx alata* no Cerrado Brasileiro. (A) Caracterização das  
 663 composições da paisagem. PLAND: porcentagem de área na paisagem, (B) Mapa de  
 664 corrente com as principais vias de menor resistência para indivíduos juvenis. As áreas  
 665 mais escuras possuem menor corrente e as mais claras tem corrente mais alta.

## 5 CONSIDERAÇÕES FINAIS

- A espécie *Dipteryx alata* conta atualmente com 19 marcadores microssatélites polimórficos publicados (Soares et al., 2012, Guimarães et al. 2017). Entretanto, estes marcadores exibem baixo polimorfismo médio, o que fez com que dez pudessem ser utilizados nos artigos deste estudo. Este fato corrobora com outros trabalhos que discutem que a espécie exibe baixa variação genética;
- As estimativas de taxa de fecundação cruzada confirmam que espécie *D. alata* apresenta um sistema reprodutivo misto, com predominância de alogamia;
- A condição *ex situ* exibe alto número de sementes que compartilham as mesmas doadoras de pólen, sendo que na condição *in situ* esse número foi maior ainda, o que indica baixo número de doadoras de pólen por árvore nas populações estudadas;
- A distância de dispersão de pólen mostra que na condição *in situ* os polinizadores estão atingindo grandes distâncias de dispersão, quando comparadas à condição *ex situ*;
- A inclusão das características da paisagem trouxe um ganho de explicação para a estrutura genética em escala local na população de *D. alata* estudada;
- Tanto nos indivíduos adultos quanto nos juvenis, foram detectados dois grupos genéticos com maior parentesco dentro deles;
- A composição da paisagem que mais contribuiu para a existência de descontinuidade genética nos indivíduos adultos foi a agricultura, enquanto nos juvenis tanto a pastagem quanto a agricultura estão interferindo. Estes resultados são relevantes para a conservação da espécie *D. alata*.

## 6 REFERÊNCIAS

- ALMEIDA, S. P. DE; PROENÇA, C. E. B.; SANO, S. M.; RIBEIRO, J. F. Cerrado: espécies vegetais úteis. **Planaltina: Embrapa-CPAC**, v. 464, 1998.
- BALKENHOL, N.; CUSHMAN, S. A.; STORFER, A.; WAITS, L. P. **Landscape Genetics: concepts, methods, applications**. 2015.
- BARLUENGA, M.; AUSTERLITZ, F.; ELZINGA, J. A.; TEIXEIRA, S.; GOUDET, J.; BERNASCONI, G. Fine-scale spatial genetic structure and gene dispersal in *Silene latifolia*. **Heredity**, v. 106, n. 1, p. 13–24, 2011.
- BATISTA, A. C. F.; RODRIGUES, H. DE S.; PEREIRA, N. R.; HERNANDEZ-TERRONES, M. G.; VIEIRA, A. T.; OLIVEIRA, M. F. DE. Use of baru oil (*Dipteryx alata* vog.) to produce biodiesel and study of the physical and chemical characteristics of biodiesel/petroleum diesel fuel blends. **Chemistry and Technology of Fuels and Oils**, v. 48, n. 1, p. 11–13, 2012.
- CANUTO, D. S. DE O.; ZARUMA, D. U. G.; MORAES, M. A. DE; SILVA, A. M. DA; MARIO, L. T. DE M.; FREITAS, M. L. F. Caracterização genética de um teste de progênies de *Dipteryx alata* Vog. proveniente de remanescente florestal da Estação Ecológica de Paulo de Faria, SP, Brasil. **Hoehnea**, v. 42, n. 2, p. 641–648, 2015.
- COLLEVATTI, R. G.; LIMA, J. S.; SOARES, T. N.; TELLES, M. P. DE C. Spatial genetic structure and life history traits in Cerrado tree species: inferences for conservation. **Natureza & Conservação**, v. 8, n. 1, p. 54–59, 2010.
- COLLEVATTI, R. G.; TELLES, M. P. DE C.; LIMA, J. S.; GOUVEIA, F. O.; SOARES, T. N. Contrasting spatial genetic structure in *Annona crassiflora* populations from fragmented and pristine savannas. **Plant Systematics and Evolution**, v. 300, n. 7, p. 1719–1727, 2014.
- COLLEVATTI, R. G.; TELLES, M. P. DE C.; NABOUT, J. C.; CHAVES, L. J.; SOARES, T. N. Demographic history and the low genetic diversity in *Dipteryx alata* (Fabaceae) from Brazilian Neotropical savannas. **Heredity**, v. 111, n. 2, p. 97–105, 2013.
- CORRÊA, G. DE C.; ROCHA, M. R. DA; NAVES, R. V. Germinação de sementes e emergência de plântulas de Baru (*Dipteryx alata* Vog.) nos Cerrados do estado de Goiás. **Pesquisa Agropecuária Tropical**, v. 30, n. 2, p. 17–23, 2000.
- COSTA, C. F.; COLLEVATTI, R. G.; CHAVES, L. J.; LIMA, J. D. S.; SOARES, T. N.; TELLES, M. P. DE C. Genetic diversity and fine-scale genetic structure in *Hancornia* (Apocynaceae). **Biochemical Systematics and Ecology journal**, v. 72, n. Supplement C, p. 63–67, 2017.
- CUSHMAN, S. A.; LANDGUTH, E. L. Scale dependent inference in landscape genetics. **Landscape Ecology**, v. 25, n. 6, p. 967–979, 2010b.
- CUSHMAN, S. A.; LANDGUTH, E. L. Spurious correlations and inference in landscape genetics. **Molecular Ecology**, v. 19, n. 17, p. 3592–3602, 2010a.

- CZEDER, L. P.; FERNANDES, D. C.; FREITAS, J. B.; NAVES, M. M. V. Baru almonds from different regions of the Brazilian Savanna: Implications on physical and nutritional characteristics. **Agricultural Sciences**, v. 3, n. 5, p. 745–754, 2012.
- DINIZ-FILHO, J.A.F.; MELO, D.B.; OLIVEIRA, G.; COLLEVATTI, R.G.; SOARES, T.N.; NABOUT, J.C.; LIMA, J.D.S. ET AL. Planning for optimal conservation of geographical genetic variability within species. **Conservation Genetics**, v. 13, n. 4, p. 1085–1093, 2012.
- ELLSTRAND, N. C. Is gene flow the most important evolutionary force in plants? **American Journal of Botany**, v. 101, n. 5, p. 737–753, 2014.
- EPPERSON, B. K. **Geographical genetics**. Princeton University Press, 2003.
- FERREIRA, R. Q. DE S.; CAMARGO, M. O.; TEIXEIRA, P. R.; SOUZA, P. B. DE; VIANA, R. H. O. Uso potencial e síndromes de dispersão das espécies de três áreas de cerrado sensu stricto, Tocantins. **Global Science and Technology**, v. 9, n. 3, p. 73–86, 2017.
- FRAGUAS, R. M.; SIMÃO, A. A.; LEAL, R. S.; SANTOS, C. M. DOS; ROCHA, D. A.; TAVARES, T. S.; MARQUES, T. R.; ET AL. Chemical composition of processed baru (*Dipteryx alata* Vog.) almonds: Lyophilization and roasting. **African Journal of Agricultural Research**, v. 9, n. 53, p. 3838–3852, 2014
- HAVENS, K.; VITT, P.; MAUNDER, M.; GUERRANT, E. O.; DIXON, K. *Ex situ* plant conservation and beyond. **BioScience**, v. 56, n. 6, p. 525, 2006.
- HOBAN, S.; SCHLARBAUM, S. Optimal sampling of seeds from plant populations for ex-situ conservation of genetic biodiversity, considering realistic population structure. **Biological Conservation**, v. 177, p. 90–99, 2014.
- KIEREPKA, E. M.; LATCH, E. K. Performance of partial statistics in individual-based landscape genetics. **Molecular Ecology Resources**, v. 15, n. 3, p. 512–525, 2015.
- LEMOES, M. R. B.; SIQUEIRA, E. M. DE A.; ARRUDA, S. F.; ZAMBIAZI, R. C. The effect of roasting on the phenolic compounds and antioxidant potential of baru nuts [*Dipteryx alata* Vog.]. **Food Research International**, v. 48, n. 2, p. 592–597, 2012
- MANEL, S.; JOOST, S.; EPPERSON, B. K.; HOLDEREGGER, R.; STORFER, A.; ROSENBERG, M. S.; SCRIBNER, K. T.; ET AL. Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. **Molecular Ecology**, v. 19, p. 3760–3772, 2010.
- MANEL, S.; SCHWARTZ, M. K.; LUIKART, G.; TABERLET, P. Landscape genetics: Combining landscape ecology and population genetics. **Trends in Ecology and Evolution**, v. 18, n. 4, p. 189–197, 2003.
- MAXTED, N.; FORD-LLOYD, B. V.; HAWKES, J. G. **Plant genetic conservation: the *in situ* approach**. Springer Science & Business Media, 2013.
- MELO, D. B.; DINIZ-FILHO, J. A. F.; OLIVEIRA, G.; SANTANA, L. L.; SOARES, T. N.; CHAVES, L. J.; NAVES, R. V. ET AL. Optimizing sampling efforts for *ex situ* conservation of genetic variability of *Dipteryx alata* Vogel. **BMC Proceedings**, v. 5, n. Suppl 7, p. P18, 2011.
- MONTEIRO, S. M.; PINTO, A. M.; BARBOSA, A. P.; CNPQ-INPA, B. P.; CPST, O.

I.; CPST, C. I. **Fenologia de *Dipteryx alata* Vogel (Fabaceae) na Reserva Florestal Adolpho Ducke (Rfad) e Estação Experimental de Silvicultura Tropical (Eest) nos últimos dez anos.** 2013.

MOTA, E. E. S. **Caracterização fenotípica e variação genética quantitativa em *Dipteryx alata* Vog . ( Barueiro ) do Cerrado.,** 2013. UFG.

NABOUT, J. C.; SOARES, T. N.; DINIZ-FILHO, J. A. F.; MARCO JÚNIOR, P.; TELLES, M. P. C.; NAVES, R. V.; CHAVES, L. J. Combining multiple models to predict the geographical distribution of the Baru tree (*Dipteryx alata* Vogel) in the Brazilian Cerrado. **Brazilian journal of biology**, v. 70, n. 4, p. 911–919, 2010.

PEREIRA, S. R.; LAURA, V. A.; SOUZA, A. L. T. Establishment of Fabaceae tree species in a tropical pasture: influence of seed size and weeding methods. **Restoration Ecology**, v. 21, n. 1, p. 67–74, 2013.

PINELI, L. D. L. DE O.; CARVALHO, M. V. DE; AGUIAR, L. DE A.; OLIVEIRA, G. T.; CELESTINO, S. M. C.; BOTELHO, R. B. A.; CHIARELLO, M. D. Use of baru ( Brazilian almond ) waste from physical extraction of oil to produce flour and cookies. **LWT - Food Science and Technology**, v. 60, p. 50–55, 2015.

ROCHA, L. S.; SANTIAGO, A. C. Implicações nutricionais e sensoriais da polpa e casca de baru (*Dipteryx Alata* vog .) na elaboração de pães. **Ciência e Tecnologia de Alimentos**, v. 29, n. 4, p. 820–825, 2009.

SANO, S. M. **Crítérios de seleção de baru para produção de amêndoas e recomposição ambiental.** Circular técnica. Planaltina, DF, Embrapa. 2016.

SANO, S. M.; BRITO, M. APARECIDA DE; RIBEIRO, J. F. Baru. In: R. F. Vieira; J. Camillo; L. Coradin (Orgs.); **Espécies Nativas da Flora Brasileira de Valor Econômico Atual ou Potencial Plantas para o Futuro - Região Centro-Oeste.** p.1–1160, 2016. MMA Brasília - DF: Ministério do Meio Ambiente.

SANO, S. M.; RIBEIRO, J. F.; BRITO, M. A. **Baru: biologia e uso.** Embrapa cerrados -Brasilia DF, 2004.

SANTOS, A. M.; ROSADO, S. C. DA S.; OLIVEIRA, A. N. Estimation of genetic parameters and verification of early selection efficiency in baru (*Dipteryx alata*). **Crop Breeding and Applied Biotechnology**, v. 14, n. 4, p. 238–243, 2014.

SCARIOT, A. O.; SERVILHA, A. C. Conservação in situ de recursos genéticos vegetais. In: L. L. Nass (Org.); **Recursos genéticos vegetais.** 1a ed., v. 1, p.473–502, 2007. Brasília, DF: Nass, Luciano Lourenço.

SEGELBACHER, G.; CUSHMAN, S. A.; EPPERSON, B. K.; FORTIN, M. J.; FRANCOIS, O.; HARDY, O. J.; HOLDEREGGER, R. ET AL. Applications of landscape genetics in conservation biology: Concepts and challenges. **Conservation Genetics**, v. 11, n. 2, p. 375–385, 2010.

SILVA, D. S. N.; VENTURIN, N.; RODAS, C. L.; MACEDO, R. L. G.; VENTURIN, R. P.; MELO, L. A. DE. Growth and mineral nutrition of baru (*Dipteryx alata* Vogel) in nutrient solution. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 20, n. 12, p. 1101–1106, 2016.

SIQUEIRA, E. M. DE A.; MARIN, A. M. F.; CUNHA, M. DE S. B. DA; FUSTINONI, A. M.; SANT'ANA, L. P. DE; ARRUDA, S. F. Consumption of baru seeds [*Dipteryx*

- alata* Vog.], a Brazilian savanna nut, prevents iron-induced oxidative stress in rats. **Food Research International**, v. 45, n. 1, p. 427–433, 2012
- SOARES, T. N.; CHAVES, L. J.; TELLES, M. P. DE C.; DINIZ-FILHO, J. A. F.; RESENDE, L.V. Distribuição espacial da variabilidade genética intrapopulacional de *Dipteryx alata*. **Pesquisa Agropecuária Brasileira**, v. 43, n. 9, p. 1151–1158, 2008.
- SOARES, T.N.; DINIZ-FILHO, J.A.F.; NABOUT, J.C.; CAMPOS TELLES, M.P. DE; TERRIBILE, L.C.; CHAVES, L.J. Patterns of genetic variability in central and peripheral populations of *Dipteryx alata* (Fabaceae) in the Brazilian Cerrado. **Plant Systematics and Evolution**, v. 301, n. 5, p. 1315–1324, 2015.
- TAKEMOTO, E.; OKADA, I. A.; GARBELOTTI, M. L.; TAVARES, M.; AUED-PIMENTEL, S. Composição química da semente e do óleo de baru (*Dipteryx alata* Vog.) nativo do Município de Pirenópolis, Estado de Goiás. **Revista Instituto Adolfo Lutz**, v. 2, n. 60, p. 113–117, 2001.
- TAMBARUSSI, E. V, SEBBENN, A. M.; ALVES-PEREIRA, A.; VENCOSKY, R.; CAMBUIM, J.; SILVA, A. M. DA; DE MORAES, M.A.; MORAES, M. L.T. DE. *Dipteryx alata* Vogel (Fabaceae), a Neotropical tree with high levels of selfing : implications for conservation and breeding programs. **Annals of Forest Research**, v. 60, n. 2, 2017.
- TARAZI, R.; MORENO, M. A.; GANDARA, F. B.; FERRAZ, E. M.; MORAES, M. L. T.; VINSON, C. C.; CIAMPI, A. Y.; ET AL. High levels of genetic differentiation and selfing in the Brazilian cerrado fruit tree *Dipteryx alata* Vog. (Fabaceae). **Genetics and molecular biology**, v. 33, n. 1, p. 78–85, 2010.
- TELLES, M. P. DE C.; DOBROVOLSKI, R.; SOUZA, K. DA S. E.; LIMA, J. DE S.; COLLEVATTI, R. G.; SOARES, T. N.; CHAVES, L. J.; DINIZ-FILHO, J. A. F. Disentangling landscape effects on population genetic structure of a Neotropical savanna tree. **Natureza & Conservação**, v. 12, n. 1, p. 65–70, 2014.
- UNAMBA, C. I. N.; NAG, A.; SHARMA, R. K. Next Generation Sequencing Technologies: the doorway to the unexplored genomics of non-model plants. **Frontiers in Plant Science**, v. 6, n. December, 2015.
- WOLF, M. J. **Roedores como agentes secundários na remoção de endocarpos de baru” (*Dipteryx alata* Vogel) em borda e interior de fragmentos de cerrado**. 2015. Universidade Federal de Mato Grosso do Sul.
- ZALAPA, J. E.; CUEVAS, H.; ZHU, H.; STEFFAN, S.; SENALIK, D.; ZELDIN, E.; MCCOWN, B.; ET AL. Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. **American Journal of Botany**, v. 99, n. 2, p. 193–208, 2012.