UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ECOLOGIA E RECURSOS NATURAIS

SANDARA NADJA RODRIGUES BRASIL

VARIAÇÃO GENÉTICA, DINÂMICA POPULACIONAL E ADAPTAÇÃO DE *Xylocopa grisescens* Lepeletier EM PAISAGENS SEMIÁRIDAS DEGRADADAS

FORTALEZA
2021

SANDARA NADJA RODRIGUES BRASIL

VARIAÇÃO GENÉTICA, DINÂMICA POPULACIONAL E ADAPTAÇÃO DE *Xylocopa grisescens* Lepeletier EM PAISAGENS SEMIÁRIDAS DEGRADADAS

Tese de doutorado submetida ao Programa de Pós-Graduação em Ecologia e Recursos Naturais da Universidade Federal do Ceará, como requisito para obtenção do grau de Doutor em Ecologia e Recursos Naturais. Área de concentração: Ecologia e Recursos Naturais

Orientadora: Prof.ª Drª Francisca Soares de Araújo
Coorientadora: Dra. Christiana M. A. Faria
Coorientador: Danilo Boscolo

FORTALEZA

2021

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

B83v Brasil, Sandara Nadja Rodrigues.
Variação genética, dinâmica populacional e adaptação de Xylocopa grisescens Lepeletier em paisagens semi-áridas degradadas / Sandara Nadja Rodrigues Brasil. – 2021.
98 f. : il. color.

Tese (doutorado) – Universidade Federal do Ceará, Centro de Ciências, Programa de Pós-Graduação em Ecologia e Recursos Naturais , Fortaleza, 2021.
Orientação: Profa. Dra. Francisca Soares de Araujo.
Coorientação: Profa. Dra. Christiana M. A. Faria.

1. Abelhas. 2. Desertificação . 3. Dispersão. 4. Estrutura populacional. 5. Diversidade genética. I. Título.
CDD 577

SANDARA NADJA RODRIGUES BRASIL

VARIAÇÃO GENÉTICA, DINÂMICA POPULACIONAL E ADAPTAÇÃO DE *Xylocopa grisescens* Lepeletier EM PAISAGENS SEMIÁRIDAS DEGRADADAS

Tese apresentada ao Programa de Pós-Graduação em Ecologia e Recursos Naturais da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Doutor em Ecologia e Recursos Naturais. Área de concentração: Ecologia e Recursos Naturais.

Aprovada em: 07/09/2021

_____

Profa. Dra. Francisca Soares de Araujo (Orientadora)
Universidade Federal do Ceará (UFC)

_____

Dra. Margarita M López-Uribe
Pennsylvania State University (USA)

_____

Dra. Silvia Helena Sofia
Universidade Estadual de Londrina (UEL)

_____

Dra. Solange Cristina Augusto
Universidade Federal de Uberlândia (UFU)

_____

Dr. Thalles Barbosa Grangeiro
Universidade Federal do Ceará (UFC)

_____

Dra. Julia Caram Sfair (Suplente)
Universidade Federal do Ceará (UFC)

_____

Dr. Breno Magalhães Freitas (Suplente)
Universidade Federal do Ceará (UFC)

Ao meu pai João e à minha mãe Tereza.

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de financiamento 001. Agradeço ao financiamento obtido através do Programa Ciência sem Fronteira (Projeto: *Evolução da perda de biodiversidade em áreas sob processos de degradação),* ao Instituto Nacional de Ciência e Tecnologia em Estudos Interdisciplinares e Transdisciplinares em Ecologia e Evolução (INCT IN-TREE) e à Fulbright Scholarship pelo financiamento do período sanduíche na University of California, Riverside.

Agradeço à minha orientadora Francisca Soares de Araujo por me permitir a liberdade de escolher uma linha de investigação científica bastante diferente da sua, confiando na minha capacidade de formar nova linha de pesquisa em seu laboratório.

Agradeço aos meus co-orientadores Christiana M. A. Faria e Danilo Boscolo por terem me proporcionado oportunidades de crescimento profissional e inestimáveis conselhos. Também agradeço ao Prof. Dr. Lorenzo Zanette pela contribuição intelectual a esse trabalho.

Grande parte da pesquisa nesta tese não poderia ter ocorrido sem a colaboração da Dra. S. Hollis Woodard, minha supervisora durante o período de doutorado sanduíche, minha amiga e mentora profissional.

Tampouco poderia ter concluído este trabalho sem a assistência e a amizade dos alunos do programa de pós-graduação em Ecologia e Recursos Naturais.

Agradeço aos colegas do laboratório de Fitogeografia do qual fiz parte por mais de cinco anos, e aos colegas do Woodard Lab, por me receberem tão bem.

Acima de tudo, gostaria de agradecer ao meus pais João Brasil e Tereza Marly, ao Diego e a minha família pelo apoio e incentivo que me deram ao longo dos anos.

O correr da vida embrulha tudo. A vida é assim: esquenta e esfria, aperta e daí afrouxa, sossega e depois desinquieta. O que ela quer da gente é coragem (ROSA, 1956, p. 293)

RESUMO

Um dos principais objetivos em biologia da conservação é entender os efeitos do habitat em determinada espécie ou grupo de espécies. Apesar das preocupações globais com os efeitos negativos da desertificação sobre a biodiversidade, ainda são limitados os estudos que abordem tais efeitos ao mesmo tempo que considera a conservação genética de populações. Ambientes áridos e semiáridos são peculiares em sua relação com a conservação genética de populações, pois exibem escassez de recursos alimentares, de nidificação e reprodutivos, mas também propiciam subsídios a adaptação local. No Brasil, a região semiárida e subsumida do Nordeste é a mais suscetível à desertificação e possui um histórico de degradação ao longo de muitas décadas até os dias atuais. Levando-se em consideração que a degradação causada pelo homem afeta negativamente a diversidade genética de populações, e que habitats altamente heterogêneos desempenham um importante papel na diferenciação genética e na adaptação local; torna-se fundamental entender como esses ambientes moldaram geneticamente suas populações. Também, devido ao longo período em que o semiárido brasileiro sofreu com a degradação de suas paisagens, é importante acessar a história demográfica das populações a fim de identificar que mudanças (divisão, fusão, gargalos ou expansão) ocorreram nestas populações ao longo do tempo. Por isso, aqui investigamos como as paisagens extremamente degradadas do semiárido brasileiro, aqui retratadas como áreas susceptíveis a desertificação (ASD), vêm afetando a variabilidade genética, a dinâmica populacional e a adaptação de um importante polinizador do semiárido nordestino, a abelha *Xylocopa grisescens*. Para isso, coletamos oitenta fêmeas de *X. grisescens* em oito locais de amostragem no Estado do Ceará, no Nordeste brasileiro, abrangendo distâncias de até 300 km. Utilizamos RADseq (*Restriction site association DNA sequencing*) para identificar 83.127 SNPs (*Single Nucleotide Polymorphism* ou polimorfismo de nucleotídeo único). Acessamos a diversidade genética, a estrutura genética, a história demográfica, o fluxo gênico e possíveis barreiras ao fluxo gênico através de técnicas de genética molecular, bioinformática e análises estatísticas. Detectamos uma baixa quantidade de diversidade genética e diferenciação entre as populações, embora a distribuição dessa diversidade genética seja heterogênea entre os locais. Inferências sobre a história demográfica de *X. grisescens* revelaram que o tamanho efetivo atual da população é de cerca de ~5.000 indivíduos, e que mesmo com a redução da população há cerca de ~60 a ~80 anos, *X. grisescens* manteve um tamanho populacional viável para manter esta espécie prosperando. Também, os resultados sobre o fluxo gênico e adaptação revelaram que as populações são mais adaptadas às variáveis de precipitação do que temperatura e altitude, e que

o fluxo gênico não é restrito devido à grande concentração de pastagens e agricultura na região semiárida brasileira, que favorece a floração de espécies nativas e culturas preferidas por *X. grisescens*. Nossos dados sugerem que, apesar do processo de desertificação ocorrendo na região semiárida brasileira impactar de forma geral na biodiversidade, essas mudanças não estão afetando geneticamente a abelha *X. grisescens*.

Palavras-chave: abelhas; desertificação; dispersão; estrutura populacional; diversidade genética.

ABSTRACT

One of the main goals in Conservation Biology is to understand the habitat effects on a deem species or species group. Despite global concerns about the negative effects of desertification on biodiversity, studies that address these effects considering the genetic conservation of populations are still limited. Globally, arid and semi-arid environments have a unique characteristic regarding the genetic conservation of their populations due to the fact that at the same time, they exhibit food, nesting, and reproductive scarcity; they also provide adaptive components to certain species. In Brazil, the area most susceptible to desertification is found in the semi-arid and sub-humid region of the Northeast, which has a history of continuous degradation over many decades until the present day. Considering that highly heterogeneous habitats play a considerable role in achieving genetic differentiation and local adaptation and that anthropogenic degradation negatively affects the genetic diversity of populations, it is essential to understand how these environments shaped the populations genetically. Environments. Also, due to the long period in which the Brazilian semi-arid region suffered from the degradation of its landscapes, it is important to access the demographic history of populations in order to identify what changes (division, merger, bottlenecks, or expansion) have occurred in these populations over time. Therefore, here we investigate how the extremely degraded landscapes of the Brazilian semiarid region, here described as areas susceptible to desertification (ASD), have been affecting the genetic variability, population dynamics, and adaptation of an important pollinator of the semi-arid northeast, the bee *Xylocopa grisescens*. For this purpose, we collected 80 females of *X. grisescens* at eight sampling sites in the State of Ceará, in Northeastern Brazil, covering distances of up to 300 km. We used RADseq (Restriction site association DNA sequencing) to identify 83,127 SNPs (Single Nucleotide Polymorphism). We access genetic diversity, genetic structure, demographic history, gene flow, and possible barriers to gene flow in populations through molecular genetics, bioinformatics, and statistical analysis techniques. We detected a low amount of genetic diversity and differentiation between different populations, although the distribution of this genetic diversity is relatively heterogeneous across sites. Inferences about the demographic history of X. grisescens revealed the current effective population size of about ~5,000 individuals and the population reduction for about ~100 to ~150 years. *X. grisescens* maintained a viable population size to keep this species thriving in such degraded habitat. Also, the results on gene flow and

adaptation revealed that populations are more adapted to precipitation metrics than temperature or altitude and that gene flow is not restricted probably due to the high concentration of pastures and agriculture in the semi-arid region of Brazil, which favors the flowering of preferred native plant species and crops *X. grisescens*. Our data suggest that despite the ongoing desertification process greatly impacting the Brazilian semiarid region, these ecological changes are not genetically affecting the genetic variability of *X. grisescens*.

Keywords: bees; desertification; dispersal; population structure; population diversity

SUMÁRIO

CAPÍTULO I

Introdução Geral

1. Diversidade e estrutura genética em populações

Os indivíduos de uma espécie não são geneticamente iguais, cada um apresenta uma pequena dissimilaridade genética. A variação genética que existe em uma espécie é representada por sequências de DNA de um local específico do genoma (*locus*) que são ligeiramente diferentes para cada indivíduo (Frankham, 2013)). Isso resulta em variação no genoma individual chamada de polimorfismo genético (Frankham, 2013). Especificamente, o polimorfismo genético (ou diversidade genética) está relacionado ao potencial evolutivo de uma determinada espécie, pois contribui para a capacidade de responder às mudanças ambientais via seleção natural e para o sucesso reprodutivo (Frankham, 2013; Hoffmann e Willi, 2008).

O papel da diversidade genética na adequação às mudanças do habitat está associado aos diferentes alelos presentes nas populações (Segelbacher *et al.*, 2010) Por exemplo, em populações com muitos indivíduos, formas raras de um alelo (ou combinações de alelos) podem conferir vantagem ao se revelar adequadas a uma certa mudança nas condições ambientais (Hoffmann e Willi, 2008). Ou seja, a chance de uma população sobreviver a um evento estocástico em seu habitat natural depende diretamente das diferenças genéticas presentes em seus indivíduos. Ocorre que quando há diminuição no tamanho dessa população, os alelos raros são perdidos por meio de deriva gênica, e, com isso, também é perdido seu potencial adaptativo (Husemann *et al.*, 2016).

As causas mais alarmantes da perda dessa diversidade em populações estão relacionadas às mudanças no seu habitat (Balkenhol *et al.*, 2016). Em caso da perda de habitat natural, deve ocorrer diminuição de recursos alimentares, reprodutivos e de nidificação, além do provável isolamento espacial das populações (Fahrig, 2003; Steffan-Dewenter, 2002). Durante o processo de perda de habitat pode surgir a formação de barreiras que limitariam a dispersão e a colonização, afetando a dinâmica populacional (Broeck, Vanden *et al.*, 2017; Steffan-Dewenter, 2002). A longo prazo, os efeitos sentidos serão a endogamia, a diminuição no tamanho populacional e no tamanho populacional efetivo (Frankham, Ballou e Briscoe, 2010), a diminuição na fertilidade, na sobrevivência e nas taxas de crescimento e resistência a doenças (ver subitem 1.3). Similarmente, a estrutura genética de uma população também é influenciada por mudanças em seu habitat. A estrutura genética é definida como a distribuição espacial da diversidade genética encontrada em uma população (Balkenhol *et al.*, 2016), e guarda informações fundamentais sobre endocruzamento, deriva gênica, dispersão de indivíduos e traços de história de vida (Ballare e Jha, 2020; López-Uribe, Jha e Soro, 2019).

Parte fundamental dos estudos sobre estrutura genética é a quantidade de fluxo gênico temporal entre populações ao longo de muitas gerações (Slatkin, 1985a), sendo possível prever os padrões de colonização e dispersão de uma espécie (Broquet e Petit, 2009) e potencialmente revelar os fatores ambientais que facilitam ou impedem o movimento em escalas espaciais. No entanto, apesar dos resultados da diversidade e da estrutura genética nas populações parecerem óbvios, sabe-se relativamente pouco sobre o potencial que a variabilidade genética tem em determinar o funcionamento do ecossistema. e, consequentemente, em ajudar a planejar estratégias de conservação da biodiversidade sustentáveis em grandes escalas.

Tabela I.1. Terminologia utilizada *sensu* (Frankham, Ballou, & Briscoe, 2008)

Adaptação	O processo pelo qual uma população ou uma espécie global ou localmente se adapta ao seu ambiente. Refere-se ao estado atual de adaptação e ao processo evolutivo dinâmico que leva à adaptação
Alelo	Uma variante genética única observada em um locus específico.

Deriva genética	Mudanças na composição genética de uma população devido à retirada aleatória de indivíduos.
Diversidade genética	Refere-se a quantidade de variabilidade genética encontrada em uma população.
Efeitos estocásticos	Refere-se a efeitos determinados aleatoriamente.
Estrutura genética	Refere-se a distribuição da variabilidade genética entre populações ou indivíduos.
Fluxo gênico	No seu sentido mais amplo, é a transferência de alelos ou genes de uma entidade para outra (por exemplo, entre populações).
Loci polimórfico	Presença, em uma espécie, de dois ou mais alelos em um locus.
Locus (pl. loci)	É uma localização particular de um gene no genoma de um organismo.
Polimorfismo de nucleotídeo único (SNPs)	Variação genética em uma única posição em um genoma (geralmente para apenas dois alelos).
Variabilidade genética	Diferenças em sequências de DNA localizados no mesmo locus de um genoma. Engloba a diversidade genética e a estrutura genética.

2. Paisagem e Fragmentação de habitats

Locais com habitat estável ao longo do tempo podem preservar alta diversidade genética em populações (Telles *et al.*, 2014), enquanto que a fragmentação e a perda de habitat funcionam como suas principais ameaças antrópicas (Fischer e Lindenmayer, 2007; Segelbacher *et al.*, 2010). Dado que esses fatores se revelam principalmente em estudos sob

ampla escala espacial e que estão intimamente relacionados à crescente crise de extinção (McKinney e Lockwood, 1999), os estudos em escala de paisagem tem recebido merecido destaque nas últimas décadas (ver (Fischer e Lindenmayer, 2007).

O conceito de paisagem é bastante difuso entre diferentes autores. Alguns adotam o modelo de Biogeografia de Ilhas de (MacArthur e Wilson, 1967), outros o modelo de Metapopulações de (Levins, 1969), e ainda os que adotam outra miríade de modelos e conceitos. Aqui, adotaremos o conceito ecológico de paisagem, que, segundo (Metzger, 2001) é "um mosaico heterogêneo formado por unidades interativas, sendo que esta heterogeneidade existe para pelo menos um fator, segundo um observador e em uma determinada escala de observação" (Metzger, 2001)). Isso significa que estudos em escala de paisagem devem considerar a heterogeneidade de um determinado ambiente segundo um determinado fator ou observador, pois um habitat que parece heterogêneo para uma espécie, pode ser homogêneo para outra (Boscolo, 2016).

A heterogeneidade espacial é ocasionada pela quebra de continuidade da distribuição espacial de um determinado ambiente (Figura I.1), podendo afetar uma variedade de processos ecológicos (Boscolo, 2016; Fahrig, 2007). Na teoria, quanto mais heterogêneo um ambiente é, há maior quebra de conectividade, e maior é a variedade de tipos de vegetação que ele abriga (Fahrig e Nuttle, 2007). Em decorrência disso, espera-se que a quantidade de espécies e o tamanho populacional que um ambiente heterogêneo consiga abrigar também seja maior. Por outro lado, quando a quebra de continuidade é devida à fragmentação da paisagem, pode ocorrer a diminuição da conectividade entre os fragmentos fazendo com que a dispersão de migrantes se torne insustentável e não haja a colonização de novos locais. Se a subdivisão desses fragmentos ocasionar a quebra de conectividade entre os fragmentos de uma paisagem, ocorre a *fragmentação de habitat* (*sensu* Fahrig, 2003), grifo nosso).

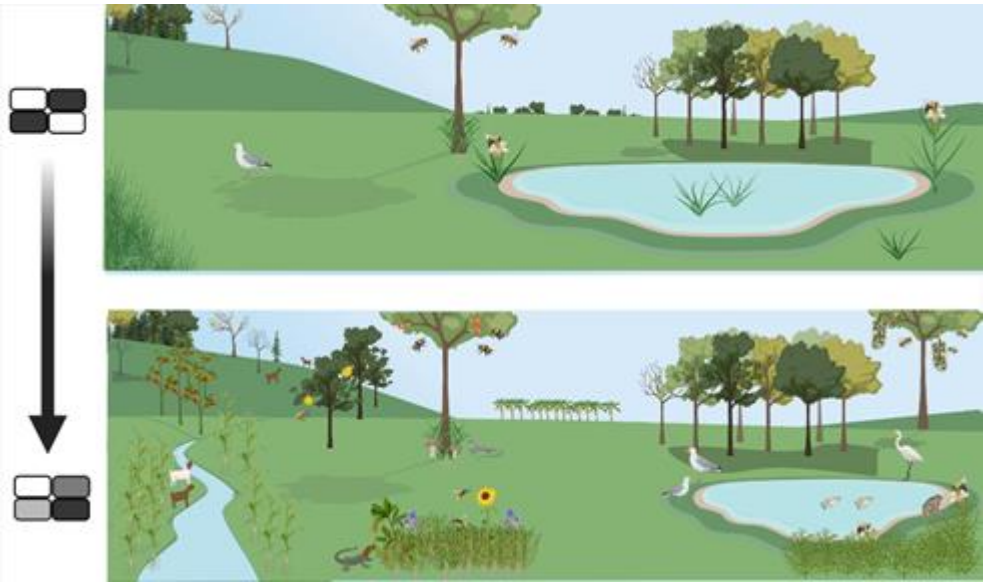


Figura I.1. Desenho esquemático dos efeitos do aumento da heterogeneidade em paisagens. Quanto mais heterogêneo um ambiente, maior a variedade de tipos de vegetação e maior a quantidade de nichos para espécies.

Comumente, o processo de fragmentação e perda de habitat (redução na área de habitat de uma paisagem, (Martin, Bennett, & Fahrig, 2021), se deve ao histórico de uso da terra, onde a retirada da cobertura vegetal é intensa. As maiores fontes contemporâneas de perda de habitat e fragmentação são a agricultura intensiva, a pecuária, a urbanização e a construção de estradas (Martin, Bennett e Fahrig, 2021). Ainda assim, é importante ressaltar que o efeito da fragmentação *per se* (sem perda de habitat) pode ser nulo ou benéfico para algumas espécies, em especial àquelas capazes de se dispersar por longas distâncias (ver (Fahrig, 2017). Aqui, utilizaremos o termo *fragmentação* para designar fragmentação + perda de habitat (Fahrig, 2017; Fahrig *et al.*, 2019; Martin, Bennett e Fahrig, 2021). Então, identificar a quantidade de habitat disponível para as populações é fundamental para conhecer os riscos de tais populações à fragmentação.

3. Consequências da fragmentação na dinâmica populacional e na diversidade genética.

A estrutura e a composição da paisagem são os principais motores da distribuição da população, adaptação local e fluxo gênico (Cushman, H. Mcrae e Mcgarigal, 2015). Se, no

processo de fragmentação, a conectividade for perdida ou diminuída, a dispersão dos indivíduos ficará restrita, o que pode acarretar a restrição do fluxo gênico entre pares de subpopulações ou entre populações (Freiria *et al.*, 2012). O fluxo gênico é um processo natural responsável pela troca de genes alelos entre populações (Slatkin, 1985a), e precisa de migração com subsequente reprodução para efetivamente ocorrer (Lenormand, 2002; Waits e Storfer, 2015). Barreiras ao fluxo gênico podem levar ao cruzamento de indivíduos muito aparentados (endogamia), e à consequente perda de diversidade genética (Waits e Storfer, 2015). Por outro lado, se não houver barreiras ao fluxo gênico, as chances de cruzamento aleatório aumentam, aumentando também o número de alelos com informação adaptativa. Por essa razão, indivíduos que conseguem se dispersar livremente pela paisagem têm mais chance de sobrevivência e mais chance de carregar "bons alelos", aumentando as chances de sucesso na colonização e manutenção de uma população (Slatkin, 1985b).

Além do exposto, outra consequência negativa da fragmentação de habitats é a redução do tamanho populacional efetivo (N_e). O N_e é descrito como o número de indivíduos capazes de se reproduzir e contribuir efetivamente para a próxima geração (Husemann *et al.*, 2016). Então, a capacidade de dispersão de indivíduos é fundamental para manter o tamanho populacional ideal e a persistência da população em determinado habitat (Lenormand, 2002). Com o isolamento, as cargas de alelos deletérios raros que existem naturalmente nas populações são facilmente disseminadas pelo excesso de endocruzamento, desfavorecendo o crescimento populacional e levando ao declínio acelerado em direção à extinção (Charlesworth e Willis, 2009; Frankham, 2019; Frankham, Ballou e Briscoe, 2008; Sexton, Strauss e Rice, 2011). Também, a população se tornaria mais homogeneizada, aumentando o risco de extinção local devido a efeitos estocásticos (Fischer e Lindenmayer, 2007; Franzén e Nilsson, 2014).

Então, qual o tamanho aceitável de uma população para que não haja a chance de extinção? Esse é um debate central em genética da conservação, pois a diminuição de 20% do tamanho populacional ou a quantidade de indivíduos menor que 1000 já categoriza uma população como vulnerável à extinção (IUCN, 2021). (Franklin e Frankham, 1998) sugerem que o tamanho populacional mínimo viável (MVP, do inglês *minimum viable population*) seja da ordem de 5000-12.500 para manter o potencial evolutivo a longo prazo. Mas como monitorar tantos indivíduos em uma população? Atualmente, é possível utilizar técnicas de biologia molecular para acessar essas informações a partir do genoma ou parte do genoma dos indivíduos.

4. Métodos de genômica populacional para medir os efeitos da fragmentação/perda de habitat

Buscando entender a relação entre a degradação da paisagem e os processos microevolutivos, (Manel, Schwartz, Luikart, & Taberlet, 2003) propuseram o termo *Genética de Paisagem*, que combina dados de ecologia de paisagens, biologia molecular e análise espacial, com foco nos efeitos da composição e estrutura da paisagem sobre o fluxo gênico. Durante muitos anos, essa relação foi estudada com a utilização de Microssatélites (*SSR, Simple Sequence Repeats*), marcadores moleculares com dezenas a centenas de *loci* polimórficos e que se tornaram bastante populares por detectarem diversidade genética entre indivíduos dentro de populações (Manel e Holderegger, 2013).

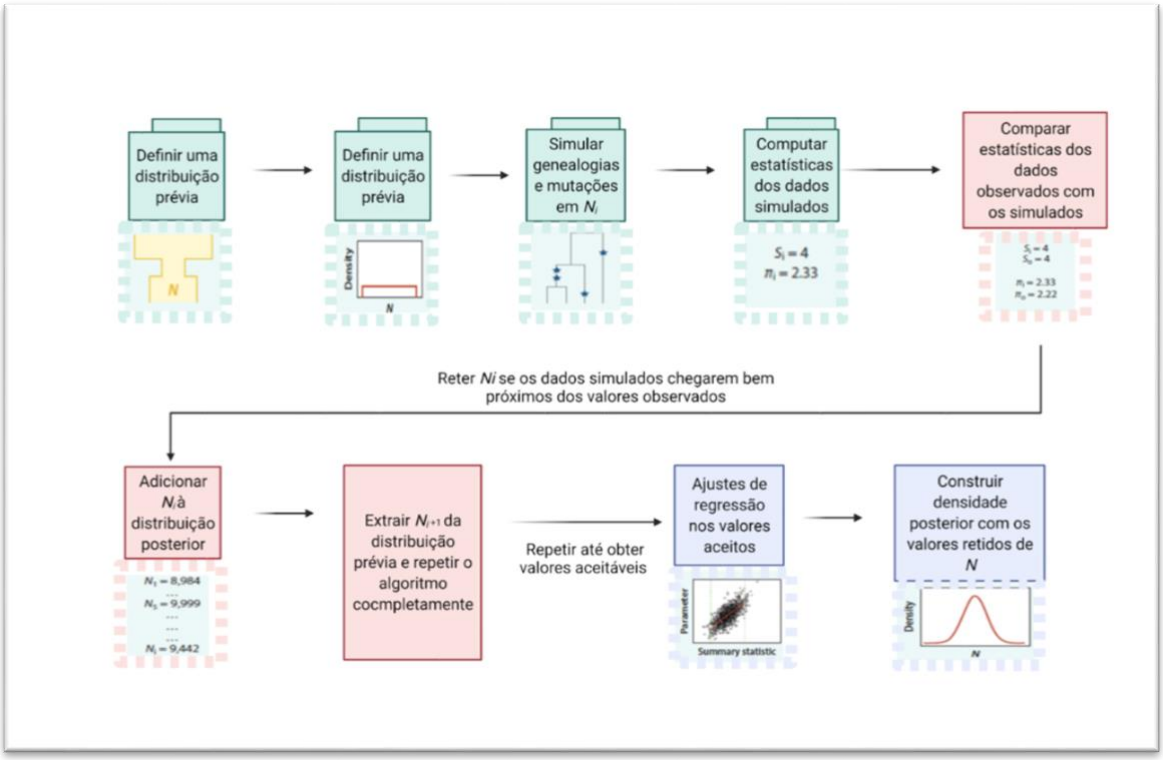
A abordagem mais recente envolve o sequenciamento de parte do genoma completo de espécies-alvo como uma forma mais robusta de medir a diversidade genética de populações (Storfer, Patton e Fraik, 2018). A técnica de RADseq (do inglês *Restriction site-associated DNA sequencing*, (Andrews *et al.*, 2016; Peterson *et al.*, 2012) utiliza milhares ou milhões de marcadores polimórficos (*SNPs, Single Nucleotide Polimorphism*) aleatoriamente distribuídos no genoma tornando possível aprofundar o conhecimento sobre processos evolutivos (Borevitz *et al.*, 2015). Essa abordagem ficou conhecida como *Genômica de Paisagem* (Storfer, Patton e Fraik, 2018). Essa abordagem também possibilita identificar *loci* adaptativos sob seleção, que potencialmente revelam seleção ambiental. Chamamos esses *loci* de *outlier loci*, que são primeiro detectados com posterior teste da associação com variáveis ambientais (Li *et al.*, 2017).

É possível também avaliar o grau de isolamento (i.e. perda de fluxo gênico) de populações em ambientes fragmentados através dos níveis de diferenciação genética (Waits e Storfer, 2015). Se o fluxo gênico é restrito, alelos neutros podem exibir altos níveis de diferenciação entre as populações, levando ao Isolamento por Distância Geográfica (*Isolation by distance, IBD*, (Wright, 1943). IBD descreve o padrão no qual a diferenciação genética aumenta com a distância geográfica e é a ferramenta mais amplamente utilizada em estudos de diversidade genética em populações e em estudos de genética de paisagem por incorporar dados em escala de paisagem (Shah e McRae, 2008).

Ainda, é possível avaliar o fluxo gênico com base no número de alelos privados em uma população (nPa, do inglês *number of private allele*). Indivíduos que experienciaram recente

redução no fornecimento de alimentos e recursos de nidificação em seu habitat natural, podem ser forçados a procurar habitats mais adequados ou lidar com a escassez de acasalamento (Slatkin e Takahata, 1985). Essa migração pode ser revelada pelo nPa, que é uma medida de quantos alelos únicos uma determinada população carrega (Slatkin, 1985b). Valores mais altos de nPa significam alto fluxo gênico, pois tem que haver dispersão frequente de diferentes indivíduos para que alelos raros se espalhem (Slatkin 1985).

Por fim, as análises demográficas são importantes para detectar possíveis variações nas taxas de crescimento, morte, razão sexual e tamanho populacional (Beaumont, 2005, 2010). A Aproximação Bayesiana Computacional (ABC) tem se mostrado eficaz em identificar possível expansão e contração demográfica nas populações, além de estimar o tamanho populacional atual (Beaumont, 2010). Na prática, a ABC funciona em três etapas: 1) há a geração dos dados observados (*loci* polimórficos), 2) geração de dados simulados com parâmetros populacionais *a priori* (quando não se sabe), que serão comparados aos dados observados e, 3) a estimativa de distribuições posteriores dos parâmetros populacionais através de regressão linear ou logística e a comparação de possíveis modelos (ou cenários) que explicam os dados observados (Figura I.2). Tais modelos são determinados em forma de eventos, podendo ser eventos de separação, desaparecimento e de mesclagem ou de mistura (Beichman, Huerta-Sanchez e Lohmueller, 2018; Cornuet *et al.*, 2014)



Adaptado de (Beichman, Huerta-Sanchez e Lohmueller, 2018)

Figura I. 2. Workflow para uma abordagem aproximada de computação baiana (ABC) para inferência demográfica. Aqui, desejamos inferir N, o tamanho das populações atuais no modelo de gargalo inferior.

5. *O semiárido brasileiro*

No Brasil, a área mais suscetível à desertificação é a região semiárida e sub-úmida do Nordeste, representando mais de 1.300.000 km² e comportando quase 35 milhões de habitantes (CGEE, 2016). A região semiárida brasileira (BSA) está entre as regiões semiáridas mais degradadas mais povoadas do mundo (Silva, da, Leal e Tabarelli, 2018). Historicamente, a vegetação do BSA sofre impactos crônicos desde a colonização brasileira pelos europeus no século XVI e aumentou durante a alta migração humana nos séculos XVIII e XIX (Silva, da, Leal e Tabarelli, 2018). A agricultura tem sido praticada de forma itinerante ou migratória: o agricultor desmata, queima, utiliza a área por alguns anos e depois deixa em pousio (tempo sem utilizar a área) para a recuperação da capacidade produtiva, enquanto inicia novo ciclo em outro local (Nunes, Araújo-Filho e Menezes, 2006). Com o tempo, a degradação induzida pelo

homem, como pecuária, extrativismo de madeira e agricultura de corte e queima foram potencialmente responsáveis pela redução da quantidade de habitat e da fragmentação do BSA (Silva, da, Leal e Tabarelli, 2018).

A consequência desse histórico de uso do BSA é um mosaico de vegetação com fragmentos de tamanhos, idades, níveis de distúrbios e conectividade diferentes. Essa mudança no habitat é especialmente elevada para o estado do Ceará, que tem a quase totalidade da sua área suscetível à desertificação (~ 148.000 km²; (CGEE, 2016). O estado do Ceará deve seu alto risco de degradação à aridez e em parte por ser o local onde as primeiras aldeias do período colonial foram estabelecidas (Jucá Neto, 2009). Devido a essas vulnerabilidades ambientais, áreas no estado do Ceará foram declaradas prioritárias para a conservação da biodiversidade (CGEE, 2016; Tomasella *et al.*, 2018). Três grandes áreas que são mais suscetíveis à desertificação no Ceará (doravante denominadas ASDs) foram delimitadas para monitoramento do uso da terra a fim de compreender os efeitos da desertificação e para prever as respostas de espécies-chave nesta região (CGEE, 2016).

6. Abelhas como modelo de estudo

Para animais que vivem em ambientes áridos e semiáridos, a mobilidade é fundamental, pois é através do movimento que conseguem forragear por longas distâncias em busca de água e recursos alimentares distribuídos esparsamente através da paisagem (Hobbs *et al.*, 2006). Também, considerando que regiões áridas e semiáridas têm condições climáticas extremas e ambientes com certa escassez de recursos, espera-se certo grau de adaptação às altas temperaturas e ao regime de baixa precipitação em populações persistentes (Willmer e Stone, 1997).

Os ambientes áridos e semiáridos contemplam a maior diversidade de abelhas (Hymenoptera: Apoidea) no nível global (Michener, 2007; Orr *et al.*, 2020). Abelhas representam um grupo monofilético estimado em 20.000 espécies distribuídas globalmente, extremamente dependentes de recursos florais (pólen e néctar) e de locais específicos para nidificação (Michener, 2007). Por essa razão, são considerados organismos bastante sensíveis às mudanças em seu habitat natural, podendo ser usadas como bioindicadores da degradação e como modelo de adaptação ao ambiente. Especificamente para o BSA, abelhas solitárias ou facultativas desempenham papel fundamental na manutenção dos serviços ecossistêmicos,

sendo consideradas polinizadores efetivos de uma variedade de espécies vegetais nativas e cultivadas (Klein *et al.*, 2020)

Dentre as abelhas que ocorrem no semiárido brasileiro (BSA), *Xylocopa grisescens* (Lepeletier) tem ampla distribuição geográfica e poliniza inúmeras espécies de plantas nativas e cultivadas, sendo vital para a manutenção da flora da vegetação dominante na região, a Caatinga. Alguns exemplos de plantas nativas polinizadas por *X. grisescens* são: *Poincianella pyramidalis, Capparis yco, Senna spectabilis, Solanum paniculatum, Eugenia rosea, Ziziphus cotinifolia, Melochia tomentosa* (Aguiar 2003) e importantes espécies cultivadas são: feijão-caupí (*Vigna unguiculata*), canavalia (*Canavalia ensiformis*), castanha do Pará (*Bertholletia excelsa*), Abóbora (*Cucurbita pepo e C. moschata*), Chuchu (*Sechium edule*), Goiaba (*Psidium guajava*), Melancia (*Citrullus lanatus*), Tomate (*Solanum lycopersicum*) e maracujá (*Passiflora spp*.) (Silva, da e Freitas, 2018). Especificamente, a polinização do maracujá requer a polinização por grandes abelhas como *Xylocopa* por ser uma espécie que apresenta proterandria e autoincompatibilidade (Freitas e Oliveira Filho, 2003; Giannini *et al.*, 2017; Silva, da e Freitas, 2018; Talles Marques, Chaves-Alves; Solange Cristina, 2011). É importante ressaltar que o Brasil é o maior produtor mundial do maracujá amarelo (*Passiflora edulis* F. *flavicarpa*, (USAID, 2014), com importância econômica destacada no Nordeste semiárido brasileiro, onde a vegetação de Caatinga é dominante. Apesar de ser reconhecidamente importante tanto para culturas agrícolas, quanto para a polinização da vegetação nativa, *X. grisescens* se encontra ameaçada devido às ações antropogênicas que vêm diminuindo seu habitat original. Estudo recente revelou que a espécie pode sofrer uma diminuição de até 35% de habitat adequado nas próximas décadas (Bezerra *et al.*, 2019).

Abelhas do gênero *Xylocopa* apresentam importantes ferramentas adaptativas a ambientes áridos e semiáridos como termorregulação e diapausa (Chappell, 1982; Michener, 2007). Da mesma forma, o corpo grande dessas abelhas auxilia na maior amplitude de voo e na maior chance encontrar recursos. Se por um lado, é importante reconhecer que abelhas grandes possuem certa adaptação à limitação de recursos, também é importante lembrar que abelhas grandes precisam de grande áreas para encontrarem a quantidade mínima de recursos e condições necessários para conseguirem manter seu metabolismo. Por isso, se torna tão importante avaliar tais respostas adaptativas nesses organismos. Mas o que significa dizer que um organismo é adaptado a um dado ambiente? Quer dizer que um dado ambiente gerou forças seletivas que moldaram as futuras gerações desse organismo fazendo com que sobrevivessem os genes mais adaptados (Darwin, 1859). Isso significa que a palavra adaptação revela não

somente os padrões observados hoje, mas a história evolutiva de um organismo. Na prática, a aparente aptidão aos ambientes em que vivem atualmente se dá apenas porque os ambientes atuais tendem a ser semelhantes aos do passado.

Devido à vasta literatura acerca da diminuição de habitat impactando negativamente a biodiversidade no semiárido brasileiro, e devido ao fato de abelhas serem extremamente sensíveis às mudanças do habitat, o objetivo desse trabalho foi acessar a potencial perda de variabilidade genética, possíveis diminuição no tamanho das populações e adaptações em *X. grisescens* vivendo em áreas extremamente degradadas do semiárido nordestino do Brasil. Hipotetizamos que: 1) As populações de *X. grisescens* apresentam baixa diversidade genética e alta estrutura genética, inferindo perda de fluxo gênico devido a possível perda de conectividade relacionada às características da paisagem; e 2) *X. grisescens* possui *loci* que estão sob seleção, o número de alelos privados são baixos, mas o número de alelos adaptativos é alto, inferindo adaptações evolutivas ao clima semiárido 3) Como reflexo da perda de habitat e diminuição dos recursos alimentares e de nidificação, as populações tiveram seu tamanho efetivo reduzido.

Em resumo, o objetivo geral da tese foi entender a influência de paisagens extremamente degradadas na variação genética de *Xylocopa grisescens*, e discutir essa relação com base na conservação genética de espécies-chave para a flora regional. Para tal fim, coletamos indivíduos dessa abelha focal em oito paisagens contemplando um gradiente de degradação em uma área do semiárido Cearense. Avaliamos a diversidade genética e a estrutura genética dessas populações, inferimos a história demográfica e identificamos possíveis sinais de adaptação ao clima semiárido. A tese está estruturada em 6 capítulos: o capítulo 1, que foi apresentado acima, de introdução/revisão teórica que apresenta e justifica a importância de medir a variação genética (diversidade e estrutura) em conjunto com as mudanças na paisagem para a conservação biológica. O capítulo 2 revisa as adaptações morfológicas, fisiológicas e comportamentais de abelhas em ambientes áridos e semiáridos, bem como as possíveis consequências da degradação desses ambientes para as abelhas. O capítulo 3, explora o status genômico das populações de *X. grisescens* levando em consideração possíveis isolamentos populacionais e barreiras ao fluxo gênico. O capítulo 4 utiliza os dados gerados a partir de simulações com *Aproximate Bayesian Computation* (ABC) para inferir a história demográfica dessas abelhas, incluindo possíveis gargalos populacionais, reduções ou extensões do tamanho efetivo populacional e tamanho populacional atual. O capítulo 5 utiliza parte dos dados dos capítulos 3 e 4 para avaliar diretamente barreiras ao fluxo gênico com base no número de alelos

privados de cada população e na composição da paisagem. Também identifica loci putativamente sob seleção e a relação desses loci com variáveis ambientais locais. O capítulo 6 apresenta as considerações finais e as conclusões gerais da tese, além de apresentar ações futuras para a conservação de abelhas.

CAPÍTULO II

Desert and semi-desert bees: adaptations, ecological interactions, and uncertain futures.

A pesquisa feita neste capítulo representa uma colaboração entre a autora, Dr. Michael Orr, Dra. Francisca Soares de Araújo e a Dra. S. Hollis Woodard.

Revista de interesse: Apidologie | ISSN 1297-9678 (online) | ISSN 0044-8435 (Print)

Esse capítulo é um artigo de revisão que retrata o atual entendimento sobre a biologia das abelhas de ambientes desérticos e semidesérticos, promovendo uma visão geral da pesquisa em evolução adaptativa que tem permitido essas abelhas sobreviverem a ambientes de extremas condições climáticas. Realçamos ainda a importância na manutenção dos ecossistemas desérticos e semidesérticos incluindo a polinização de plantas nativas e culturas agrícolas, e discutimos como a desertificação e outras ameaças atuam na manutenção das populações desses ambientes.

Abstract

Deserts have among the richest bee communities in the world being able to support speciation and adaptation. While a wide variety of adaptations have been identified in desert-dwelling bees, including physiological, morphological, and behavioral strategies, no study has compiled all this information. This review highlights the current knowledge about desert and semideserts bees with a detailed description of their biology and evolutionary adaptations that have allowed bees to address the biotic and abiotic challenges of desert life, with an emphasis on thermoregulation and synchronization with floral food resources. Finally, we use existing knowledge about their biology to explore how they might be impacted by desertification, a growing threat in the Anthropocene.

Keywords: *Desertification; Pollination; Semiarid; Synchronization; Thermoregulation.*

Introduction

Deserts and semi-deserts are among the most bee biodiverse ecosystems in the globe and contain an abundance of species that are uniquely adapted to the harsh conditions of xeric habitats (Michener, 2007). One might predict that the characteristic of xeric ecosystems, such as very low precipitation, high daily temperatures, and solar radiation, and the unpredictable and often sparse spatial and temporal patterns of floral resource availability (Whitford, 2002), would inhibit survival, especially of strongly interdependent mutualists. However, global bee surveys consistently agree that deserts have among the richest bee communities in the world (Michener, 2007; Minckley, 2014; Orr *et al.*, 2020).

Records from the Southwestern USA, Mediterranean Basin, and Australia provide important information on bee richness (Orr *et al.*, 2020). For example, in the northwestern Chihuahuan desert, bee richness reaches >500 species (Minckley e Ascher, 2013), with a recent survey showing the remarkable data of 473 bee species only in San Bernadino County (Minckley e Radke, 2021). Other spots in North America that have reported extreme bee richness are Utah's Grand Staircase Escalante National Monument, which is only 7,689 Km^2, and identified > 650 species (Carril *et al.*, 2018); Pinnacles National Park in California, which houses ~450 bee species within a modest 109 km^2 (Meiners, Griswold e Carril, 2019), Palm Springs (more than 500 species, (Michener, 1979). As for semiarid areas, the Mediterranean basin also houses significant diversity, representing 20% of the world's floristic richness (Médail e Quézel, 1999). Also, semiarid areas in Israel (Orr *et al.*, 2020), Riverside County in

California (~400 species, (Linsley, 1958), the Brazilian semiarid region (270, unpublished data), and central Chile, Argentina, much of Australia, and western parts of southern Africa (Michener 2007).

A key to understanding why bees are so abundant in deserts and semideserts might be found by exploring their various adaptations to desert life. A wide variety of adaptations have been identified in desert-dwelling bees, including physiological, morphological, and behavioral strategies. Behavioral strategies often rely on avoiding overheating or excessive cooling by positioning, flying or moving, and nesting in unusual substrates that may appear unfavorable (Danforth, n.d.; Heinrich & Esch, 1994; Michener, 2007; Orr, Parker, & Woodard, 2018). Physiological and morphological adaptations may include changes in plumosity, fur coloration, or body size (Heinrich e Esch, 1994; Michener, 2007; Orr *et al.*, 2018). Other studies have examined biotic relations responsible for maintaining bee biodiversity in xeric areas (Minckley, Cane e Kervin, 2000). Flowering events in deserts are often triggered by rainfall and consequently, bees that also coordinate their emergence or activity with humidity cues, and spend time in a state of developmental arrest when conditions are unfavorable, are probably better able to persist in these environments (Danforth 1999; Minckley et al. 2000; Orr et al. 2016). Floral specialization is also another strategy prevalent amongst desert bees and may contribute to their diversification and success (Wcislo 1996). Understanding adaptations in these bees is also important because globally, many parts of the world are warming and drying (Maestre *et al.*, 2012), and physiological and other traits are integral for conservation efforts (Seebacher e Franklin, 2012). The effects these changes will have on these bees are currently unknown but might be better predicted in light of their unique biology.

In this review, we discuss our current understanding of desert and semi-desert bee biology. We provide an overview of research on evolutionary adaptations that have allowed bees to address the biotic and abiotic challenges of desert life, with an emphasis on thermoregulation and synchronization with floral food resources. Throughout, we highlight the importance of bees to xeric ecosystems, including native plant populations. Finally, we use existing knowledge about their biology to explore how bees might be impacted by land degradation and desertification, a growing threat in the Anthropocene (Andrew *et al.*, 2013).

Unique adaptions in desert and semideserts bees

Xeric habitats (here including desert and semidesert regions) pose many challenges to small, highly mobile, largely diurnal insects. These environments are typified by their lack of precipitation (between 250 and 800 mm), aridity index between 0.05 and 0.50, and high daytime temperatures (e.g., up to 56°C in Death Valley, CA, with averages of 38°C, WorldClimate 2019). Thus, desert-dwelling animals must be able to withstand this lack of water and extremely high temperatures (Ward 2009). Willmer et al. (2009) classified animals according to two strategies to deal with extreme temperatures: evaders and endurers. Bees are generally classified as *evaders*, which refers to animals that prevent overheating of their body by using behavioral and other strategies related to active thermoregulation, although some species perform as *endurers* and can more directly tolerate extreme temperatures via enhanced physiological tolerance or other mechanisms.

Behavioral thermoregulation helps desert bees to prevent overheating. One way that this is achieved is by changing foraging activity patterns, primarily to avoid flight during the hottest parts of the day (Linsley 1978; reviewed in Willmer and Stone 1997, table II.1). Crepuscular or matinal flights may also provide additional competitive advantages for bees when they arrive early at flowers that have higher amounts of nectar and pollen earlier in the day (Linsley and Cazier 1970; Bloch et al. 2017). An interesting example is the matinal behavior of *Ptiloglossa arizonensis* foraging on *Solanum* flowers. Before dawn, those bees arrive at flowers and begin to force their way into the still-closed flowers in order to extract pollen, and during several minutes while the first sunlight comes, they collect pollen frantically while bumping any competitor from the flowers. By the time the sun rise, activity is essentially over, suggesting that pollen was completely consumed (Linsley and Cazier 1970). A similar pattern occurs in *Peponapis*, which gathers pollen exclusively from *Cucurbita* flowers from sunrise until the flowers wilt or until pollen is completely consumed (Linsley 1958). In addition to the timing of flight activity, bees might also position themselves in cooler microhabitats during periods of activity to achieve temporary cooling.

Table II.1. Behavioral adaptations of desert and semideserts bees

Genera/Species	Activity	Advantage	Authors

Xylocopa (Proxylocopa) olivieri	Crepuscular activity	Distinct temporal niche during times of low solar intensity, and possibility to allocate most of the daytime to nest construction and to guard	Gottlieb et al. 2005
*Perdita (*subgenus *Xerophasma*	Nocturnal species with enlarged ocelli	Collect pollen from the evening flowers. An example is the primrose *Oenothera*	Linsley 1958
Ptiloglossa and *Caupolicana*	Foraging before sunrise (19-20°C)	Distinct temporal niche if in optimal temperatures	Linsley 1978
Andrena (Diandrena) chalybaea	Delayed foraging flight	Avoid anthesis mismatching in host plants during overcast days	Thorp 1969; Estes and Thorp 1975
Lasioglossum (*Hemihalictus) lustrans*	Delayed foraging flight. Gathers pollen almost exclusively from *Pyrrhopappus* flowers	Avoid anthesis mismatching in host plants during overcast days and overlapping foraging activity within the period where flowers are open	Thorp 1969; Estes and Thorp 1975. Linsey 1978
Anthophora pauperata	Bimodal activity pattern (morning and afternoon foraging activity)	Wider thermal window in which can forage	Stone et al. 1999

Because of their relatively small body sizes, bees have a large surface area to volume ratio, which exposes more of their body surface to radiation but also allows for greater convective cooling, especially during flight (Heinrich and Esch 1994; Corbet and Huang 2016). As desert areas tend to contain more sparsely distributed resources for bees, such as water, food,

and nesting habitats, bees may be forced to fly long distances to find these resources, particularly when host plants may be rare or patchily distributed. Hence, physiological thermoregulatory mechanisms related to flight are crucial for maintaining sustained desert bee flight activity. This phenomenon can be seen especially in bees of the genus *Anthophora pauperata,* a desert solitary bee able to regulate thoracic temperatures during flight and capable of considerable endothermic heat generation independent of external sources (Stone et al. 1999). Many bees are also able to control their temperature by alternative means related to flight. For example, large bees of the genus *Xylocopa* dwelling in hot deserts use convective cooling from their anteriorly-flattened heads, which lose heat faster than the thoraces or abdomens, resulting in a six-fold increase in heat loss (Heinrich; Esch 1994).

Chappel (1982) stated that large bees face additional challenges in thermoregulating because of their high muscular activity required for flight, although carpenter bees well overcome these challenges by virtue of high thermal conductance and ability to tolerate high body temperature, being able to forage until they reach 44°C. Some desert bees may thus prevent overheating by making cooling flights (Willmer and Stone 1997; Corbet and Huang 2016). Bees dwelling in hot areas tend to be fast flyers and may use this high flight speed to avoid overheating, even enabling flight during peak heat midday (Orr et al. 2018). This mechanism is made possible in part through metabolic water gains during flight, which compensate for evaporative water losses, maintaining water balance during flight (Willmer and Stone 1997). This phenomenon has been shown in honeybees (*Apis mellifera*) under moderate-desert simulated conditions (Louw and Hadley 1985) but requires further investigation regarding desert-adapted bees.

Given that water balance is crucial for maintaining homeostasis despite high temperatures, individuals or colonies experiencing hot, warm weather are likely to face physiological constraints due to water scarcity. Solitary bees do not forage on water, and so they must overcome water loss by obtaining all water needed from nectar collecting. This strategy is relatively under-studied, although Willmer (1986) found a decrease in hemolymph sugar concentrations after nectar ingestion when studying the semi-desert bee *Megachile* (*Chalicodoma*) *sicula,* suggesting that these bees osmoregulate through nectar-derived water. It is unknown how bees are impacted by these changes in floral sugar concentration and whether they might preferentially forage on nectars that are lower or higher in sugar concentration to meet the conflicting demands of energetics and water balance.

There are also unusual aspects of the nesting biology of desert bees. Nectar and pollen are imperative for the growth and development of bee larvae and are also consumed by adult bees (Michener 2007; Willmer and Stone 2004). The overwhelming majority of bees (including those dwelling in deserts) are ground-nesters, who store their floral provision in cells excavated in soil and often thinly lined with secreted waxy or cellophane-like material (Michener 2007). Orr et al. (2016) found a new species of bee that excavates its nests in sandstone. This desert bee, named *Anthophora (Anthophoroides) pueblo* Orr, nests in sandstone throughout the Southwest, which is believed to minimize mortality due to moisture, fungi, and pathogens (Orr et al. 2016).

Desert bee and plant interactions

Whereas abiotic challenges are important for shaping the diversity and evolution of desert bees over evolutionary time scales, biotic relations are responsible for maintaining bee biodiversity in these areas. Plant-pollinator relations are mostly specific in desert areas, with these floral specialists often reliant on just one plant group or even individual plant species. (e.g., *Anthophora abroniae* on *Abronia villosa*, Orr et al. 2018). Floral specialization is believed to evolve in systems where the use of a single resource is more efficient or otherwise beneficial to an organism than a more generalist foraging strategy (Minckley et al., 2000; Waser and Ollerton 2006). The improved ability to reliably track specific resources, demonstrated by specialists (Minckley et al., 2013), is especially beneficial in extreme environments, where floral resources may be less reliable across years.

One of the primary ways that desert and semi-desert bees track their floral resources is by synchronizing their life history transitions with flowering phenology cues (Beatley 1974), often across multiple timescales (Wcislo 1996). One mechanism through which this synchronization is achieved is by using humidity (or precipitation) cues to trigger emergence. Some bees are able to stay in a state of developmental arrest, rather than eclosing, during harsher years with fewer floral resources available (Danforth 1999, Orr et al. 2016). Another approach is to time emergence with plant blooms, which minimizes catastrophic losses in unfavorable years (Danforth 1999). In semiarid regions, the slightly higher precipitation assures trees, shrubs, and herbs to develop at the same time (Noy-Meir 1973), which creates several spatial niches, potentially reducing competition and favoring the coexistence of more species (MacArthur & MacArthur 1961). For example, in the Brazilian semiarid, the bee's high

coexistence is mainly due to the low overlap in the use of floral resources (Araújo *et al.*, 2021; Santos *et al.*, 2013).

To achieve this synchrony between desert bees and the plants they visit, plants might evolve to secrete nectar and release pollen at the same time of pollinator activity in order to achieve successful pollination (Willmer and Stone 1997), or flowers might modulate the temporal organization of pollinator activity (Bloch et al. 2017). For example, plants pollinated by *X. olivieri* have the peak of nectar production, mirroring the bee's flight activity. Another approach is to exhibit facultative multivoltine, where more than one generation may be produced per year, rather than the more common univoltine lifestyle, where species produce only one generation per year. Multivoltinism allows bees to take advantage of years with a long flowering and resource-abundant season.

Consequences of desertification for bees

Desertification is among the greatest contemporary environmental problems threatening global ecosystem sustainability, biodiversity conservation, and human welfare (Adeel et al., 2005). The United Nations Convention to Combat Desertification defines desertification as "reduction or loss of the biological or economic productivity and complexity resulting from various factors including climatic variations and human activity in arid, semi-arid and dry sub-humid areas" (UNCCD 1994). Previous studies have found that the global magnitude of dryland degradation ranges from 10-20% of the Earth's surface, which translates to an estimated total area of 6-12 million km² that are impacted by desertification (Adeel et al. 2005).

For animals that live in desert habitats, desertification poses many potential threats, and it may drastically change entire ecosystems (Brown and Paxton 2009). At the landscape scale, some bees may face increasingly fragmented landscapes where suitable habitat (including nesting and feeding sites) may be more difficult to locate. This would be predicted to result in smaller local population sizes, which might face an increased risk of local extinction due to stochastic effects (Cane *et al.*, 2006; Franzén e Nilsson, 2014). These threats might be particularly strong for specialist bees whose population dynamics are highly dependent on host plant populations. In this context, Cane et al. (2006) found that ground-nesting native desert bees strongly associated with the creosote bush (*Larrea tridentata)* were less abundant in smaller fragments, where there are fewer bushes and consequently fewer floral resources for specialist bees (Cane et al. 2006). Another study of a desert bee performed near the Mojave

Desert in Nevada found that *Perdita meconis* Griswold (Andrenidae), a native solitary ground-nesting bee strongly specialized on the rare dwarf bear poppy (genus *Arctomecon)* and the closely related *Argemone* (Griswold, 1993), is now locally extinct in sites that had few flowering *Arctomecon californica* plants (Portman, Tepedino e Amber, 2018). In this system, *P. meconis* faces not only sparse food resources but also competition with the introduced European honeybee (*Apis mellifera* L.).

Changing resource distributions and habitat fragmentation in arid and semi-arid regions may also translate to range shifts in some species. For instance, (Bezerra *et al.*, 2019) evaluated the effects of climate change on crop pollination in Brazil and confirmed a reduction of ~35% of suitable habitat for the semi-arid carpenter bee *Xylocopa grisescens* under future climate change scenarios. This species is an important pollinator of Passion fruit (*Passiflora edulis*), a crop of which Brazil is the world's largest producer. Similarly, (Giannini *et al.*, 2012) modeled likely distributions of ten Brazilian bee species under different future warming scenarios and found that nine of the species were expected to show range contractions under future conditions. These results substantiate how some bees, including some important crop pollinators, are prone to extinction due to desertification-driven landscape degradation.

Interestingly, an alternative hypothesis is that desert and semidesert bees might be advantaged by increasing desert area because these bees will have more suitable habitat available in future warming scenarios, allowing them to shift their areas of occurrence and successfully colonize new areas. Supporting this idea, Bezerra *et al*. (2019) found that the Brazilian semiarid native bee *Xylocopa frontalis* would gain new potential suitable areas and present a potential shift of areas in future scenarios of climate change. Likewise, Silva et al. (2018) found a minor increase in distribution under future warming scenarios for the Australian allodapine *Exoneurella tridentata* Houston, 1976 *(Apidae: Allodapini)*, a bee that nests in only two arid/semi-arid tree species and whose current distribution is restricted to arid and semi-arid regions of Australia.

However, it appears that many desert bee species exist at very low population sizes, with high genetic structured populations driven by patchily distributed resource availability (Cane et al. 2006). Therefore, models that predict range expansions for these species based on changes in suitable habitat may be inaccurate because these species are not solely limited by habitat availability. Further, dryland degradation is considered to be the result of both climate change and human impact (Adeel et al., 2005). Adeel et al. (2005) created four future scenarios

addressing desertification and human well-being in these areas, and under all four scenarios, population growth and increasing food demand will drive an expansion of cultivated land, often at the expense of woodlands and rangelands (Adeel et al. (2005). This means that even if species ranges increase due to climate change, projections for human development must also be accounted for.

The threats that land degradation poses to bees are also intertwined with increasing agricultural demands in arid and semiarid areas. These bees are endangered by agricultural demands in multiple ways, including loss of foraging habitat via the conversion of heterogeneous multifloral resources into homogeneous monofloral areas and decreases in survival due to harmful agrochemical insecticides. Although desert agricultural pollinators have been less well studied (but see Pitts-Singer and Cane, 2011), studies in temperate bee systems broadly suggest that agricultural intensification has detrimental impacts on pollinator communities when compared to natural habitats (Kremen et al. 2002; Senapathi et al., 2015). These relationships are less understood in xeric areas, but we expect no different responses since deserts present discontinuities that harm bee biodiversity.

Hence, knowing that desert and semidesert systems have the richest bee community in the globe and that this community often rely on specific plant species of their habitat, needing diversity of plants to maintain temporal and spatial niche to avoid competition, and acknowledging how agricultural and urbanization has grown in the drylands, is important to have habitat management schemes for conserving these areas and to develop policies to maintain bee biodiversity in arid habitats.

CAPÍTULO III

Population genomics of a large carpenter bee reveals patterns of high dispersal in a tropical semi-arid region susceptible to desertification

A pesquisa feita neste capítulo representa uma colaboração entre a autora, a Dra. Francisca Soares de Araújo, Dra. Christiana Faria, Dr. Alan Brelsford e a Dra. S. Hollis Woodard.

Revista de interesse: Proceedings of the Royal Society B | ISSN :1471-2954 (online)

Esse capítulo apresenta uma investigação do status gênomico atual das populações de *X. grisescens* e o potencial fluxo gênico ocorrendo entre essas populações. Aqui, utilizamos a ténica de RADseq (Restriction site associate DNA sequencing) para avalliar a potencial relação da diversidade e estrutura genética dessas populações com o isolamento geográfico em áreas suceptiveis a desertificação no semiárido brasileiro.

Abstract

Current and future scenarios predict that drylands biodiversity is highly threatened due to desertification. On a landscape scale, one consequence of this degradation is habitat fragmentation and habitat loss, which can lead to reduced animal movement and, consequently, loss of gene flow and population isolation. Here, we investigated the population genomics of an ecologically and economically important carpenter bee (*Xylocopa grisescens* Lepeletier) dwelling in a region highly susceptible to desertification in Brazil. For this, we used a genome-wide approach (RAD-seq) to gather genetic data (83,127 SNPs) for assessing genetic diversity and structure in this species. Data were collected from a set of eight sites, spanning distances of up to 300 km, and distributed within and between areas of extreme aridity in this desertification hotspot. The amounts of population genetic diversity and differentiation we detected are consistently low for this carpenter bee species across the entire study region, although the distribution of genetic diversity was relatively heterogeneous across sites. Our data suggest that despite the ongoing desertification process impacting the Brazilian semiarid region, these ecological changes have not strongly affected the genetic variability of *X. grisescens*, a highly mobile pollinator species.

KEYWORDS: bees, desertification, dispersal, population structure, RADseq.

Introduction

Desertification is one of the primary processes shaping our planet today. This process is defined as land degradation taking place in arid, semiarid, and dry sub-humid areas (collectively called "drylands") (UNCCD, 1994). Together, drylands represent 41% of the world's land cover, of which some 10-20% are already severely degraded (Reynolds et al., 2007), translating to an estimated total area of 6-12 million km² affected by desertification worldwide. Future scenarios built for the probable conditions in 2050, which are driven by the ongoing impacts of human-mediated CO_2 emissions and global warming, predict a much larger area under threat (Millenium Ecosystem Assessment [MEA], 2005; Reynolds et al., 2007). Thus, as desertification is likely to increase worldwide, it has become increasingly important to understand and predict species responses to this expanding force shaping our planet and the biodiversity it harbors.

A large body of literature has documented the adverse effects of human-mediated land degradation on native species persistence (Angelo, de *et al.*, 2013; Filgueiras, Iannuzzi e Leal, 2011; Kline e Joshi, 2020; Leal *et al.*, 2012)). Broadly, when original land cover is anthropogenically converted to less suitable habitat, this can reduce the total amount of habitat available for species (Fahrig et al., 2015) and also increase habitat fragmentation by increasing the number and isolation of patches of suitable habitat (Hadley & Betts, 2012; Haila, 2002). When habitat is highly fragmented, resources such as food and nest sites can exist only in small, isolated areas; this can hinder individual dispersal and consequently increase the chances of isolation in species that are unable to move between these areas (Segelbacher et al., 2010). In arid landscapes, movement and dispersal can be further impacted by inter-annual differences in rainfall, which can alter dispersal-related species interactions (Louthan et al., 2018; F. M. P. Oliveira et al., 2019), and also have direct effects on behavior and physiology that limit movement (Araújo, Castro, & Albuquerque, 2007). If dispersal is restricted for a sufficient duration of time, whether by habitat unsuitability or other factors, neutral loci may exhibit high levels of differentiation between populations. This differentiation can lead to Isolation-by-Distance (IBD, Wright 1943), a term that describes the pattern in which genetic differentiation increases with geographical distance (Wright 1943). A strategy to identify individual dispersal and colonization ability is by measuring the genetic structure of a population, which can reflect a combination of genetic drift, amount of gene flow, and demographic history (Bohonak, 1999; Slatkin, 1985a). If reproductive individuals cannot disperse readily, due to fragmented habitats or other constraints, after a sufficient amount of time has passed since the occurrence of these constraints, we would expect relatively high levels of genetic structure and population inbreeding, low effective population sizes, and even local extinctions (Balkenhol, Cushman, Storfer, & Waits, 2016; Pardini, de Bueno, Gardner, Prado, & Metzger, 2010).

The impacts of habitat fragmentation can be harmful to smaller-bodied animals such as bees (Anthophila) because it can reduce already sparsely-distributed food and nesting resources in some habitats, such as deserts (Cane, Minckley, Kervin, Roulston, & Williams, 2006). For example, wood-nesting bees (as opposed to ground-nesting) may face more substantial limitations due to the reduced abundance and diversity of plants on which their nesting relies. This can make it even more challenging to travel the distances required to search for appropriate nesting sites and food resources in a large, unrewarding matrix. For instance, there is evidence of species presenting strong local site fidelity for building nests in a region largely modified by anthropogenic actions as reported for the carpenter bee *Xylocopa virginica* L. (Ballare & Jha,

2020). However, there is also increasing evidence that bees are relatively mobile organisms (Ballare & Jha, 2020; Exeler, Kratochwil, & Hochkirch, 2008; Jaffé et al., 2016; Suni, 2017), and within the group, traits such as body size and sociality appear to mediate environmental effects on genetic structure (López-Uribe, Jha, & Soro, 2019). Thus, bees have emerged as an important system for understanding the genetic consequences of anthropogenic habitat degradation in light of their shared but also unique, lineage-specific traits. Concerning desertification, some studies exist on how native bees in drylands are potentially harmed due to natural vegetation reduction (Cane et al., 2006; Danforth, Ji, & Ballard, 2003; Martins, de Siqueira, Kiill, Sá, & Aguiar, 2014), although a study had also found a desert cavity-nesting bee benefitting from urbanization possibly due to an increase in sites for nesting in the landscape matrix (Cane *et al.*, 2006). Yet, the effects of desertification on genetic structure and genetic diversity in native bees are still broadly not understood for the overwhelming majority of bee taxa, given the high diversity seen in this insect group (an estimated > 20,000 species), especially in arid regions (Michener, 2007; Minckley, 2014; Orr et al., 2020).

In Brazil, the most critical desertification hotspot is in the Northeast (Vieira et al., 2015). This region encompasses a semiarid landscape known as the Caatinga ecoregion (Andrade-Lima, 1981). Northeastern Brazil, where caatinga is located, is one of the world regions that are most vulnerable to climate change (IPCC, 2007). Since Brazilian colonization in the sixteenth century, the caatinga ecoregion has declined from originally occupying ~900,000 km^2 of the area to currently existing in remnants that correspond to only ~40% of the original coverage (Silva, da, Leal e Tabarelli, 2018). Human land use (such as cattle overgrazing, slash-and-burn agriculture, replacement of native vegetation by crops, and natural wood removal for charcoal production) has collectively established a continuum of degradation in Caatinga native vegetation that has dramatically increased the rate and risk of desertification (E. V. de S. B. Sampaio, Menezes, Sampaio, & de Freitas, 2018; E. V. de S. B. Sampaio & Sampaio, 2006). Concerning the impacts of global warming, caatinga is reported to face vegetation reduction and increase environmental susceptibility to degradation (Bezerra et al., 2019; Vieira et al., 2015). Such habitat change is especially high for the Brazilian state of Ceará, which belongs to the northeast portion of caatinga and is susceptible to desertification throughout its entire area (~148,000 km^2; CGEE, 2016). Ceará state owes its high degradation risk in part to being the location where the first villages in Ceará were established (Jucá Neto, 2009), and it has been declared one of the priority conservation areas in the Brazilian northeast (CGEE, 2016; Tomasella et al., 2018). As such, three main large areas (totaling over ~28,000 km^2) that are

more susceptible to desertification in Ceará (hereafter called ASDs, which refers to "areas susceptible to desertification") have been delimited for land use monitoring in order to understand and ameliorate the negative effects of desertification, and to predict key species and genera responses to it in this region (CGEE, 2016).

Understanding the impacts of desertification on the biodiversity of Brazil in the caatinga and beyond is especially important in the context of economically and ecologically important species. Large carpenter bees of the genus *Xylocopa* (family Apidae, subfamily Xylocopinae) are uniquely responsible for effectively pollinating a significant number of Caatinga plant species (Neves, Teixeira, Silva, & Viana, 2006) and important crops such as common bean, passion fruit, pumpkin, watermelon, tomato cotton, coffee, sunflower, and many others (Klein et al., 2020). The group's importance is partly due to their relatively large body size and floral sonication for pollen harvesting, which helps to meet the pollination requirements of some crops (Buchmann, 1983; Gerling, Velthuis, & Hefetz, 1989). For instance, passion fruit (*Passiflora spp.*) has physiological barriers to avoid self-pollination as well as a certain intrafloral distance that makes it impossible for smaller insects to make contact with reproductive parts while collecting nectar (Bezerra et al., 2019; Giannini et al., 2012; Silva et al., 2014). The *Xylocopa* species from Brazil are the only effective pollinators of the yellow passion fruit (*Passiflora edulis flavicarpa*) crop for which Brazil is the world's largest producer (Freitas & Oliveira Filho, 2003; USAID, 2014), with marked economic importance in northeastern Brazil (S. R. Silva, Almeida, de Siqueira, Souza, & Castro, 2019). The large-sized carpenter bee *Xylocopa grisescens*, easily recognized by its hairy body and dense white hairs in the mesosoma, is a solitary bee that uses dead wood for building its nests, needs clear space for perching and mating, and whose female bee offspring potentially exhibit high philopatry (Klein *et al.*, 2020; Marchi e Alves-dos-santos, 2013). Nests are usually excavated in standing or fallen dead wood, especially in dry trees or branches with solid wood, but poles and construction wood can also be used (Klein et al. 2020, Martins et al. 2014). *X. grisescens* is an important generalist native pollinator found throughout Brazil, particularly in the Brazilian Caatinga (Marchi & Alves-dos-santos, 2013), where it is predicted to experience significant reductions in suitable habitat in the coming years. Native Caatinga remnants are abundant in nesting sites and food resources for *X. grisescens* (Martins et al., 2014; S. R. Silva et al., 2019), but desertification scenarios create difficulties for keeping these bees when habitat is lost. A recent study predicted that *X. grisescens* would lose ~35% of suitable habitat in the coming decades due to climate change (Bezerra et al., 2019), which could jeopardize the essential pollination services of this species

provides (Klein et al., 2020). Given the international priority of preserving valuable ecosystem services of pollinators for food production (IPBES, 2019), it is important to predict how these patterns in increasing desertification will influence patterns of genetic variation in pollinators such as *X. grisescens.*

Here, we assess for the first time the population genomic status of *X. grisescens,* a carpenter bee that may be facing desertification effects. Our study addresses the question of whether desertification has thus far imposed negative effects on genetic diversity and population structure of *X. grisescens* populations in a Caatinga ecoregion under desertification. To answer this, we used restriction-site associated DNA sequencing (RAD-seq) to generate >80,000 molecular markers for population genomic analyses for a set of bees collected from eight sites, with distances ranging from ~20 to ~300 km. These sites were located within a matrix of two distinct areas susceptible to desertification (ASDs), separated by a semiarid (but not ASD) landscape. This sampling design allowed us to assess the effects of geographic distance, including across intensive aridity (within ASDs), on genetic patterns in this system. We then used these data to infer gene flow and to assess genetic differentiation and population structure in our study region, to explore how (and if) desertification has shaped population genomic patterns and how old this modifying event is. We hypothesized that desertification has reduced genetic diversity in *X. grisescens* across our study region and driven population differentiation by hindering gene flow during a recent event of intensive degradation. Specifically, we predicted that *X. grisescens* populations would (1) exhibit high levels of homozygosity due to prolonged isolation, mediated by increased aridity in ASDs; and (2) exhibit significant genetic differentiation and be highly structured (non-panmictic) as a response to inbreeding and dispersal limitations, across the study region.

Materials and Methods

Study area and sampling

The study was carried out in the State of Ceará in Northeast Brazil (Figure III.1). The climate in this region is a Köppen-Geiger Aw climate (Peel, Finlayson, & McMahon, 2007) with aridity index values ranging between 0.21-0.50, annual rainfall varying from ~400-800 mm, and mean temperatures between 26-28 ºC (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005; Nimer, 1989; Thornthwaite, 1948). Water availability is defined by short rainfall pulses during the rainy season, concentrated at 3-5 months per year, for which reason the region

exhibits strong seasonality and high spatial-temporal resource variability (Nimer, 1989). The predominant vegetation in the Caatinga ecoregion is characterized by a mostly deciduous xerophilous thorn woodland/shrubland with ephemeral seasonal herbaceous-stratum, mainly composed of therophytic life-form species (Costa, Araújo, & Lima-Verde, 2007; E. V. S. B. Sampaio, 1995).

We sampled eight field sites (n = 10 individuals per site) separated by at least 16 km throughout the State of Ceará, which was our focus due to its historical disturbance (see in Supplemental Information) and priority on monitoring desertification (PAE-Ceará 2010). Our study sites were clustered in two distinct regions of ASDs (CGEE, 2016) near the cities of Irauçuba, Santa Quitéria (two sample sites), and Sobral (cluster ASD I), Jaguaretama, Jaguaribe (two sample sites), and Morada Nova (cluster ASD II) (hereafter referred to as IRA, SQI, SQII, SO, JM, JGI, JGII, MN, respectively; Table III.1), separated by a landscape that is arid but not classified as ASD. Details describing the characteristics of the ASDs can be found in the Supplementary Material. This sampling design allowed us to explore genetic patterns within areas of more extreme aridity and habitat degradation (within two separate ASDs), separated by areas of relatively lower aridity (between two ASDs). The number of sample field sites consisted of an attempt to cover each ASD entirely (Figure III.1) while avoiding collecting bees from non-independent sites within these areas. For determining the minimum distance between sites, we set thresholds above the foraging ability of the genus (up to 13 km for some *Xylocopa* species; Greenleaf et al., 2007), given that data for dispersal distances for our focal species is currently unknown. The complete species range of *Xylocopa grisescens* is distributed across Brazil and in parts of Paraguay, and thus our sampling design focuses specifically on the part of the species range that spans the state of Ceará (in the northern caatinga). Females of *X. grisescens* studied herein were captured during the rainy season (March to May) due to the higher food availability for pollinators in caatinga at this time of the year. Ten bees from each site (for a total of 80 individuals) were collected while foraging using a sweep net, immediately preserved in 95% ethanol, and posteriorly stored at -20°C. Geolocation data for sampling points are provided in Table V.1 (Chapter V)**.**

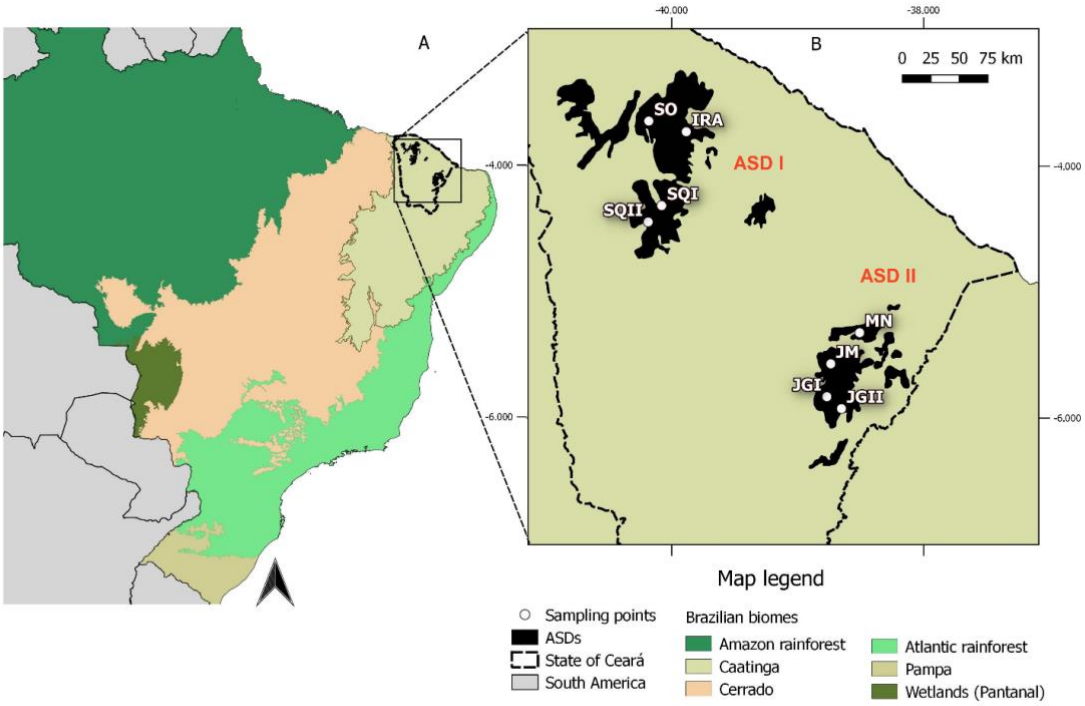


Figure III.1. Map of geographic sampling locations of *X. grisescens*. A) Location of Brazil within South America and its biomes. B) Details of the State of Ceará delimitation (Brazilian northeast) and its respective areas susceptible to desertification as black polygons (ASDI and ASDII). ASDI comprises localities IRA, SO, SQ, and SQII, and ASDII includes localities MN, JM, JGI, and JGII. White dots represent the sampling points of each of the eight populations, with names (acronyms) provided in Table III.1.

DNA extraction and library preparation

DNA was extracted from one middle leg of each individual using a Qiagen DNeasy Blood and Tissue kit (Qiagen, 2006), following the manufacturer's instructions with minor modifications. We used 70 % ethanol instead of AW2 and used 30 µl of AE elution buffer. DNA quality and concentration were verified using a Qubit 3.0 (Life Technologies). We generated restriction site-associated DNA (RADseq) libraries for all 80 bees using the method of Brelsford et al. (2016), which incorporates elements of those proposed by Parchman et al. (2012) and Peterson et al. (2012). Briefly, genomic DNA was digested using *Mse*I and *EcoR*I restriction enzymes (Illumina, Inc.), followed by the annealing of two unique adapters for each DNA fragment (forward and reverse). Ligated samples were purified with Agencourt AMPure

beads (Beckman Coulter, Inc.) to remove small DNA fragments, and then two PCR amplifications were performed. All libraries were multiplexed in a single lane and sequenced on an Illumina HiSeq X-Ten paired-end 150bp platform at Novogene Corporation (https://en.novogene.com).

Bioinformatic data processing

Illumina reads in FASTQ format were filtered to remove any reads with an unnamed base and low overall quality. The upper limit for the sequencing error rate was set to 0.05. Following this, we used the PROCESS_RADTAGS tool in STACKS v2.3b (Catchen, Amores, Hohenlohe, Cresko, & Postlethwait, 2011; Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013) to demultiplex raw data by barcode, filter for low quality reads, and call single nucleotide polymorphisms (SNPs). After assessing read quality, we used the DENOVO_MAP pipeline to perform a *de novo* assembly of loci for all populations and the POPULATION tool to filter SNPs including only loci present in at least 75% of individuals, a quality score > 20, minor allele frequency (MAF) > 0.1, and to export SNP data in VCF format. Then, on VCFtools (Danecek et al., 2011), we used the flag --hardy to check on heterozygosity excess and excluded all sites with an excess of heterozygosity using the flag --exclude-positions. VCFtoolswere also used to convert VCF files to different file formats and to assess locus and sample coverage depth statistics.

Population genetic diversity and structure

After filtering, we used the POPULATIONS tool in Stacks v2.3b (Catchen et al., 2011, 2013) to generate genome-wide population statistics, such as expected heterozygosity (H_e), observed heterozygosity (H_o), private allele number (A_p), percentage of polymorphic loci (%P), and inbreeding coefficient (F_{IS}) for each population. To investigate the genetic structure, we used three main approaches. First, Weir-Cockerham F_{ST} was calculated using VCFtools (Danecek et al., 2011) to assess the proportion of the total genetic variance among populations relative to the variance within populations. F_{ST} values were plotted as a heatmap using the R package ggplot2 v.3.3.0 (Wickham et al., 2020). Second, we used PLINK v.2.0 (Purcell et al., 2007) to perform a Principal Component Analysis (PCA) by extracting PC coordinates for each

individual and visualized the results using the R packages ggplot2 v.3.3.2 and tidyverse v.1.3.0 (Wickham et al., 2019, 2020). For our third approach, we generated bed, bim, and fam files in Plink v.2.0 to be analyzed on the Bayesian model-based clustering program fastStructure (Raj, Stephens e Pritchard, 2014). We evaluated prior clusters from $K = 1$ to $K = 8$, with 20 independent runs for each K (where K equals the number of population clusters) to map genetic structure and common kinship. The optimal value of K was obtained by running the flag c*hooseK.py* according to two metrics to obtain a range of values for the number of populations that explain structure in data: 1) model complexity that maximizes the marginal likelihood of the entire data, and 2) model components used to explain structure in data. We further visualized the observed admixture for chosen values of K using the flag *distruct.py*.

Isolation by distance

To test Isolation by distance (IBD), we constructed two distance matrices (genetic distance and geographic distance) and tested their correlation using a MRDM (Multiple regression on distance matrices) approaches in the R package ecodist (Goslee e Urban, 2007) with 10,000 permutations. The matrix of genetic distances between populations was constructed by calculating pairwise F_{ST} using VCFtools (Danecek et al., 2011) and set as the response variable, while the geographic distance matrix was calculated from the sample site coordinates using the function vegdist in vegan package (Oksanen et al., 2018) and set as the exploratory variable. To better visualize this correlation, we plotted pairwise F_{ST} against geographic distances as a scatterplot using the package ggplot2 v.3.3.2 in R.

Results

SNP Calling and Filtering

RAD sequencing for the 80 individual carpenter bees resulted in a total of ~917 million raw reads, and ~815 million retained high-quality reads with a mean effective coverage per individual ~36x (7.5 - 81.6x). The *de novo* assembly generated a total of ~1,370,000 loci with a mean length of 253 bp for the entire dataset. Quality filtering resulted in 83,127 single nucleotide polymorphisms (SNPs). SNPs were then used to calculate genome-wide estimates of genetic diversity and population structure.

Population genetic diversity

We observed low levels of variance in genetic diversity amongst all localities (Table III.1). Among all sites, no significant difference was found when analyzing the percentage of polymorphic loci (%*P*). This ratio, which was measured by dividing the number of polymorphic loci by the total number of loci, ranged narrowly from 33,7% (SQI) to 35,2% (JGII), denoting low variation among sites (1.5 %) (Table III.1).

Expected heterozygosity under Hardy-Weinberg Equilibrium (H_E) consistently exceeded observed heterozygosity (H_O) over localities ($Ho < He$), suggesting heterozygote deficiency, except for IRA (Table III.1). H_O ranged from 0.316 (SQI) to 0.333 (IRA), and similarly, H_E ranged from 0.325 (SQI) to 0.334 (SQII). Our data showed low F_{IS} values among all localities ($\bar{x} = 0.0704$), revealing little evidence of population inbreeding. Among sites, IRA showed the lowest mean values (F_{IS}= 0.054, 5.4%) followed by a slightly higher value in JGII (F_{IS}= 0.059, 5.9%), whereas MN and JGI showed the highest mean values (F_{IS} = 0.085, 8.5% and F_{IS} = 0.088, 8.8%, respectively). The number of private alleles (NPa) varied from 31 (IRA) to 121 (SQII).

Table III.1. Summary statistics of estimates of population genetics parameters based on 83,127 SNPs from all 80 individuals of *X. grisescens*. Abbreviations: NPa = Number of private alleles; *P%* = percentage of polymorphic loci; H_O = observed heterozygosity; H_E = expected heterozygosity; F_{IS} = inbreeding coefficient.

ASD	**Population**	**NPa**	**P (%)**	***Ho***	***HE***	***FIS***
I	IRA	31	0.340	0.333	0.331	0.054
I	SO	73	0.338	0.331	0.332	0.060
I	SQI	72	0.337	0.316	0.325	0.081

I	SQII	121	0.344	0.331	0.334	0.068
II	JGI	108	0.342	0.321	0.333	0.088
II	JGII	72	0.352	0.332	0.333	0.059
II	JM	67	0.341	0.330	0.333	0.068
II	MN	107	0.343	0.322	0.331	0.085

Population genetic structure

Pairwise comparisons of genetic differentiation index between sites (F_{ST}), although low, were significantly different from zero (t = 3.9703, df = 27, p-value <0.001), with an average of 0.0012 and a range from -0.003 to 0.0040 (Figure III.2). Locations JM and JGII, two sites that are > 70 km apart, showed the strongest divergence (Table III.1).

Both complementary genetic clustering analyses we employed (fastStructure and PCA) also revealed some genetic structure. Admixture distributions are observed as *distruct* plots (Figure III.4), showing a few dominant components of ancestry but not consistent with prior population designations). Accordingly, the results from fastStructure suggested the optimal value of K to be between $K = 1$ for the model complexity that maximizes marginal likelihood and $K = 3$ for the model used to explain structure in the data, with marginal likelihood values decreased when raising K (Table III.2). Principal component analysis (PCA) showed a structured grouping of clusters, where we noticed two genetically similar individuals from the same site (JGII) far apart from the group cluster along the first PC axis, and along the second PC axis, we noticed two genetically similar individuals (SQI and IRA) although belonging to two different sample sites (Figure III.3).

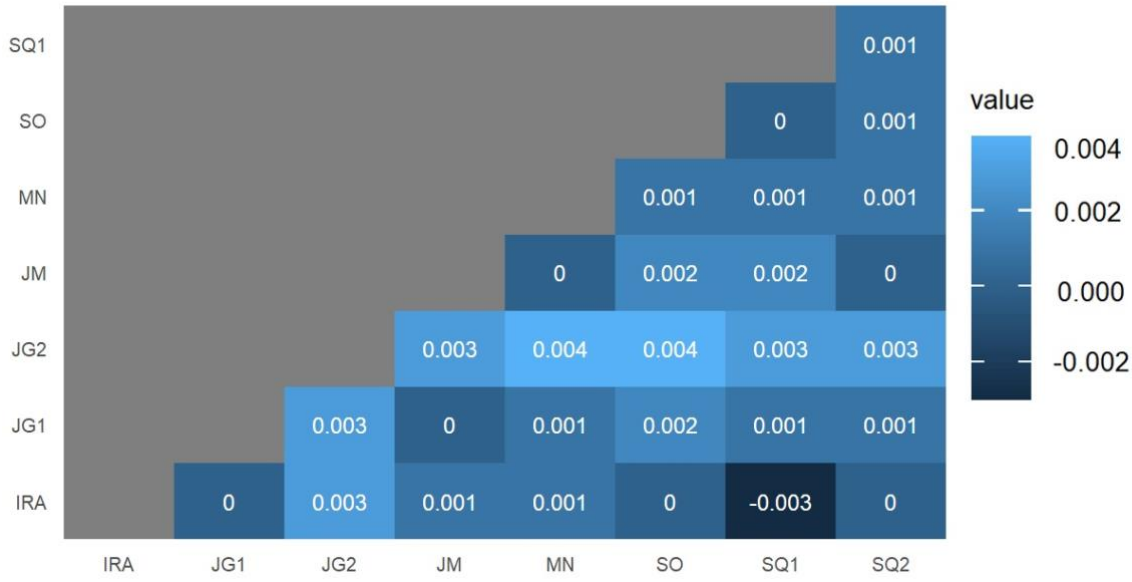


Figure III.2. Weir-Cockerham F_{ST} correlation matrix. Values from each pairwise analysis are displayed inside the grid cells. Heatmap colors represent the strength of each correlation, corresponding to the color scale of the box on the right.

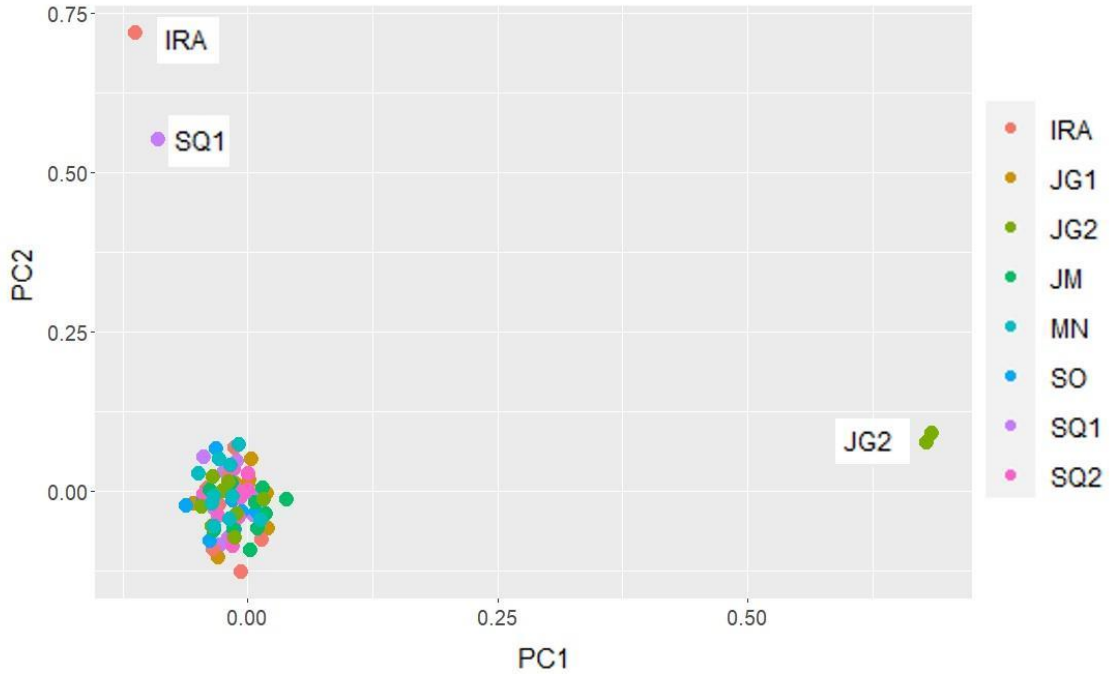


Figure III.3. Population structure of X. *grisescens a*s revealed by PCA analysis based on the total dataset of 83,127 SNPs. Scatter plot of PCA revealing three genetic clusters. Individuals are represented as colored dots, whereas each color indicates a specific sample site.

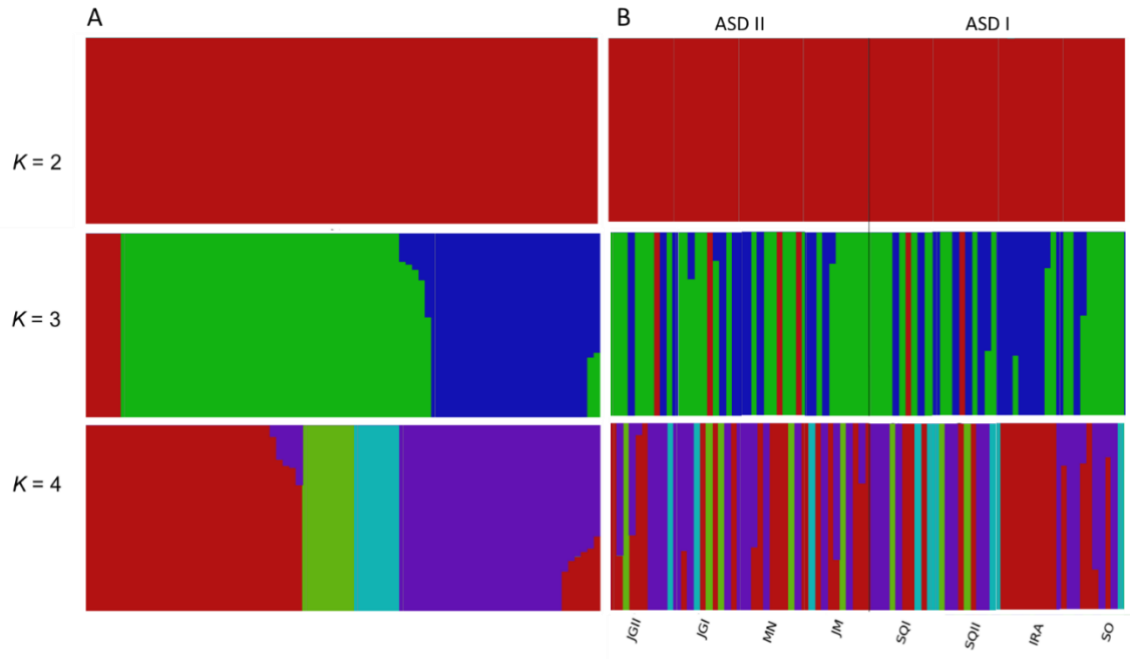


Figure III.4. Population structure results were obtained from fastSTRUCTURE analysis for eight sites based on genetic distance. Each vertical bar represents a single individual of *X. grisescens*, and colors represent unique ancestry proportions (membership) from K=2 to K=4 (marginal likelihoods are provided in Supplemental Information). A) admixture proportions as inferred with no prior population designation; B) admixture proportions according to each geographic location, including ASD's and sample sites. The black line separates ASD II and ASD I sites.

Table III.2. Mean Likelihoods for each cluster of the genetic structure analysis

CLUSTER	**VALUE OF MARGINAL LIKELIHOOD**
K2	-0.6732233154
K3	-0.6859930717
K4	-0.6799925389
K5	-0.6861214889

CLUSTER	VALUE OF MARGINAL LIKELIHOOD
K6	-0.6774970910
K7	-0.6765810352
K8	-0.6732945109

Isolation by distance

In all pairwise populations, genetic differentiation expressed by F_{ST} did not increase with geographic distance. RMDM results of IBD showed low values of correlation coefficient (R^2) = 0.0038, F-value = 0.0719, and no statistical significance (p-value = 0.7021 0.25400). This means that only about 0.38% of genetic differentiation is explained by geographic distance and that the association between differentiation and distance is not statistically significant, reinforcing no Isolation by Distance in our data. Plotted values are provided in Figure III.5.

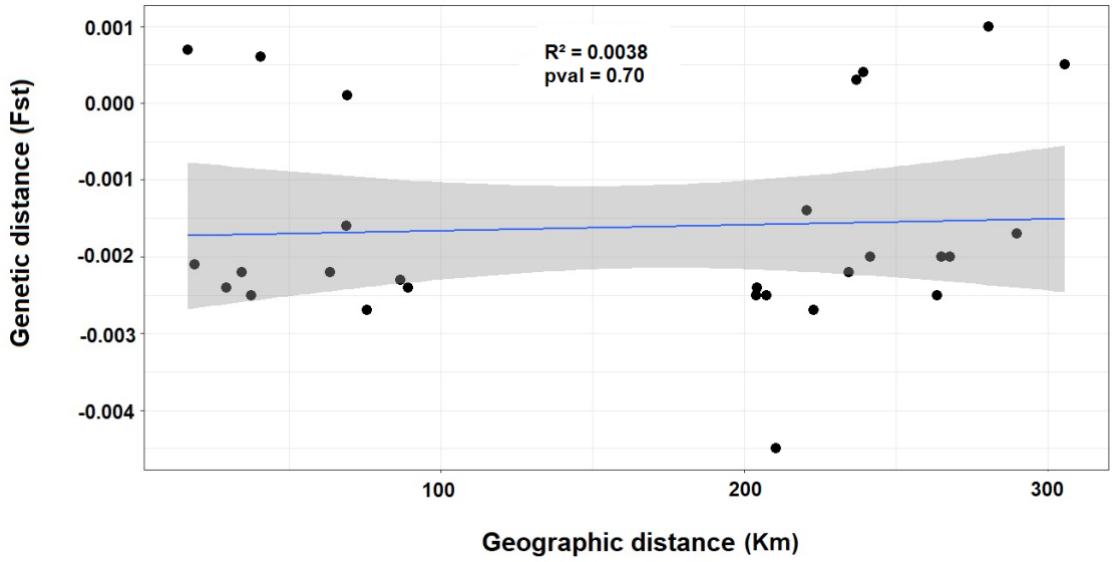


Figure III.5. Scatter plot of MRDM analysis for isolation by distance showing the pairwise relationship between genetic distances and geographic locations of populations of *X. grisescens*. Genetic distances were obtained by calculating pairwise F_{ST}. Geographic distances are

displayed in kilometers (km). The test revealed no significant correlation ($R^2 = 0.0038$, pvalue= 0.7021) between analyzed matrices. A regression line is shown with a 95% confidence shadow.

Discussion

We examined genome-wide data from a Brazilian carpenter bee to evaluate genetic diversity, structure, differentiation, and demographic history across dryland habitats subject to intensive and ongoing desertification in northeast Brazil. Collectively, our results paint a picture that mobility and gene flow in *X. grisescens* is quite high in the Brazilian semiarid region, despite examining this across a large area in a region that is prone to ongoing desertification. Our study demonstrated that the variation of population genetic diversity is consistently low in this region and that the distribution of genetic diversity (i.e., population structure) is significant, but not related to the geographic distribution, even across the study area, which included sites separated by up to 300 km distance. These results indicate that populations are consistently sharing genetic material, which is an indirect measure of high gene flow and individual dispersal. Although previous studies have explored genetic variability and IBD in desertified areas among other taxa (e.g., plants, mammals, and reptiles; El Mousadik & Petit, 1996; Epps et al., 2005; Shaffer et al., 2017), our study is the first one to present spatial genetic structure based on many thousands of neutral genetic markers of a pollinator species throughout a region susceptible to desertification.

Habitat degradation can lead to species loss and can contribute in part to the loss of genetic variation (Balkenhol et al., 2016). One reason why this phenomenon is especially harmful to species dwelling in drylands is that it makes already sparsely distributed food resources and nest sites even less accessible (Galvin, Reid, Behnke, & Hobbs, 2008). However, organisms can be robust to these effects if they can move readily between habitat patches. Our results suggest that *X. grisescens* is likely mitigating the issue of intensive desertification in our study region through its relatively high mobility between different fragments of the landscape. Given the spatial distances that we detected evidence of gene flow across (up to 300 km distance), it appears that some individuals of this species have the capacity to migrate across both ASDs and the semiarid landscape between them. The individuals who manage to overcome these landscape barriers and increase access to food and other requirements for reproduction (e.g., nesting sites) are likely those who maintain gene flow and a continuously distributed population even in such degraded areas.

Fragmentation, urbanization, and agricultural land use are largely assumed to have negative effects on populations' genetic structure, leading to populations with higher genetic structure by limiting dispersal (Davis et al., 2010; Jha & Kremen, 2013). Our findings are consistent with other studies, which have found that despite their relatively small body size, bees are highly mobile and able to maintain movement and ongoing gene flow. For example, Jaffé et al. (2016) found the tropical stingless bee species *Trigona spinipes* capable of dispersing over remarkably long distances (up to a 200 km range), even across agricultural landscape mosaics and human-altered forest fragments in Brazil. Likewise, Suni (2017) found evidence of weak population differentiation (suggesting unrestricted dispersal) for the orchid bee *Euglossa imperialis* over degraded landscapes in Costa Rica. Ballare & Jha (2020) found that human-altered habitats are likely conducting gene flow in *Xylocopa virginica* species, providing evidence that anthropogenically altered habitats can enhance the dispersal of pollinators. Although it is unclear what is facilitating movement in *X. grisescens*, our data broadly support the idea that bees are capable of overcoming many factors that are generally considered barriers to dispersal. More specifically, our detection of two highly related individuals belonging to distant sample sites (IRA and SQI) suggests that highly mobile individuals of *X. grisescens* are maintaining high gene flow and dispersal ability in this study region, even across highly fragmented and desertified areas of great distance. Indeed, there is evidence of high foraging efficiency among carpenter bees due to their generalized resource use, advanced cognitive abilities, the capability to trapline between isolated patches, and thermoregulation (Somanathan et al., 2019).

Due to their patchy distribution of food resources, xeric habitats are predicted to generate high levels of population structure in bees (Danforth et al., 2003). Our findings suggest that there is some consistent genetic structure among sites, but geographic isolation is not the limiting factor. Contemporary population structure could be affected by many other variables despite geographic distance, as suggests Wang and Bradburd (2014). For example, Isolation by Environment (IBE, Wang and Summers, 2013) is often reported to describe population structure in contrast with geographic distance, including in human-induced habitat changes (e.g., Quintela et al. 2021). As our data describe a relatively strong structured population, although not related to distance, we believe some of these bees are dispersing only through selected sites encompassing preferred resources or paths.

Interestingly, the lowest value of genetic distance (F_{ST}) was found between IRA and SQII (at least 80 km apart). By its geography, IRA contains a landscape naturally covered by

rocky outcrops and historically experienced overgrazing disturbance (J. G. B. de Oliveira & Sales, 2015), making it one of the most severely degraded areas among our sampled sites. As both natural and anthropogenic factors potentially reduce the availability of local floral and nesting resources (Fahrig, 2003), we suggest that this resulted in intensified migration from this site towards more rewarding areas, reducing genetic differentiation in neighboring areas. This rationale is supported by the smallest number of private alleles in IRA, suggesting high gene flow and individual migration (Slatkin, 1985a). As for the more distant locations, gene flow may be facilitated by the lack of conspicuous geographic or environmental barriers constraining individual movement. Indeed, 95% of the remaining Caatinga fragments are accessible to species able to cross as few as 1 km of a matrix (Antongiovanni, Venticinque e Roberto, 2018).

According to our PCA and fastStructure results, clusters of individuals were not grouped according to their geographic sampling location, revealing that genomic variation presented here is not geographically structured. Yet, our PCA analysis identified two individuals that share the highest membership coefficient. Those two individuals belong to the same sample location (JGII), and we assume them to be closely related, potentially sharing either the same nest or nearby nests. Nest sharing has been previously reported for *X. grisescens* (Camillo & Garófalo, 1989).

The relatively large body size of our focal bee might make it especially mobile and able to traverse great distances without being limited by desertification. Larger-bodied bees forage disproportionally farther than smaller bees (Gathmann & Tscharntke, 2002; Greenleaf, Williams, Winfree, & Kremen, 2007), and specifically, large carpenter bees of the genus *Xylocopa* are reported to have reached up to 13 km flying distance (Greenleaf, Williams, Winfree, & Kremen, 2007). It has been hypothesized that this is because a larger body size enables stronger flight muscles, thus permitting greater long-distance foraging flight (Somanathan, Saryan, & Balamurali, 2019). In disturbed areas, the ability to travel for large foraging distances is a considerable advantage, as small bees may not be able to collect sufficient food resources during their shorter foraging trips (Gathmann & Tscharntke, 2002; López-Uribe et al., 2019).

Besides large body size, another factor supporting the movement of *X. grisescens* is that in some parts of the Caatinga ecoregion, the connectivity between habitat fragments is eased by remnants of large-sized patches (Antongiovanni et al., 2018), increasing functional connectivity of the landscape. This means that these bees possibly have a high patch encounter

rate, as previously reported by Healey & Hovel (2004). It is also possible that these bees harbor high regional adaptation and have improved the ability to track specific resources, as it is reported for some bee species occurring in arid regions (Minckley, Roulston, & Williams, 2014). Thus, individuals of *X. grisescens* might be easily finding food resources throughout the landscape, successfully colonizing new areas, then shifting their areas of occurrence. Further, carpenter bees tend to be speciose in deserts (Michener 2000, Orr 2020) and have been studied as models of water balance in arid environments, given that they are able to remain active in desert regions even during the hottest parts of the day (Chappell, 1982; Nicolson & Louw, 1982; Willmer, 1988). Physiological adaptations to desert environments might help facilitate movement and persistence in this bee group, despite increasing aridity.

The high gene flow we detected in this system might also be explained by food scarcity, which is characteristic of our study area. In the Caatinga ecoregion, food availability occurs as short rainfall pulses only for a few months a year (Araújo et al., 2007; Nimer, 1989), providing a seasonal food shortage for some species. Perhaps, when the landscape-level floral cover is low, *X. grisescens* travel farther from their nesting sites to reach patches of high floral reward. Bees performing longer flights towards more resource-rewarding landscapes have been previously reported by Westphal et al. (2006) and by Pope & Jha (2018). Also, as *Xylocopa* bees are wood-nesting (Gerling et al., 1989), dispersal could be facilitated by the high availability of nesting substrates even in habitats with seasonally scarce food resources. For instance, *Xylocopa virginica* is largely known for nesting in wooden built structures such as fence posts and house barns (Gerling & Hermann, 1978), and our focal species *X. grisescens* nests in natural substrates such as wooden trunks, driftwood logs, dry tree branches and bamboo stems (Camillo & Garófalo, 1982; Chaves-Alves, Junqueira, Rabelo, de Oliveira, & Augusto, 2011).

Overall, our study sheds light on the genomic status of important carpenter bees in a degraded semiarid region. We suggest that for this species, under the conditions and scale that we explored, gene flow is maintaining populations with individuals that are resistant to geographic isolation and also to ongoing desertification. Moreover, we believe our data will be useful for future studies with *X. grisescens,* given its biological and economic importance, and will lend support to future studies on large carpenter bees' genetic responses to human-altered landscapes.

CAPÍTULO IV

Inferring demographic history and effective population size of a large carpenter bee dwelling in a degraded semi-arid landscape.

A pesquisa feita neste capítulo representa uma colaboração entre a autora e a Dra. Francisca Soares de Araújo.

Revista de interesse: Heredity (*Heredity*) | ISSN 1365-2540 (online) | ISSN 0018-067X (print)

The main idea of this chapter is to continue the investigation on how genomic population data is shaped by desertification. This chapter's focus is to investigate possible changes in landscape configuration and composition that possibly affected the population size of our focal bee.

Abstract

The demographic history of a given species population describes changes in population size through time and split and merge events such as bottlenecking and expansion. Here, we inferred for the first time the demographic history of a Xylocopa species, a carpenter bee performing as an outstanding pollinator of several native plant species and agricultural crops. The large carpenter bee *Xylocopa grisescens* dwelling in degraded semiarid areas of the Brazilian Northeaster have had likely shrinkage in effective population size in an event occurring about 100 to 150 years ago. After the shrinkage event, this bee maintained its population size at, what we believe, to be a viable population size to maintain this species thriving in such degraded habitat.

Introduction

Information about demographic history is essential to disentangle the mechanisms that have shaped a given population. It is possible to infer the demographic history of a population by assessing a suitable model describing changing events that have occurred in a population, like changes in size through time and split and merge events (Beichman, Huerta-Sanchez e Lohmueller, 2018). It is also possible to obtain information on the current effective population size (N_e) and on the most likely mechanisms of speciation and differentiation (e.g., isolation and migration, (Wang *et al.*, 2019). Thus, an accurate estimation of a history of an entire wild population might help identify species and populations at risk of extinction (Allendorf, Hohenlohe e Luikart, 2010; Frankham, 2019; Portman *et al.*, 2017) and thus worthy of conservation effort.

The effective size of a population, N_e, is described as the number of individuals effectively capable of reproducing and effectively contributing to the next generation (Balkenhol *et al.*, 2016). Evolutionarily, it follows that populations with large N_e lean increase their variability and effectiveness of selection (Lanfear, Kokko e Eyre-Walker, 2014), whereas low N_e populations tend to become genetically eroded across generations (Frankham, 2019). The reason is that small-isolated populations hold a load of rare (generally recessive) harmful alleles that would be easily spread through a population by an excess of relatives mating (Frankham, 2019). Through time, this high frequency of harmful alleles might cause negative population growth and lean-to an accelerated decline towards extinction and turning into a more homogenized population (Frankham 2019)

Under natural conditions, a population is likely to maintain its size due to genetic drift, hereafter defined as “changes in the genetic composition of a population (allele frequency) due to random sampling in finite populations.” (Frankham, Ballou e Briscoe, 2008). It means that, under random mating, migration and birth and deaths, a population is likely maintaining a suitable size in suitable habitat. Inversely, when changes such as habitat loss and landscape fragmentation occur, causing patch discontinuities and isolation, a directional selection of alleles also occurs. An example of such patch discontinuity is taking place in the most populous semiarid region in the world (over 54 million inhabitants), located in the Brazilian Northeast (hereafter called BSA, (Silva Pinto Vieira, da *et al.,* 2014), and its highly degraded vegetation called Caatinga (da Silva et al., 2018). The prominent human population density over the eighteenth and nineteenth centuries has intensified anthropic activities in BSA, leading to reduced natural vegetation and an accelerated desertification process (J. M. C. da Silva et al., 2018; Leal & Tabarelli, n.d.).

For many years, assessing the demographic history of non-model species has been hampered by the limited availability of genomic data. However, due to advances in molecular resources such as next-generation sequencing and the availability of reliable metadata for many taxa, it became routine to discover what is shaping a given population. Single Nucleotide Polymorphism data (SNPs) has a fundamental role in such discovery for holding great potential in accurately describe evolutionary data (Peterson *et al.*, 2012). Also, for having decreased computational burden and time for analysis that usually took weeks or months to be done. Indeed, there are plenty of methods in which demographic history can be inferred from SNP data, including one of the most used Approximation Bayesian Approach (ABC, Beaumont, et al. 2002).

ABC computes summary statistics from the real data set (SNP data) and compares to a simulated data set by assembling coalescent simulations (i.e., “a method of reconstructing the history of a sample of alleles from a population by tracing their genealogy back to their most recent common ancestral allele,” (Charlesworth e Willis, 2009). Simulated demographic parameters (prior inferred) are then accepted if they generate summary statistics close to the values seen in real data, thus indicating a best-fit model (Beaumont, 2010; Beichman, Huerta-Sanchez e Lohmueller, 2018). This process is repeated at least for a thousand (or millions) of runs. An advantage of the ABC approach is that as it uses a coalescent simulation from summary statistics, one can explore very complex and realistic models to accurately match a model-fit of a non-model species (Beichman et al., 2018). Thus, assessing demographic history

from a set of few hundreds or thousands of SNPs accurately describes the scheme of events a population has experienced for a large set of different taxa.

Among the native bees of the BSA, the large carpenter bee *Xylocopa grisescens* (Lepeletier) pollinates numerous species of native plants (*Poincianella pyramidalis, Capparis yco, Senna spectabilis, Solanum paniculatum, Eugenia rosea, Ziziphus cotinifolia, Melochia tomentosa* , Aguiar 2003), and important crop species (cowpea (*Vigna unguiculata*), canevalia (*Canavalia ensiformis*), Brazilian nuts (*Bertholletia excelsa*), Squash (*Cucurbita pepo* and *C. moschata*), Chayote squash (*Sechium edule*), Guava (*Psidium guajava*), Watermelon (*Citrullus lanatus*), Tomato (*Solanum lycopersicum*) and *Passiflora spp*., Silva; Freitas 2018). In the absence of some of these BSA native and crop plants due to habitat change, it is very likely that this species would experience a reduction in population size.

Here, we expand the knowledge of *X. grisescens'* evolutionary history by inferring their demographic history and effective population size by using an ABC approach as implemented in the DIYABC software (Cornuet *et al.*, 2014). We introduce, for the first time, an estimation of demographic events occurring in a pollinator dwelling in a highly degraded area that's facing ongoing desertification. We predict that these populations had experienced a recent event of population decreasing due to the high migration of human population from rural areas to the cities in BSA in the eighteenth and nineteenth century, as reported by (Silva, da, Leal e Tabarelli, 2018). We hypothesize that during the past two centuries, *X. grisescens* 1) has decreased its population size, and 2) it is not sustaining an ideal population size to maintain enough genetic diversity.

Materials and Methods

To investigate the effective population size and the population demographic history, we performed a coalescent-based Approximate Bayesian Computation (ABC) analysis in the software DIYABC v.2.1.0 (Cornuet *et al.*, 2014). Based on the low genetic structure identified by fastStructure, PCA, and F_{st} analysis (Chapter III) [*remove for publication]*, we used only one population to design all model scenarios. Using VCFTools, we filtered SNPs with a *–max-missing* of 1.0 to reduce computational burden and to avoid impacts of missing data and sequencing errors. This resulted in high-quality data of 2,914 polymorphic loci.

Historical model

We tested seven competing demographic scenarios, which included constant population size (scenario 1), expansion (scenarios 3 and 5), shrinkage (scenarios 2 and 4), expansion + subsequent shrinkage (scenario 6), and bottlenecking (scenario 7) (Figure IV.1). *Scenario 1* assumed a population from which samples have been taken from a constant effective population size (null model); *scenario 2* assumed a population with a decline in a recent time event (t1) until current days (0); *scenario 3* a population increased its size in an old-time event (t2) until current days (0); *scenario 4*: population declined its size at an old-time event (t2) until the current day (0); *scenario 5*: population increased its size in a recent time event (t1); *scenario 6*: a population increased at an old-time event followed by decreasing at a recent time event (t1) until now; and *scenario 7*: a population bottleneck, with increasing at an old-time event (t2) and back to decline at a recent time event (t1) until now. We then characterized three time-events with uniform distributions for all scenarios, setting unit time in 1, 50, and 100 generations (0, t1, and t2 respectively, table IV.1). We employed a generation time of 3 years, considering carpenter bees to reach up to two or three years of active life (Silva and Freitas 2018).

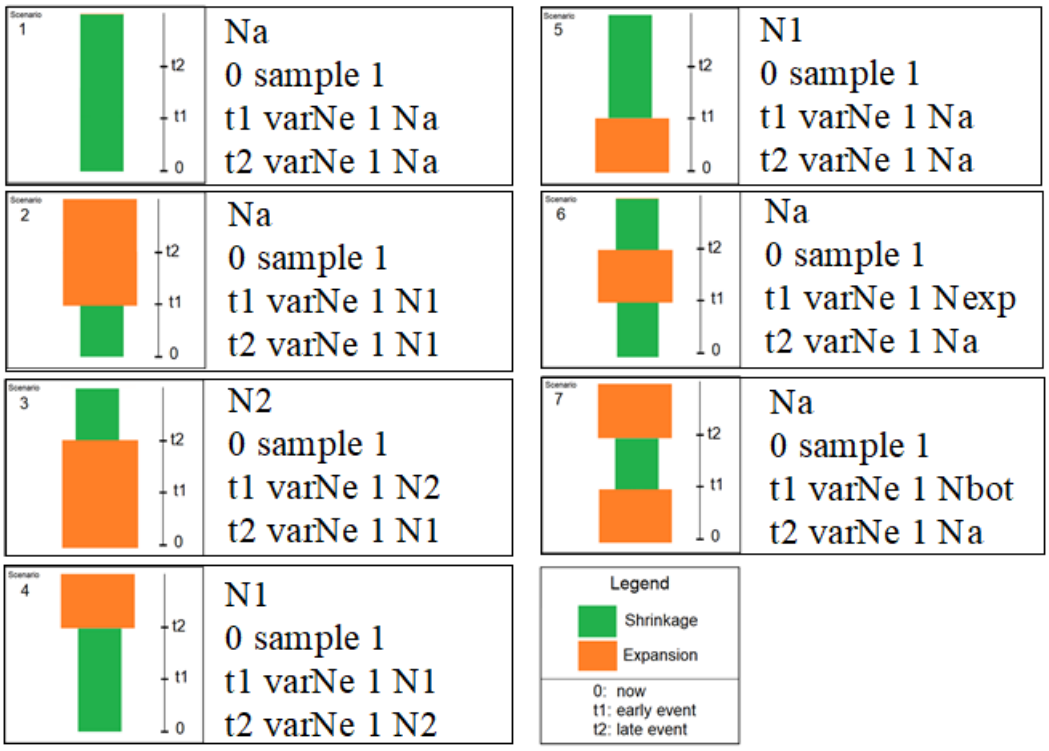


Figure IV.1: Schematic drawing of tested scenarios. 1) constant population size, 2) early population shrinkage, 3) old population expansion, 4) old population shrinkage, 5) early

expansion, 6) old expansion + early shrinkage; and 7) old shrinkage + early expansion (bottleneck). The color green represents population shrinking, whereas the color orange represents population expansion. Time events set to 0 (now), t1 (early event) and t2 (old event).

Prior distributions

We set two-time parameters t1 and t2, where t2 is always larger or equal than the second one t1 (t1 ≤ t2). This means that for every analysis, t2 will be the old-time event and t1 the recent time event (Table IV.1, Figure IV.1). To test for effective population size (N_e), we simulated three different combinations of N varying from 100 – 10,000 UN (Table IV.1). We set "NA" as the current population size, "N1" as the size of the population after first shrink or expansion, and "N2" the size of the population before the first shrink or expansion, and that NA < N1; N1 < N2; and N2 < NA. For each model, we tested the following summary statistics available for a single-population analysis: (1) mean gene diversity across polymorphic loci (Nei, 1972), (2) variance of gene diversity across polymorphic loci, and (3) mean gene diversity across all loci.

Table IV.1: Demographic parameters used in DIYABC.

DEMOGRAPHIC PARAMETER	**DESCRIPTION**	**PRIOR DISTRIBUTION**	**PRIOR CONDITION**
NA	current population size	10^2 - 10^4	NA < N1
N1	size of population before first shrink/expansion	10^2 - 10^4	N1 < N2
N2	size of population after first shrink/expansion	10^2 - 10^4	N2 < NA
T1	time at second population event	1- 50 generations	
T2	time at first population event	5 – 100 generations	t1 ≤ t2

Pre-evaluation and posterior evaluation of scenarios

We simulated 1,000,000 data for each scenario (7×10^6) to test whether at least one combination of scenarios and priors adequately assembling data close enough to our observed data, as suggested by the authors (Cornuet et al., 2014). The best scenario was the one that showed the highest posterior probability with a nonoverlapping 95% credibility interval (CI) according to both direct and logistic regression (Cornuet *et al.*, 2014). In a further analysis, the algorithm ranks each summary statistic of the observed data set against the distribution of the corresponding summary statistics from the posterior predictive distribution.

To lend support to our analysis, we checked for Type I and Type II errors by assessing parameters of confidence in scenario choice. Type I error represents the probability with which one scenario is rejected although it is the true scenario, whereas Type II errors correspond to the probability of deciding for a model scenario when it is not the true scenario (Cornuet et al., 2014). No mutation rate is required for our SNP genotype data analysis since we excluded all non-polymorphic loci for this analysis, thus assuming all loci have only two alleles (an ancestral and a recent one).

Results

Population demographic history

Using DIYABC, we tested a set of seven possible scenarios (Figure IV.1), as follows: 1) constant population size, 2) early population shrinkage, 3) old population expansion, 4) old population shrinkage, 5) early expansion, 6) old expansion + early shrinkage; and 7) old shrinkage + early expansion (bottleneck) to understand what historical demographic modeling is more accurate for these populations. Pre-evaluation of model scenarios showed four scenarios with values close to zero (scenarios 1, 3, 5, and 6) and three scenarios with probability to describe the observed data (scenarios 2, 4, and 7, Figure IV.2). The higher posterior probability based on our data was identified in scenario 4 (i.e., old population shrinkage), performing as the best fit for our data (logistic approach PP = 0.7847, 95% Cl = 0.6593-0.9101 (Table IV.2). This demographic model corresponds to a population declining its size at t2 and maintaining until now.

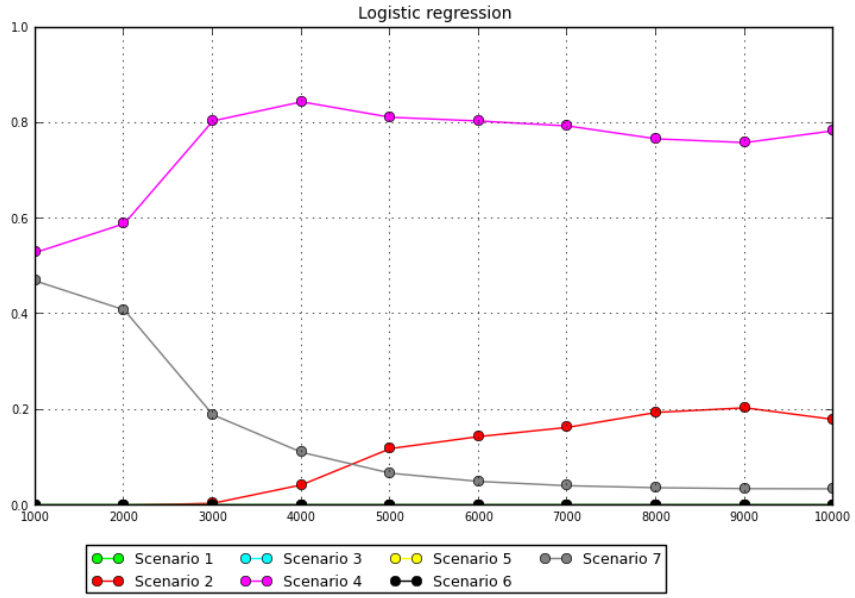


Figure IV.2: Comparative logistic regression for all tested scenarios, showing scenario four as the higher probability.

We performed a model checking analysis to assess the goodness-of-fit of scenario 4, including a principal component analysis (PCA) based on 10,000 simulations indicating the fit between the observed data and the posterior predictive distribution (Figure IV.3). Also, scenario 4 presented relatively low values of Type I and Type II errors (0.296 and 0.431, respectively, Table IV.2), which correspond to the proportion of 1,000 tests in which scenario four is rejected although it is the true scenario, and to the probability of deciding for scenario 4 when it is not the true scenario, respectively (Cornuet et al. 2014).

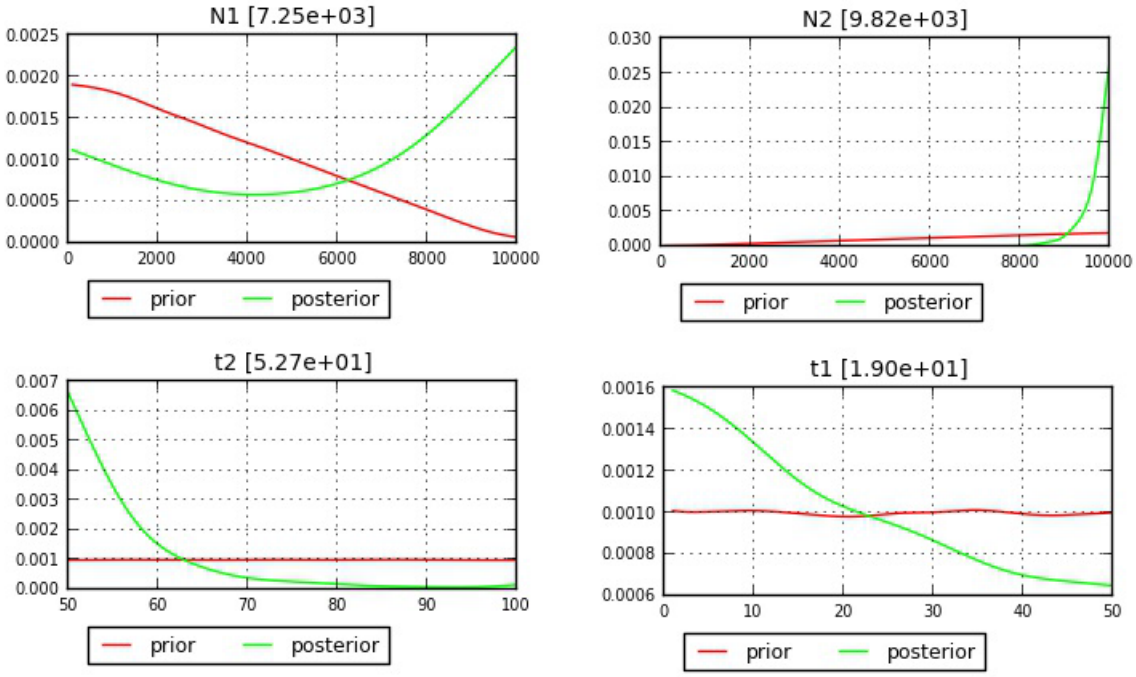


Figure IV.3: Estimate posterior distributions of parameters for scenario 4 based on the number of total simulated data.

Table IV.2: Posterior probabilities of all scenarios. The most probable scenario is highlighted in grey.

	scenarios						
	1	**2**	**3**	**4**	**5**	**6**	**7**
Direct approach	0.000	0.0080	0.000	0.5280	0.000	0.000	0.4640
Logistic approach	0.000	0.1802	0.000	0.7847	0.000	0.000	0.0351
Type I error	-	-	-	0.297	-	-	-
Type II error	-	-	-	0.431	-	-	-

Taken together, both analysis and error checking agree that scenario 4 is best describing the population history of *X. grisescens*. The mean value of probable current population effective size (NA) was 6.170 UN, whereas the ancestral population (N1) seemed to have 9.700 UN. The

older event (t2) occurred about 56 generations ago, and the earlier event (t1) occurred 21 generations ago (Table IV.3)

Table IV.3: Demographic parameters and posterior distributions of our data.

Parameter	Mean	Median	Mode	q025	q050	q250	q750	q950	q975
N1	9700	9820	9990	8720	9010	9560	9940	9990	10000
NA	6170	7250	9910	185	320	2830	9450	9920	9950
T2	56.5	52.7	50	50	50	50.7	58.5	76.4	86.1
T1	21	19	7.19	1.65	2.31	8.27	31.8	45.8	47.6

Discussion

We examined SNP data from a Brazilian carpenter bee dwelling in the world's most populated semiarid area to shed unprecedented light on their demographic history and population size. Our demographic modeling approach revealed a historical population decline, suggesting these populations experienced a possible habitat constraint about ~60 generations ago and stabilized their size until now. Considering carpenter bees to reach up to two or three years of active life (Silva and Freitas 2018), we assume such event to have occurred about ~120–180 years ago.

Historically, the Caatinga ecoregion has been chronically impacted since the Brazilian colonization by Europeans on the sixteenth century and have increased during the high human migration on the eighteenth and nineteenth century (Andrade et al. 2005; da Silva et al. 2018). Over time, human-induced degradation such as cattle-ranching, timber removal, and slash-and-burn agriculture (da Silva et al. 2018) have potentially been responsible for reducing *X. grisescens* population size. In fact, cattle supply was the economic force pushing BSA's initial occupation, and it is still the most important single land use in the region (Sampaio et al., 2009).

Specifically, the state of Ceará, where we collect our data, was mainly unexplored until the end of the seventeenth century (Jucá Neto 2009). From the first half of the eighteenth

century, the Ceara territory became highly explored by livestock activities and related products (Jucá Neto 2009). As the population increased, it became a problem to transport cattle and its related products throughout the region, so crops were established as a primary subsistence production, restricted to small farms (Sampaio *et al.*, 2018). Until close to the end of the eighteenth century the production system remained the same, exploring a larger area due to the natural expansion of the domestic animal and human populations, which increased five to ten times across each century (Souza 1979). Based on these historical records, we believe the boom of human migration from the past two has shaped populations of *X. grisescens*, making it decrease due to the boom, but then stabilizing because the occupancy embedded continuous/constant throughout the years, slightly affecting the landscape. Yet, we deem they also continue to fluctuate stochastically.

According to (Traill, Bradshaw e Brook, 2007), the minimum viable population (MVP) size is, on average, ~4,100 individuals. Thus, from a conservation standpoint, although population size was reduced (from ~9.000 to ~6,000 UN), it stabilized to an estimated size which is maintaining these populations viable. The MVP concept considers the levels of genetic diversity necessary to maintain adaptation and evolution in a species (Shaffer, 1981). According to Traill et al. (2007), to consider the number of individuals in a population that declined deterministically (or catastrophically) to a smaller size but then stabilized, is an appropriate measure of its viability. Similarly, Frankham (1995) recommended affective population size to be an average of 4500 UN based on genetic data. Considering these concepts, we understand that at this timeframe (two to three centuries), *X. grisescens* populations are securely maintaining viable populations according to the number of individuals and its genetic diversity.

Considering that MVP is highly considered on conservation management programs (see Traill et al. 2007), including the IUCN's red list criteria (www.iucnredlist.org), our study is valuable as a starting point for future research in extinction risk and population changing hypothesis in *X. grisescens* and other large carpenter bees. Even though our results showed that our focal bee stabilized their population size, we understand it is of high necessity to apply these results on a real-world scenario and to consider other ecological predictors such as wider studies on the genera, related landscape elements analysis, and/or environmental adaptation. Intrinsically, one needs specific factors such as variability of the local environment to determining the ideal population size.

CAPÍTULO V

How is landscape modification affecting the gene flow among populations of the large carpenter bee Xylocopa grisescens?

A pesquisa feita neste capítulo representa uma colaboração entre a autora, a Dra. Francisca Soares de Araújo, a Dra. S. Hollis Woodard, Dra, Christiana Faria e o Dr. Danilo Boscolo.

Revista de interesse: Evolutionary Applications | ISSN:1752-4571 (online)

A ideia principal deste capítulo é continuar a investigação sobre como os dados populacionais genômicos são moldados pela desertificação. O foco deste capítulo é investigar as possíveis causas ecológicas para a diferença existente nas variáveis genéticas entre nossos oito locais de amostragem levando em consideração a estrutura e a composição da paisagem.

Introduction

Gene flow is a natural process responsible for enabling some alleles to spread throughout different populations (Slatkin 1985). Successful gene flow occurs when individuals disperse and subsequently reproduces, moving alleles between populations as a result of active random mating (Waits and Storfer 2016). It can be affected by many factors, including habitat loss, local management practices, and geographical and reproductive barriers (Fahrig e Nuttle, 2007). These barriers draw different genetic effects on populations being either negative when mating of highly related individuals generate inbreeding and accumulation of deleterious alleles, or positive when population differentiation increase alleles adapted to environmental variation (Charlesworth and Willis 2009; Slatkin 1975, Sexton et al. 2011).

A general prediction over gene flow among populations is that individuals with higher dispersal potential would be less affected. Dispersal over large areas increases chances of random mating and colonization of new habitats, increasing chances of survival in a population. If gene flow is high among populations from different environments (e.g., high dispersing individuals), it roughly indicates these populations are not isolated by geographical distances or environmental barriers. More precisely, these populations would have less locally adapted alleles because no infusion of new alleles is being provided, although they would also be less prone to extinction for the same reason (Sexton et al., 2013). Although this assumption is real, many widespread or cosmopolitan species show some level of adaptation to their habitat landscape.

The structure and composition of the landscape are key drivers of population distribution, local adaptation, and gene flow (Cushman, McRae, and McGarigal, 2016). Much of the quantity of gene flow a population experiences can be estimated by analyzing its genetic differentiation against its habitat information (Waits and Storfer, 2016). In altered habitats with a recent shrunk provision of food and nesting resources, small-bodied animals may face establishment constraints at a finer scale, being forced to either search for more suitable habitats or deal with mating scarcity (cit,). Such search for more suitable habitats can be revealed by assessing the number of private alleles in a population. The number of private alleles (nPa) is a measure of how many unique alleles a given population carries (Slatkin 1985b). It can be used for estimating the level of gene flow in a subdivided population, where the highest values of nPa mean high gene flow. This assumption relies on the difficulty of rare alleles to be carried out if there is not frequent individual dispersal (Slatkin 1985b).

The large carpenter bee *Xylocopa grisescens* is endemic from the Neotropics (Moure 2012) and currently inhabits the most degraded semi-arid region in the world, the Brazilian semiarid region (BSA). This species is potentially being affected by climate change together with anthropogenic exploitation that has been reducing its original habitat, and it is expected to lose up to 35% of suitable habitat in the coming decades (Bezerra et al. 2019). *X. grisescens* is an important pollinator of native plants and crops in the BSA, being responsible for maintaining plant species. There is particular interest in understanding how populations of important pollinators are coping with land degradation given its importance worldwide on produces and croplands and in maintaining native vegetation.

To understand how our focal bee is thriving the land degradation effects ongoing on the Brazilian semiarid region, one should gather information on different levels of land degradation in its main habitat and measure levels of gene diversity and gene flow values. Given this, here we assess the quantity of gene flow these bees are experiencing by measuring the number of private alleles in eight populations dwelling in the most degraded semi-arid in the world. We aim to answer if the number of private alleles is somehow related to the amount of habitat *X. grisescens* rely on. We hypothesize that 1) in areas with lesser suitable habitat, there will be a higher number of private alleles; and that 2) Individuals that do not migrate a lot (lower number of private alleles) present a higher ability to adapt to their local site. We predict that, because of the resource's scarcity, individuals do not stay for a time long enough to keep a record in its genome and end up dying or migrating to other more favorable habitats, showing a higher number of private alleles; likewise, individuals that stayed for a longer period in more suitable habitat, will show more outlier loci related to adaptative environmental conditions.

Methods

Study site and sample collection

The Brazilian semi-arid region (BSA) is considered the most degraded semiarid in the world, corresponding to over 40% of habitat loss. It is located between 2.5°S and 16.1°S and 34.8°W and 46°W, encompassing a total area of ~1,500,000 km2 covering 18.26 % of the total area of Brazil (Magalhães et al. 1988). The predominant vegetation is called Caatinga, where the annual rainfall index varies from 300 to 600 mm (FUNCEME, 2016), and the mean annual temperature is between 26° and 28° (Nimer, 1972). Water availability is expressed as short rainfall pulses concentrated in 3 to 5 months per year (Nimer, 1972).

The study was conducted in two large areas classified as susceptible to desertification in separated sites of Ceará State in the BSA. These areas have the worst historical anthropogenic vegetation removal (Andrade et al., 2005) and include one of the most populous semiarid areas in the world. We separated two different clines in ASDI and ASDII after the Brazilian Classification of Areas Susceptible to Desertification (ASD). ASD I comprises four sample sites SO, IRA, SQI, SQII, whereas ASDII also comprises four samples sites hereafter named MN, JM, JGI, JGII (Figure III.1 and V.II). Eighty females of *X. grisescens* were collected in each location (ten for each site) using a sweep net. After collection, the bees were immediately preserved in 95° ethanol and stored at -20°C. Geolocation data for sampling points are provided in table V.1.

Table V.1 Geolocation data for sampling points. Acronyms for Irauçuba, Jaguaretama, Jaguaribe, Morada Nova, Santa Quitéria, Sobral: IRA, JM, JGI, JGII, MN, SQI, SQII, SO, respectively.

ASD	**Acronyms**	**Locality**	**x**	**y**
I	IRA	Irauçuba	-39.88466	-3.72628
I	SO	Sobral	-40.182301	-3.64164
I	SQI	Santa Quitéria	-40.079976	-4.31299
I	SQII	Santa Quitéria	-40.187629	-4.44283
II	JGI	Jaguaribe	-38.768824	-5.8308
II	JGII	Jaguaribe	-38.65274	-5.92564
II	JM	Jaguaretama	-38.73726	-5.56964
II	MN	Morada Nova	-38.50815	-5.32244

DNA extraction

For DNA extraction, we used the Qiagen DNeasy Blood and Tissue kit. This procedure includes 5 steps: preparation, dissection, digestion, purification, and elution. Before starting the procedure, each bee was removed from the tube containing ethanol and completely dried with sterile filter paper. Then, the removal of the leg of each bee was performed using sterile dissecting tools flamed with ethanol. In the digestion step, each leg is placed individually in a 1.7ml Eppendorf tube and the biological content is macerated with the help of a sterile plastic pistil and liquid nitrogen. Each tube received 180 µl of ATL buffer followed by 20 µl of Proteinase K, being incubated at 56°C overnight for tissue breakage. For purification, we added 200 µl of AL buffer and 200 µl of 100% ethanol. Then, 500 µl of buffer AW1 was added followed by centrifugation. For elution, we added 30 µl of elution buffer AE.

RADseq library preparation

The first step in preparing the RADseq libraries was the quantification of genetic material from 10% of the samples using the Qubit 3.0 protocol (Life technologies). Then, libraries were prepared using the protocol described by Brelsford et al. (2016) with minor modifications (Figure V.1). First, we performed double digestion of the genomic DNA by the restriction enzymes *MseI* and *EcoRI* (Illumina Inc.), followed by the annealing of two unique adapters for each individual bee (Forward and Reverse) at the edges of the DNA fragments. After ligation, the sample was purified with Agencourt AMPure (Beckman Coulter, Inc., Brea, CA, USA) to remove small DNA fragments. Then, the 80 samples were multiplexed in a lane and sent for sequencing with the company Novogene Corporation. Samples were sequenced using the Illumina HiSeq X Ten paired-end 150 bp platform.

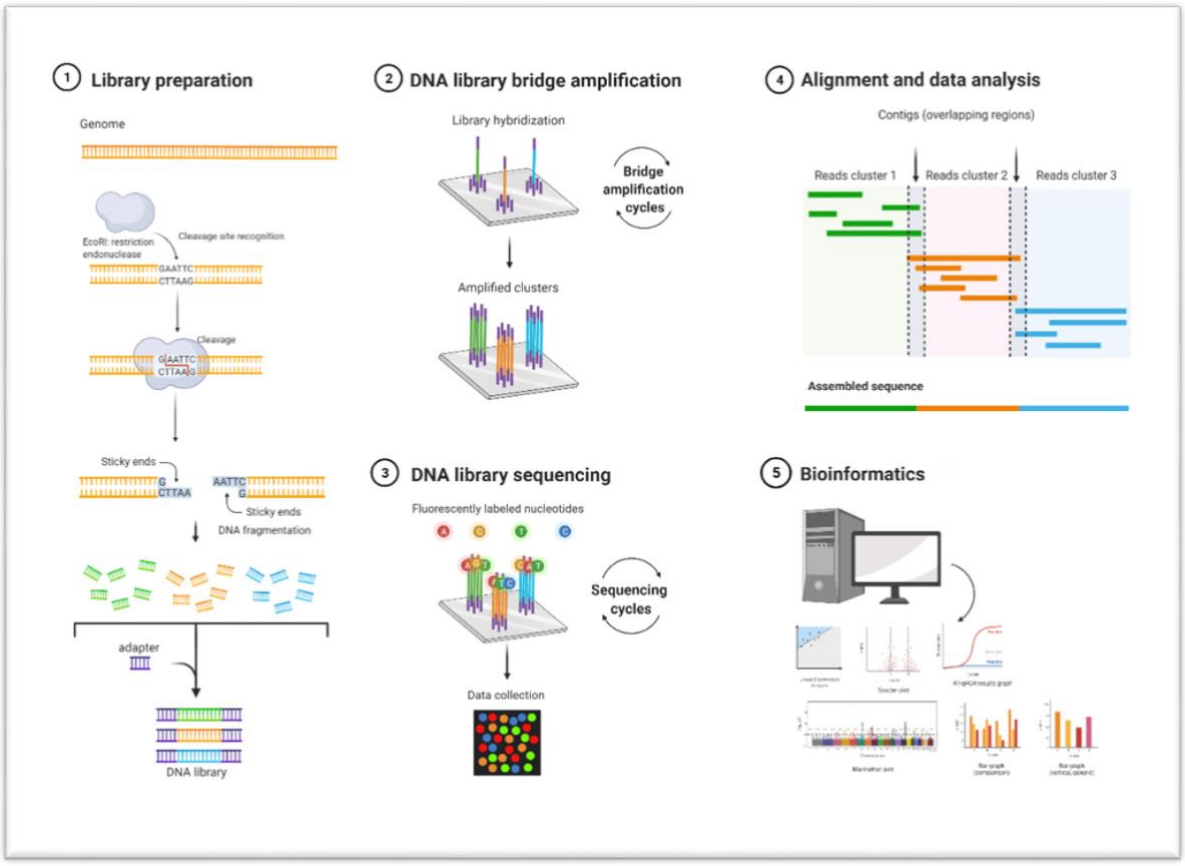


Figure V.1. workflow of RADseq from library preparation to the downstream bioinformatic analysis

Downstream analysis and filtering

The raw data were demultiplexed by barcode using the PROCESS_RADTAGS tool in the STACKS v2.3b program (Catchen et al., 2013). Sequences of reads were filtered to remove any reads with an unnamed base, discard low-quality reads as defined by default settings, and retrieve barcodes with up to an expected barcode error (codes provided). Also, the upper bound for the sequencing error rate was set at 0.05. After checking the quality of the reads, we used the DENOVO_MAP pipeline in STACKS to reassemble all population loci. We then used POPULATIONS tools to filter for loci present in at least 75% of individuals, a quality score > 20, and a minimum minor allele frequency (MAF) > 0.1. We also used POPULATIONS to generate statistics regarding the number of private alleles at the population level.

Environmental variables

Environmental variables from each point were represented by the bioclimatic variables obtained from the WorldClim platform (period: 1950-2010). This procedure was performed using the "raster" package v.2.9-5 (Hijmans 2019) in R (R CoreTeam, 2015). We also collected in-point elevation data. Highly correlated variables were excluded from the analysis, and we kept only five variables: Altitude (ALT), precipitation of the coldest quarter (PCQ), precipitation of the warmest quarter (PWQ), precipitation seasonality (PS), and temperature of annual range (TAR).

Table V.2. Sets of bioclimatic variables used in statistical analysis. Bioclimatic variables represent either annual measures or extreme environmental factors. A quarter is a period of three months.

ACRONYM	**VARIABLE**
ALT	altitude
PCQ	precipitation of the coldest quarter
PWQ	precipitation of the warmest quarter
PS	precipitation seasonality
TAR	temperature of annual range

Mapping and Landscape analysis

A land cover classification from collection 5 of the MapBiomas project was used to evaluate the landscape structure. The MapBiomas maps cover the entire Brazilian territory. They were obtained from the classification of LANDSAT image mosaics, with 30m spatial resolution, using a random forest classification algorithm from the Google Earth Engine (Souza et al. 2020). From the 29 original land cover classes of the nationwide map, 23 are present in the studied region including "Forest Formation", "Savanna Formation", "Mangrove", "Forest Plantation", "Wetland", "Grassland", "Other non-Forest Formations", "Pasture", "Agriculture",

"Temporary Crop", "Sugar Cane", "Mosaic of Agriculture and Pasture", "Beach and Dune", "Urban Infrastructure", "Other Non-Vegetated Areas", "Rocky Outcrop", "Mining", "Aquaculture", "Salt Flat", "River and Lake", "Perennial Crop" and "Other Temporary Crops" (Figure V.2).

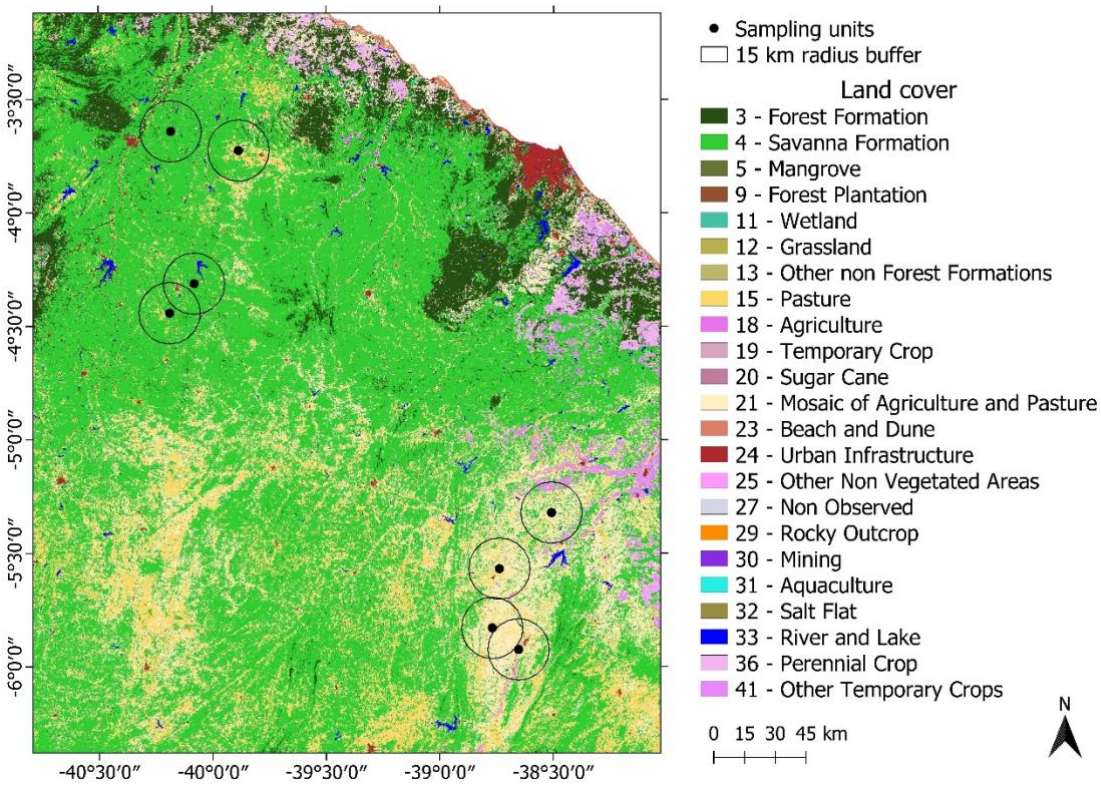


Figure V. 2. Land cover map and distribution of sampling units in the studied region.

From the MapBiomas land cover classification, the classes “Forest Formation,” “Savanna Formation,” and “Grassland” were considered natural and semi-natural environments and used as a proxy for the landscape habitat proportion. All land cover classes were used to calculate the landscape Shannon’s diversity index (McGarigal et al., 2012). The above-mentioned landscape metrics were calculated using circular buffers around the sampling units with a varying radius ranging from 0.5 to 15 km. We then delimited the radius into three scales: local (2 to 3 km from the sampling unit), regional (6 to 8 km from the sampling unit), and broad (12 to 15 km from the sampling unit) and compared which scale is facilitating gene flow. The landscape metrics calculations were performed in the R software version 3.6.3, with the package ‘landscapemetrics’ version 1.5.2 (Hesselbarth et al. 2019).

Detection of outlier SNPs and regions potentially under selection

To assess whether our data show candidate genes under selection, we scanned our entire dataset (83,127 SNPs) for outliers SNPs using two F_{st}-based approaches and one Principal Component Analysis (PCA) approach. First, we used the Fst-based model Bayescan, which implements a Bayesian likelihood method via reversible-jump Monte Carlo Markov chain (RJMCMC). First, we used STACKS v2.3b to generated *–vcf* format files, then the software PGDspider v2.1.1.5 (Lischer e Excoffier, 2012) to convert *–vcf* file format to BayeScan readable format. We used the software BayeScan v2.1 (Foll, 2012) to estimate a model parameter by performing 20 pilot runs with 5,000 iterations each, with a burn-in of 100,000 iterations and a thinning interval of 10. We set the prior odds of 1, which means that we are assuming that the model including selection is as likely as the neutral model (Foll 2012).

For our second F_{st}-based approach, we used the R package Fsthet (Flanagan e Jones, 2017), which identifies outlier loci by comparing F_{st} values relative to the heterozygosity, where extreme F_{st} values are selected. We first parsed genepop format (Rousset, Lopez e Belkhir, 2017) file into R. Then, we calculated allele frequencies, heterozygosity, and F_{st}, compared these parameters with each other, and identified loci lying outside of the deeming threshold. We also tested for SNP outliers using the R package PCAdapt v. 4.3.2 (Luu, Bazin e Blum, 2017). This approach detects outlier loci based on a Principal Component Analysis (PCA), regressing each SNP against each PC by using z-scores and Mahalanobis distance. We used MAF = 0.05 and a more conservative *q*-value threshold of 0.005 as a cutoff.

Statistical analysis

We use redundancy analysis (RDA) in the R package vegan v. (Oksanen *et al.*, 2015) to assess candidate adaptive loci with strong associations to environmental variables. RDA is a multivariate ordination approach to model response variables against the explanatory variables for many loci and environmental predictors simultaneously (Forester *et al.*, 2018). First, we randomly filtered our dataset downing to 8,189 SNPs to account for less time-consuming analysis, afterward filtering for missing data. We ran RDA using the SNPs dataset as the explanatory variable and bioclimatic data as predictor variables (Table V.2). Using the loadings

of SNPs in the ordination space, we determined which SNPs are candidates for local adaptation and identified those that were shared among RDA and the outlier analysis. We then excluded the duplicates of further analysis. We also tested for which environmental variables had the strongest associations with each candidate SNP using an adjusted Pearson's correlation coefficient (*r*).

For the landscape genomics statistical analysis, we assessed the potential relation of the number of private alleles to landscape composition and heterogeneity using GLM (generalized linear models), model: Poisson, link: log, and using the number of private alleles as response variables and landscape metrics as predictor variables. After That, we tested each GLM result with ANOVA (test = Chisq).

Results

Landscape composition and number of private alleles

Our landscape analysis showed a very different percentage of land covers throughout the eight sample sites (Table S1 and S2, ANEXO). In ASD I, the higher percentage of land cover measures was mosaic Agriculture and Pasture cover for all three scales (local, regional and broad) (table S2, ANEXO). In ASD II, the higher values were found in savanna formation for both local, regional, and broad-scale (table S3, ANEXO). The GLM results revealed that the type of habitat cover had some effect on the gene flow, here measured as the number of private alleles (Table V.3).

We detected a total of 651 private alleles throughout our eight sampled populations. A lower number of private alleles was found in IRA (31), followed by JM (67), JGII (72), SQI (72), SO (73), MN (107), JGI (108), and SQII (121). The number of private alleles was positively correlated to rocky outcrops in both local scales ($p<0.05$) and broad-scale ($p<0.001$). At the local scale, nPa was also slightly correlated to Grassland ($p = 0.0517$). On a regional scale, nPa was more related to the pasture cover ($p<0.05$) and slightly correlated to savanna formation ($p = 0.0542$) and the mosaic Agriculture and Pasture ($p = 0.0543$). We also tested for the best explanation of variation in nPa throughout all scales according to the Shannon Diversity Index (SHI). The best fit was both local scale (~3000m) and broad-scale (~10000m, ANEXO).

Table V.3. The relative importance of landscape categories as calculated against the number of private alleles. SE refers to standard error, and values of p highlighted in bold refer to the best fit for each model.

Landscape categories			
Local scale	**Estimate**	**SE**	**p**
Agri and Past	0.09061	0.05622	0.107
Forest	1.31289	0.705	0.0626
Grassland	-0.84319	0.43345	0.0517
Pasture	0.09286	0.0583	0.1112
Rocky outcrops	**1.23552**	**0.62602**	**0.0484***
Savanna	0.08772	0.05507	0.1112
Regional-scale			
Agri and Past	0.00446	0.002341	0.0568
Forest	-0.00633	0.003291	0.0543
Grassland	-0.00853	0.005195	0.1006
Pasture	**0.003442**	**0.001706**	**0.0437***
Rock outcrops	0.073476	0.063323	0.2459
Savanna	0.004624	0.002402	0.0542
Urban	0.022237	0.012528	0.0759
Broad-scale			

Agri and Past	7.23E-05	0.000192	0.707
Forest	-0.00016	0.000211	0.444
Grassland	0.000334	0.000209	0.11
Pasture	0.00019	0.00016	0.236
Rocky outcrops	**-0.0191**	**0.003802**	**0.001****
Savanna	9.26E-05	0.000191	0.628
Urban	-0.00033	0.000561	0.553

*Signif. Codes 0.001 '**' 0.01 '*'*

Outlier SNPs detection

We found four different values of outlier detections, according to each approach. RDA analysis found 227 outlier loci, BayeScan found 18, Pcadapt found 473, and Fsthet found 14 (Figure V.4). None of the three later approaches overlapped in identifying the loci. We then tested all loci outlier to detect *c*andidate SNPs putatively under selection. First, we tested the correlation of candidate loci to all sample sites according to the environmental variables (ALT, PCQ, PWQ, PS, and TAR). A higher correlation was found in SQ2 with PWQ and to PCQ in JM and G1 (Figure V3.a). Overall, candidate SNPs on axis 1 represent SNPs associated with precipitation of the warmest quarter and to precipitation of the coldest quarter, whereas axis 3 represent SNPs associated with altitude, precipitation seasonality, and temperature (Figure V.3b).

We also analyzed the outlier loci against the bioclimatic variables. RDA analysis revealed 33 loci associated with ALT, 30 to PCQ, 133 to PS, 35 to PWQ, and 6 to TAR (Figure V.4). The five environmental variables explained 24% of variance within the SNP data (adj.r.squared = 0.2430323). SNPs most strongly correlated with precipitation seasonality (PS) showed higher loadings around both axis 1 and 2, as well as the second most correlated

precipitation of the warmest quarter (PWQ, Figure 4). The other SNPs related were more centered throughout the axis.

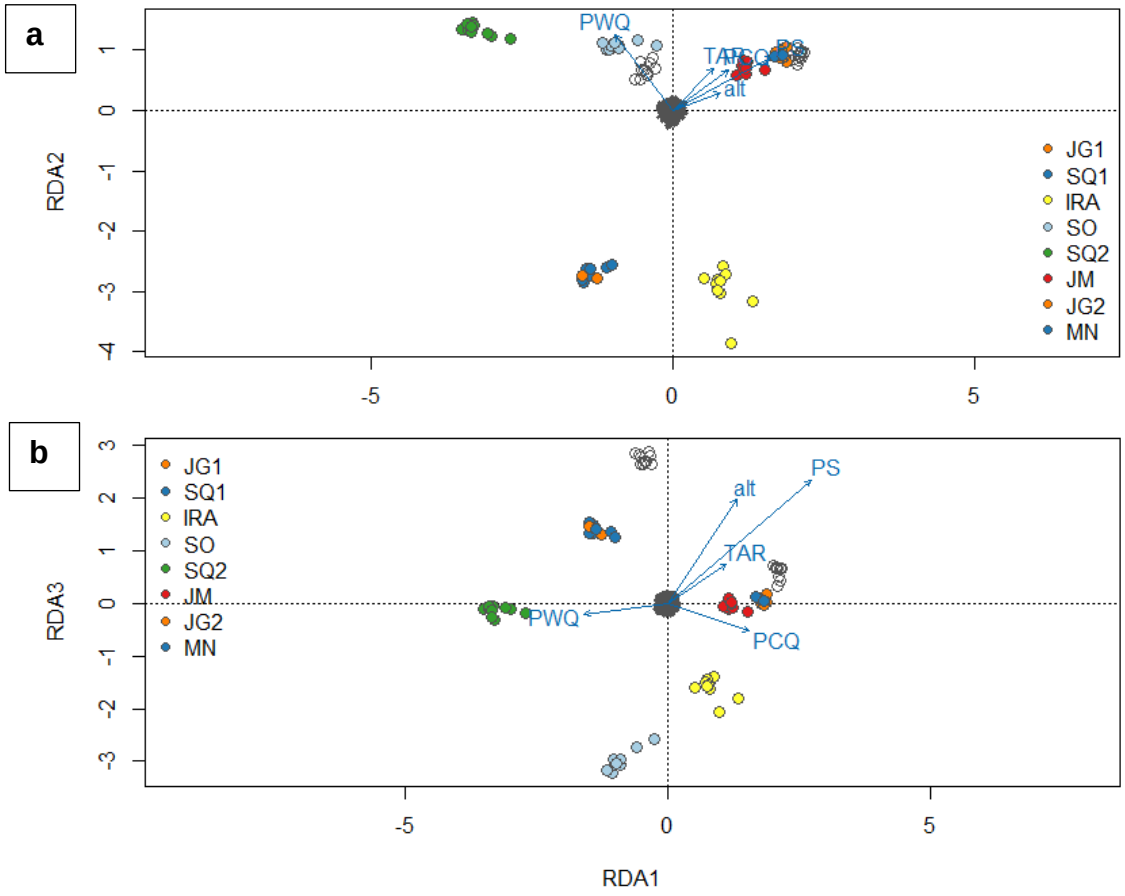


Figure V.3. RDA (Redundancy analysis) on a dataset of over 8,000 neutral loci. a) axis 1 and 2 of the RDA showing all eight populations sampled for this study and the variables correlated. b) axis 1 and 3 of the RDA. Explanatory variables (arrows): ALT (altitude), PCQ (precipitation of the coldest quarter), PS (precipitation seasonality), PWQ (precipitation of the warmest quarter), and TAR (temperature of annual range). Locations: ASDI (JM, MN, JGI, JGII) and ASDII (SQI, SQ2, IRA, SO).

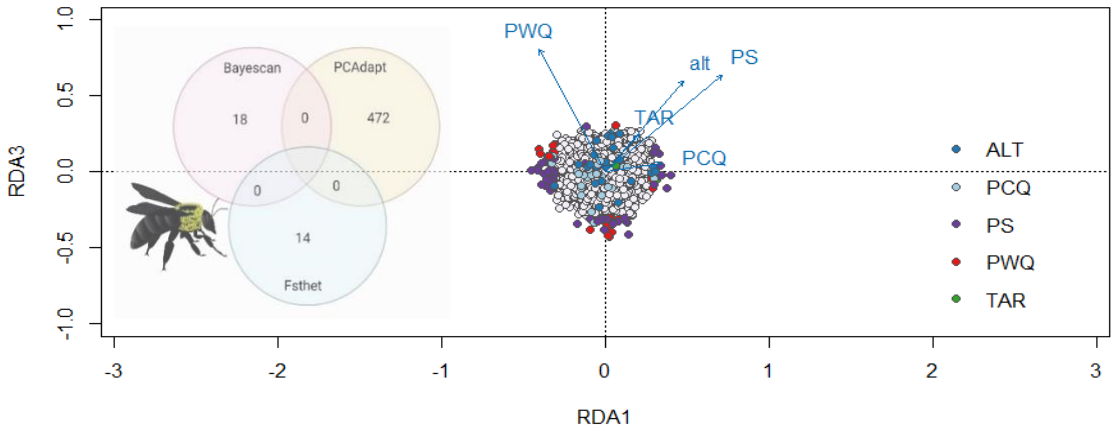


Figure V.4. RDA (Redundancy analysis) and Venn diagram of loci under selection detected by three approaches on a dataset of over 8,189 neutral loci. SNP genotypes are in grey. Axis 1 and 3 of the RDA showing all the environmental variables used in this study. Explanatory variables (arrows): ALT (altitude), PCQ (precipitation of the coldest quarter), PS (precipitation seasonality), PWQ (precipitation of the warmest quarter), and TAR (temperature of annual range).

Discussion

Here, we evaluated the relation between landscape composition (here measure as the habitat and non-habitat amount) and the gene flow for eight populations in a highly degraded and fragmented semiarid area. Our data supports high gene flow between populations of the large carpenter bee *X. grisescens* as related to some types of vegetation cover. Also, based on the strongest correlations, most SNPs were associated with our two precipitation variables (precipitation seasonality and precipitation of the warmest quarter), followed by altitude, precipitation of the coldest quarter, and temperature.

Higher levels of the number of private alleles are indicated of high gene flow (Slatkin 1985b). The rationale underlying this is that, for a given population to have a high number of private alleles, it must have a lot of migrants coming and introducing new alleles in this population (Slatkin 1985). Also, to achieve longer distances, migrants need a habitat that supports such long flights. It seems that even in a highly fragmented landscape, our sampled individuals are receiving enough resources to either forage for longer distances or stay still. We found no barriers to gene flow in our dataset, although we did find a strong relation of landscape composition to the gene flow for some locations. The effects of habitat fragmentation can be

reduced by the high dispersal abilities of individuals in a population (Cane 2001). Looking only at fragmentation and gene flow, we presume that *X. grisescens* is counteracting fragmentation negative effects due to its large foraging range.

Our scale-dependent analysis revealed that at both broad and local scales, individuals of X. *grisescens* are mostly affected by rocky outcrops. Broad-scale had negative effects whereas local scale had positive. In light of the species biology, we would expect such a negative effect because rocky outcrops severely down food and nesting opportunities for this species, as *X. grisescens* is a wood-nesting and a resource generalist bee. The lower value of nPa was in IRA, which is the landscape with the higher number of rocky outcrops. This indicates that the number of migrants is effectively low due to lower available resources, while individuals search for more rewardable areas even over large areas. There is some evidence that wood-nesting solitary or primitively social bees are more likely to cross soil barriers (water and rocks) than are those that nest in the ground (Michener 2007). Our regional scale results suggest that pasture is positively affecting the gene flow. In the rainy season of caatinga vegetation, the predominant vegetation in BSA, there is a rich herbaceous cover. Specifically, a plant species highly present in pastures of the sampling points is *Senna obtusifolia* (L.) H. Irwin & Barneby (*author's information)*, which is one of the ten most preferred plan species pollinated by *X. grisescens* (Inês Da Silva, 2009).

We observed a higher number of outlier SNP loci related to the precipitation seasonality. Precipitation seasonality is a measure of the irregular distribution of rainfall during a normal year, meaning that that most of the rainfall occurs in specific months (Hijmans *et al.*, 2005). The Brazilian semiarid region has a spatially and temporally irregular precipitation, usually concentrated from January through May (Nimer, 1989). For the rest of the year, the region has a water deficit that can vary between 7 and 11 months (Souza 1994). During the rainy season, the BSA has a flowering bloom where plant species present highly differentiated growth forms like trees, shrubs, grass/herbs, and lianas (Costa, Araújo e Lima-Verde, 2007).

Our findings provide new insights into the genetic conservation of the large carpenter bee *X. grisescens*. We provided a framework for assessing the relation between gene flow and landscape composition and how it is affecting their populations. One should be careful, though, with the extrapolation of our results to other bee faunas experiencing habitat fragmentation because it must be considered both current habitat availability and historical land use, as suggested by Cane et al. (2006).

CONCLUSÕES

Este trabalho abordou a genômica de paisagem de uma importante abelha para o semiárido brasileiro. Os dados aqui encontrados sugerem que as populações de *X. grisescens* sofreram uma redução em seu tamanho populacional em algum momento de suas gerações (provavelmente há 150 anos), mas que, atualmente, o seu tamanho populacional é viável para manter suas populações sem o risco de extinção. Também, encontramos que existe certo nível de diversidade genética entre as populações, mas que essa diversidade não é explicada pelas distâncias geográficas entre cada paisagem. O que mais parece favorecer a manutenção das populações de *X. grisescens* no semiárido cearense é o mosaico de plantas nativas e de culturas agrícolas, juntamente com o estrato herbáceo que cobre os pastos durante o período de chuvas. Paisagens formadas por muitos afloramentos rochosos, ao contrário, dificultam o fluxo gênico, como é o caso de Irauçuba (IRA), município com maior quantidade de afloramentos rochosos nas nossas análises e o mais degradado do Ceará.

Os dados obtidos com esse estudo trouxeram luz à lacuna de conhecimento que existe sobre a genômica populacional em regiões semiáridas extremamente degradadas e sobre a adaptação de polinizadores generalistas em áreas degradadas. É importante ressaltar que a obtenção e a análise dos dados só foi possível graças ao fortalecimento das técnicas de genômica de paisagem. Seguindo essa abordagem de forma pioneira, conseguimos entender como as paisagens semiáridas e o clima semiárido influenciam a diversidade genética, a distribuição da diversidade genética, a adaptação e o fluxo gênico das populações de *X. grisescens*.

Esse conhecimento pode ser útil para o desenvolvimento de outros modelos ou impulsionar novos trabalhos para entender como essa ou outras espécies podem responder a mudanças no seu habitat no futuro, pois durante muito tempo, a conservação foi voltada para a manutenção da maior quantidade de habitat, mas mostramos aqui que a melhor abordagem é a prevalência do habitat que melhor funciona para cada espécie.

REFERÊNCIAS

ALLENDORF, F. W.; HOHENLOHE, P. A.; LUIKART, G. Genomics and the future of conservation genetics. **Nature Reviews Genetics 2010 11:10**, v. 11, n. 10, p. 697–709, 17 set. 2010.

ANDREW, N. R. *et al.* Assessing insect responses to climate change: What are we testing for? Where should we be heading? **PeerJ**, v. 1, n. 1, p. e11, 12 fev. 2013.

ANDREWS, K. R. *et al.* Harnessing the power of RADseq for ecological and evolutionary genomics. **Nature Reviews Genetics**, v. 17, n. 2, p. 81–92, 2016.

ANGELO, C. DE *et al.* Understanding species persistence for defining conservation actions: A management landscape for jaguars in the Atlantic Forest. **Biological Conservation**, v. 159, p. 422–433, 2013.

ANTONGIOVANNI, M.; VENTICINQUE, E. M.; ROBERTO, C. Fragmentation patterns of the Caatinga drylands. **Landscape Ecology**, 2018.

ARAÚJO, T. N. *et al.* Individual-resource network between <scp> *Xylocopa* </scp> bees and plant resources: generalist species, specialist individuals? **Ecological Entomology**, p. een.13072, 7 ago. 2021.

BALKENHOL, N. *et al.* **Landscape genetics : concepts, methods, applications**. [s.l.] John Wiley & Sons Ltd Registered, 2016. v. 1

BALLARE, K. M.; JHA, S. Genetic structure across urban and agricultural landscapes reveals evidence of resource specialization and philopatry in the Eastern carpenter bee, Xylocopa virginica L. **Evolutionary Applications**, n. February, p. 1–14, 2020.

BEAUMONT, M. A. Adaptation and speciation: What can Fst tell us? **Trends in Ecology and Evolution**, v. 20, n. 8, 2005.

___. Approximate Bayesian computation in evolution and ecology. **Annual Review of Ecology, Evolution, and Systematics**, v. 41, 2010.

BEICHMAN, A. C.; HUERTA-SANCHEZ, E.; LOHMUELLER, K. E. **Using genomic data to infer historic population dynamics of nonmodel organismsAnnual Review of Ecology, Evolution, and Systematics**Annual Reviews Inc., , 2 nov. 2018.

BEZERRA, A. D. M. *et al.* Agricultural area losses and pollinator mismatch due to climate changes endanger passion fruit production in the Neotropics. **Agricultural Systems**, v. 169, n. 2019, p. 49–57, 2019.

BOHONAK, A. J. THE QUARTERLY REVIEW OF BIOLOGY DISPERSAL, GENE FLOW, AND POPULATION STRUCTURE A CENTRAL challenge for organismal. **The Quaterly review of Biology**, v. 74, n. 1, p. 21–45, 1999.

BOREVITZ, J. O. *et al.* Tansley review Genomic variation across landscapes : insights and applications. p. 953–967, 2015.

BOSCOLO, D. Da matriz à matiz : em busca de uma abordagem funcional na Ecologia de Paisagens. p. 157–187, 2016.

BROECK, A. VANDEN *et al.* Gene flow and effective population sizes of the butterfly Maculinea alcon in a highly fragmented, anthropogenic landscape. **Biological Conservation**, v. 209, p. 89–97, 2017.

BROQUET, T.; PETIT, E. J. Molecular estimation of dispersal for ecology and population genetics. **Annual Review of Ecology, Evolution, and Systematics**, v. 40, n. December, p. 193–216, 2009.

CANE, J. H. *et al.* Complex responses within a desert bee guild (Hymenoptera: Apiformes) to urban habitat fragmentation. **Ecological Applications**, v. 16, n. 2, p. 632–644, 2006.

CARRIL, O. M. *et al.* Wild bees of Grand Staircase-Escalante National Monument: richness, abundance, and spatio-temporal beta-diversity. **PeerJ**, v. 6, n. 11, p. e5867, 7 nov. 2018.

CGEE. **Desertificação, degradação da terra e secas no Brasil**. Brasilia -DF: [s.n.].

CHAPPELL, M. A. . Temperature Regulation of Carpenter Bees (Xylocopa californica) Foraging in the Colorado Desert of Southern California Published by : The University of Chicago Press . Sponsored by the Division of Comparative Physiology a. v. 55, n. 3, p. 267–280, 1982.

CHARLESWORTH, D.; WILLIS, J. H. **The genetics of inbreeding depressionNature Reviews Genetics**, 2009.

CORNUET, J. M. *et al.* DIYABC v2.0: A software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. **Bioinformatics**, v. 30, n. 8, p. 1187–1189, 2014.

COSTA, R. C. DA; ARAÚJO, F. S. DE; LIMA-VERDE, L. W. Flora and life-form spectrum in an area of deciduous thorn woodland (caatinga) in northeastern, Brazil. **Journal of Arid Environments**, v. 68, n. 2, p. 237–247, 2007.

CUSHMAN, S. A.; H. MCRAE, B.; MCGARIGAL, K. Basics of Landscape Ecology: An Introduction to Landscapes and Population Processes for Landscape Geneticists. *Em*: **Landscape Genetics: Concepts, Methods, Applications**. [s.l: s.n.]. .

DANFORTH, B. N. Emergence dynamics and bet hedging in a desert bee, Perdita portalis. [s.d.].

DARWIN, C. **On the Origin of Species, 1859**. [s.l.] Routledge, 1859.

FAHRIG, L. Effects of Habitat Fragmentation on Biodiversity. **Annual Review of Ecology, Evolution, and Systematics**, v. 34, p. 487–515, 2003.

___. **Non-optimal animal movement in human-altered landscapesFunctional Ecology**, 2007.

___. Ecological Responses to Habitat Fragmentation per Se. **Annual Review of Ecology, Evolution, and Systematics**, v. 48, 2017.

___. Is habitat fragmentation bad for biodiversity? **Biological Conservation**, v. 230, n. October 2018, p. 179–186, 2019.

FAHRIG, L.; NUTTLE, W. K. Population Ecology in Spatially Heterogeneous Environments. *Em*: **Ecosystem Function in Heterogeneous Landscapes**. [s.l: s.n.]. .

FILGUEIRAS, B. K. C.; IANNUZZI, L.; LEAL, I. R. Habitat fragmentation alters the structure of dung beetle communities in the Atlantic Forest. **Biological Conservation**, v. 144, n. 1, p. 362–369, 2011.

FISCHER, J.; LINDENMAYER, D. B. **Landscape modification and habitat fragmentation: A synthesisGlobal Ecology and Biogeography**, 2007.

FLANAGAN, S. P.; JONES, A. G. Constraints on the F ST -Heterozygosity Outlier Approach. **Journal of Heredity**, v. 108, n. 5, 2017.

FOLL, M. BayeScan v2.1. **Ecology**, v. 20, n. 11, 2012.

FORESTER, B. R. *et al.* Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. **Molecular Ecology**, v. 27, n. 9, 2018.

FRANKHAM, R. **Genetic Management of Fragmented Animal and Plant Populations**. [s.l: s.n.]. v. 53

___. Conservation Genetics. **Encyclopedia of Ecology**, p. 382–390, 1 jan. 2019.

FRANKHAM, R.; BALLOU, J. D.; BRISCOE, D. A. **Fundamentos de Genética da Conservação**. [s.l: s.n.].

FRANKHAM, R.; BALLOU, J. D.; BRISCOE, D. A. Loss of genetic diversity in small populations. *Em*: **Introduction to Conservation Genetics**. Cambridge: Cambridge University Press, 2010. p. 231–259.

FRANKLIN, I. R.; FRANKHAM, R. How large must populations be to retain evolutionary potential? **Animal Conservation**, v. 1, n. 1, 1998.

FRANZÉN, M.; NILSSON, S. G. Both population size and patch quality affect local extinctions and colonizations Subject collections Articles on similar topics can be found in the following collections. n. December, 2014.

FREIRIA, G. A. *et al.* Population structure and genetic diversity of the orchid bee Eufriesea violacea (Hymenoptera, Apidae, Euglossini) from Atlantic Forest remnants in southern and southeastern Brazil. **Apidologie**, v. 43, n. 4, 2012.

FREITAS, B. M.; OLIVEIRA FILHO, J. H. DE. Ninhos racionais para mamangava (Xylocopa frontalis) na polinização do maracujá-amarelo (Passiflora edulis) Rational. **Ciência Rural, Santa Maria**, v. 33, n. 6, p. 1135–1139, 2003.

GIANNINI, T. C. *et al.* Pollination services at risk : Bee habitats will decrease owing to climate change in Brazil Pollination services at risk : Bee habitats will decrease owing to climate change in Brazil. **Ecological Modelling**, v. 244, n. October 2017, p. 127–131, 2012.

GIANNINI, T. C. *et al.* Protecting a managed bee pollinator against climate change: strategies for an area with extreme climatic conditions and socioeconomic vulnerability. **Apidologie**, v. 48, n. 6, p. 784–794, 2017.

GOSLEE, S. C.; URBAN, D. L. **The ecodist package for dissimilarity-based analysis of ecological data**. [s.l: s.n.]. v. 22

GREENLEAF, S. S. *et al.* Bee foraging ranges and their relationship to body size. **Oecologia**, v. 153, n. 3, p. 589–596, 2007.

HEINRICH, B.; ESCH, H. Thermoregulation in bees. **American Scientist**, v. 82, n. 2, p. 164–170, 1994.

HIJMANS, R. J. *et al.* Very high resolution interpolated climate surfaces for global land areas. **International Journal of Climatology**, v. 25, n. 15, p. 1965–1978, 2005.

HOBBS, R. J. *et al.* Novel ecosystems: Theoretical and management aspects of the new ecological world order. **Global Ecology and Biogeography**, v. 15, n. 1, p. 1–7, 2006.

HOFFMANN, A. A.; WILLI, Y. Detecting genetic responses to environmental change. **Nature Reviews Genetics**, v. 9, n. 6, p. 421–432, 2008.

HUSEMANN, M. *et al.* Effective population size in ecology and evolution. **Heredity**, v. 117, n. 4, p. 191–192, 2016.

INÊS DA SILVA, C. DISTRIBUIÇÃO ESPAÇO-TEMPORAL DE RECURSOS FLORAIS UTILIZADOS POR ESPÉCIES DE XYLOCOPA (HYMENOPTERA, APIDAE) E INTERAÇÃO COM PLANTAS DO CERRADO SENTIDO RESTRITO NO TRIÂNGULO MINEIRO. 2009.

IUCN. **The IUCN Red List of Threatened Species. Version 2020-3https://www.iucnredlist.org. Downloaded on 12 January 2021.**, 2021.

JUCÁ NETO, C. R. Primórdios Da Rede Urbana Cearense. **Mercator**, v. 8, n. 16, p. 77–102, 2009.

KLEIN, A. *et al.* Insect pollination of crops in Brazil. n. February, 2020.

KLINE, O.; JOSHI, N. K. Mitigating the effects of habitat loss on solitary bees in agricultural ecosystems. **Agriculture (Switzerland)**, v. 10, n. 4, 2020.

LANFEAR, R.; KOKKO, H.; EYRE-WALKER, A. Population size and the rate of evolution. **Trends in Ecology and Evolution**, v. 29, n. 1, p. 33–41, jan. 2014.

LEAL, I. R. *et al.* Effects of habitat fragmentation on ant richness and functional composition in Brazilian Atlantic forest. **Biodiversity and Conservation**, v. 21, n. 7, p. 1687–1701, 2012.

LEAL, I. R.; TABARELLI, M. Ecologia e Conservação da Caatinga. [s.d.].

LENORMAND, T. **Gene flow and the limits to natural selectionTrends in Ecology and Evolution**, 2002.

LEVINS, R. Some Demographic and Genetic Consequences of Environmental Heterogeneity for Biological Control1. **Bulletin of the Entomological Society of America**, v. 15, n. 3, p. 237–240, 1 set. 1969.

LI, Y. *et al.* Ten Years of Landscape Genomics : Challenges and Opportunities. v. 8, n. December, p. 1–7, 2017.

LINSLEY, E. G. The ecology of solitary bees. **Hilgardia**, v. 27, n. 19, p. 543–599, out. 1958.

LISCHER, H. E. L.; EXCOFFIER, L. PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. **Bioinformatics**, v. 28, n. 2, 2012.

LÓPEZ-URIBE, M. M.; JHA, S.; SORO, A. A trait-based approach to predict population genetic structure in bees. **Molecular Ecology**, v. 28, n. 8, p. 1919–1929, 2019.

LUU, K.; BAZIN, E.; BLUM, M. G. B. pcadapt: an R package to perform genome scans for selection based on principal component analysis. **Molecular Ecology Resources**, v. 17, n. 1, p. 67–77, 2017.

MACARTHUR, R. H.; WILSON, E. O. **IntroductionThe theory of island biogeography**, 1967.

MAESTRE, F. T. *et al.* Plant species richness and ecosystem multifunctionality in global drylands. **Science**, v. 335, n. 6065, p. 214–218, 2012.

MANEL, S. *et al.* **Landscape genetics: Combining landscape ecology and population geneticsTrends in Ecology and Evolution**Elsevier Ltd, , 1 abr. 2003.

MANEL, S.; HOLDEREGGER, R. **Ten years of landscape geneticsTrends in Ecology and Evolution**, out. 2013.

MARCHI, P.; ALVES-DOS-SANTOS, I. The bees of the genus Xylocopa Latreille (Xylocopini , Apidae) of Sao Paulo. n. June 2014, 2013.

MARTIN, A. E.; BENNETT, J. R.; FAHRIG, L. Habitat fragmentation. *Em*: **The Routledge Handbook of Landscape Ecology**. [s.l.] Routledge, 2021. p. 118–139.

MCKINNEY, M. L.; LOCKWOOD, J. L. Biotic homogenization: A few winners replacing many losers in the next mass extinction. **Trends in Ecology and Evolution**, v. 14, n. 11, p. 450–453, 1999.

MÉDAIL, F.; QUÉZEL, P. Biodiversity Hotspots in the Mediterranean Basin: Setting Global Conservation Priorities. **Conservation Biology**, v. 13, n. 6, p. 1510–1513, 1 dez. 1999.

MEINERS, J. M.; GRISWOLD, T. L.; CARRIL, O. M. Decades of native bee biodiversity surveys at Pinnacles National Park highlight the importance of monitoring natural areas over time. **PLoS ONE**, v. 14, n. 1, 1 jan. 2019.

METZGER, J. P. O que é ecologia de paisagens. **Biota Neotropica**, v. 17, n. 12, 2001.

MICHENER, C. D. Biogeography of the Bees. **Annals of the Missouri Botanical Garden**, v. 66, n. 3, p. 277, 1979.

MICHENER, C. D. **The bees of the world**. 2nd ed ed. Baltimore, Maryland: The Johns Hopkins University Press, 2007.

MINCKLEY, R. L. Maintenance of richness despite reduced abundance of desert bees (Hymenoptera: Apiformes) to persistent grazing. **Insect Conservation and Diversity**, v. 7, n. 3, p. 263–273, 2014.

MINCKLEY, R. L.; ASCHER, J. S. **Preliminary survey of bee (Hymenoptera: Anthophila) richness in the northwestern Chihuahuan Desert**Merging science and management in a rapidly changing world: Biodiversity and management of the Madrean Archipelago III: 2012 May 1-5; Tucson, AZ. Proceedingings. RMRS-P-67. **Anais**...2013

MINCKLEY, R. L.; CANE, J. H.; KERVIN, L. Origins and ecological consequences of pollen specialization among desert bees. **Proceedings of the Royal Society B: Biological Sciences**, v. 267, n. 1440, p. 265–271, 2000.

MINCKLEY, R. L.; RADKE, W. R. Extreme species density of bees (Apiformes, Hymenoptera) in the warm deserts of North America. **Journal of Hymenoptera Research**, v. 82, 2021.

NEI, M. Genetic Distance between Populations. **The American Naturalist**, v. 106, n. 949, p. 283–292, 25 ago. 1972.

NUNES, L. A. P. L.; ARAÚJO-FILHO, J. A. DE; MENEZES, R. I. DE Q. IMPACTO DA QUEIMADA E DO POUSIO SOBRE A QUALIDADE DE UM SOLO SOB CAATINGA NO SEMI-ÁRIDO NORDESTINO. **Revista Caatinga**, v. 19, n. 2, 2006.

OKSANEN, A. J. *et al.* Package 'vegan'_Ordination methods, diversity analysis and other functiuons for community and vegetation ecologists. **R package "veagan" version 2,3-0**, n. September, 2015.

ORR, M. C. *et al.* A new bee species sandstone nests. **Current Biology**, v. 26, n. 17, p. R792–R793, 2018.

ORR, M. C. *et al.* Global Patterns and Drivers of Bee Distribution. **Current Biology**, p. 1–8, 2020.

ORR, M. C.; PARKER, F. D.; WOODARD, S. H. A tale of two cliff dwellings: Parallels between humans and native bees. **American Entomologist**, v. 64, n. 4, p. 218–223, 2018.

PETERSON, B. K. *et al.* Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. **PLoS ONE**, v. 7, n. 5, 2012.

PORTMAN, Z. M. *et al.* Local extinction of a rare plant pollinator in Southern Utah (USA) associated with invasion by Africanized honey bees. **Biological Invasions**, 2017.

PORTMAN, Z. M.; TEPEDINO, V. J.; AMBER, D. Persistence of an imperiled specialist bee and its rare host plant in a protected area. 2018.

RAJ, A.; STEPHENS, M.; PRITCHARD, J. K. FastSTRUCTURE: Variational inference of population structure in large SNP data sets. **Genetics**, v. 197, n. 2, p. 573–589, 2014.

ROUSSET, F.; LOPEZ, L.; BELKHIR, K. Package “genepop”. **Cran**, 2017.

SAMPAIO, E. V. DE S. B. *et al.* Sustainable agricultural uses in the Caatinga. *Em*: **Caatinga: The Largest Tropical Dry Forest Region in South America**. [s.l.] Springer International Publishing, 2018. p. 413–428.

SANTOS, R. M. *et al.* The larval provisions of the crop pollinator Centris analis: Pollen spectrum and trophic niche breadth in an agroecosystem. **Apidologie**, v. 44, n. 6, p. 630–641, nov. 2013.

SEEBACHER, F.; FRANKLIN, C. E. Determining environmental causes of biological effects: the need for a mechanistic physiological dimension in conservation biology. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 367, n. 1596, p. 1607–1614, 2012.

SEGELBACHER, G. *et al.* Applications of landscape genetics in conservation biology: Concepts and challenges. **Conservation Genetics**, v. 11, n. 2, p. 375–385, 2010.

SEXTON, J. P.; STRAUSS, S. Y.; RICE, K. J. Gene flow increases fitness at the warm edge of a species’ range. **Proceedings of the National Academy of Sciences**, v. 108, n. 28, p. 11704–11709, 12 jul. 2011.

SHAFFER, M. L. Minimum Population Sizes for Species Conservation. **BioScience**, v. 31, n. 2, p. 131–134, 1 fev. 1981.

SHAH, V. B.; MCRAE, B. H. Circuitscape : A Tool for Landscape Ecology. **Proceedings of the 7th Python in Science Conference**, n. SciPy, 2008.

SILVA, C. I. DA; FREITAS, B. M. **Rearing carpenter bees (Xylocopa spp.) for crop pollination: A case study with passion fruit (Passiflora edulis)**. [s.l: s.n.].

SILVA, J. M. C. DA; LEAL, I. R.; TABARELLI, M. **Caatinga: The largest tropical dry forest region in South America**. [s.l: s.n.].

SILVA PINTO VIEIRA, R. M. DA *et al.* Identifying areas susceptible to desertification in the Brazilian Northeast. **Solid Earth Discussions**, v. 6, n. 2, p. 3227–3260, 2014.

SLATKIN, M. Gene flow in natural populations. **Annual review of ecology and systematics. Vol. 16**, p. 393–430, 1985a.

___. Rare alleles as indicators of gene flow. **Evolution**, v. 39, n. 1, 1985b.

SLATKIN, M.; TAKAHATA, N. The average frequency of private alleles in a partially isolated population. **Theoretical Population Biology**, v. 28, n. 3, p. 314–331, 1985.

STEFFAN-DEWENTER, I. Landscape context affects trap-nesting bees, wasps, and their natural enemies. **Ecological Entomology**, v. 27, n. 5, p. 631–637, out. 2002.

STORFER, A.; PATTON, A.; FRAIK, A. K. Navigating the Interface Between Landscape Genetics and Landscape Genomics. v. 9, n. March, 2018.

TALLES MARQUES, CHAVES-ALVES; SOLANGE CRISTINA, A. Recursos ecológicos utilizados por Xylocopa spp. em área urbana, Uberlândia-MG, Brasil. p. 2–3, 2011.

TELLES, M. P. DE C. *et al.* Disentangling landscape effects on population genetic structure of a Neotropical savanna tree. **Natureza a Conservacao**, v. 12, n. 1, 2014.

TOMASELLA, J. *et al.* Desertification trends in the Northeast of Brazil over the period 2000–2016. **International Journal of Applied Earth Observation and Geoinformation**, v. 73, n. August 2017, p. 197–206, 2018.

TRAILL, L. W.; BRADSHAW, C. J. A.; BROOK, B. W. Minimum viable population size: A meta-analysis of 30 years of published estimates. **Biological Conservation**, v. 139, n. 1–2, p. 159–166, 2007.

USAID. The US market for fresh passion fruit. **Acceso Market Brief 16**, n. December, p. 1–8, 2014.

WAITS, L. P.; STORFER, A. Basics of Population Genetics: Quantifying Neutral and Adaptive Genetic Variation for Landscape Genetic Studies. *Em*: **Landscape Genetics: Concepts, Methods, Applications**, 2015 .

WANG, X. *et al.* Genetic, phenotypic and ecological differentiation suggests incipient speciation in two Charadrius plovers along the Chinese coast. **BMC Evolutionary Biology 2019 19:1**, v. 19, n. 1, p. 1–18, 27 jun. 2019.

WHITFORD, W. G. Ecology of Desert Systems. San Diego, California, 343 pp. **Academic Press**, v. 84, n. 3, p. 1122–1124, 29 ago. 2002.

WILLMER, P.; STONE, G. Temperature and water relations in desert bees. **Journal of Thermal Biology**, v. 22, n. 6, p. 453–465, dez. 1997.

WRIGHT, S. ISOLATION BY DISTANCE. **Genetics**, v. 28, n. 2, p. 114 LP – 138, 29 mar. 1943.

ANEXO A – INFORMAÇÃO SUPLEMENTAR SOBRE AS ÁREAS SUSCETÍVEIS A DESERTIFICAÇÃO

Details describing the characteristics of the ASDs

ASD's are mainly degraded by the disorderly occupation of the soil, deforestation, and slash-and-burn practices. Agricultural establishments occupy the soil mainly with temporary crops (e.g. beans, maniot and corn) and animal husbandry (PAE-Ceará 2010). According to Funceme (2014), the proportion of different land uses for ASD I are: 40% of the territory (considering 12.202,41 km^2) is occupied by degraded Caatinga, with the primary vegetal cover transformed and inclusions of subsistence crops and extensive pastures. Thirty-nine percent are covered by Caatinga shrub-tree and shrub-herbaceous vegetation, in which predominates family farming and livestock farming combined with natural or cultivated pastures. Nine percent are covered by Caatiga partially conserved with natural characteristics remaining from the primary vegetation cover (it may present a shrub and herbaceous strata already devoid of the original vegetation as a result of logging). Seven percent are covered by intensely degraded Caatinga subjected to desertification processes and with irreversibly compromised soils and biodiversity, with dispersed presence of shrub-trees, shrub-herbaceous, cactus, rocky exposures, eroded soils with intense use by extensive grazing. The remaining 5% of the territory are occupied by moderately or intensely degraded riparian wood with extractivism and/or agriculture activities, seasonal flood areas, alluvial plains, and surface rock exposure slabs. Unfortunately, the same land use data is not available yet for ASD II at Funceme repository (http://www.funceme.br/) but a general overview of land use for this area can be found on page 113 in PAE-Ceará (2010).

ASD I includes the cities of Iracuba, Santa Quiteria, and Sobral, which are part of the "Iraucuba and the North-Centre Desertification Nucleus", located in the north western part of the state (Figure III.1). It spans an area of 12.202,41 km^2 of which mostly is affected by extreme and moderate levels of desertification (PAE-PAE-Ceará 2010, Funceme 2014). Its high propensity to desertification is particularly associated with the predominance of the solodic and natric haplic planosols, a type of mineral soil that is poorly drained (Perez-Marin et al. 2012). The rainwater accumulated in the soil is easily lost by evapotranspiration, favouring the concentration of crystalline salts (calcium, magnesium and sodium) both in the surface and along the soil profile (Perez-Marin et al. 2012). The excess of these salts harm the plant growth and provoke soil compaction problems hindering water movement and root development

(Perez-Marin et al. 2012). Thus, there is a restriction on the establishment of tall species and predominance of herbaceous species. Historically, over more than three centuries, the soils in this region have been compromised by the extraction of wood for various purposes, and agriculture has been one of the biggest anthropic contributors to the degradation of the environment (Facundo and Frota, 2020). To these days, remaining vegetation is generally used for pasture, most of with overloading, leading to the consumption of the whole herbaceous stratum, which renders the soil uncovered and lacking seed bank (Sampaio et al. 2013). The soils, already very susceptible to erosion, are left to the action of wind and rainwater. The main agriculture practices occur on the steep slopes, contributing to the loss of the few existing centimeters of soil (Barreto 2015). Temporary agriculture (beans, corn and maniot plantations) and animal farming correspond to 53% and 42%, respectively, of the agricultural establishments distributed in the ASD I area. Extraction of raw material for coal production has also been an important activity in ASD I (PAE-Ceará 2010).

ASD II includes the cities of Jaguaribe, Jaguaretama and Morada Nova, which are part of the "Medio Jaguaribe Desertification Nucleus", located in the south eastern part of the state (Figure III.1). It spans an area of 8.421,72 km^2 (Ceará, 2010) of which mostly is affected by extreme and moderate levels of desertification. The cattle and cotton activities, present in the region since the colonial period, are the main responsible for the degradation of this nucleus (Guerra et al. 2009). Livestock occupied the banks of the Jaguaribe River, which began to be populated, contributing to the occupation of the area and foundation of the Captaincy of Ceará during the eighteenth century (Guerra, 2009). This period was marked by an acceleration of cattle activity, which declined from the second half of the century, due to the occurrence of great droughts but also due to overgrazing and degradation of native pastures (PAE-Ceará 2010). Cotton was one of the main products traded internationally in the nineteenth century, but perished with a severe pest disrupting the economy. This crop was also developed in the Jaguaribe hydrographic basin (Guerra 2009). Both activities contributed to the desertification process of the region due to the opening of clearings first for settlement of cattle and, later, for the implantation of cotton farming (Guerra et al. 2009). As these activities did not follow a sustainable approach they resulted in desertified landscapes (PAE-Ceará 2010). Currently, temporary agriculture (mainly bean, maniot and corn plantations) and animal farming are the main activities, corresponding to 42% and 51%, respectively, of the agricultural establishments distributed in the ASD II area. Extraction of raw material for coal production is also one of the most important activities in ASD II (PAE-Ceará 2010).

Table S2. Land-use classes over each sample site of ASDI. Lo: local scale; Re: regional scale, and Ab: broad scale.

	JGI			JGII			JM			MN		
Classes	**Lo**	**Re**	**Ab**	**Lo**	**Re**	**Ab**	**Lo**	**Re**	**Ab**	**Lo**	**Re**	**Ab**
Forest		0	17	6	31	**241**	0	5	19	1	5	58
Grassland				9	364	831		71	131			**902**
Agri and Past	749	8020	**31295**	752	6985	28151	973	6761	27772	1036	7750	26920
past	466	2570	8721	404	2409	8436	232	3433	**10703**	74	1416	6124
rocky		9	**25**	1	6	24	1	2	10			5
savanna	34	658	4982	79	1097	6684	42	819	6211	142	2012	**10677**
urban					**299**	452		160	181			8
water	7	44	170	6	31	208	8	48	179	4	117	**433**

Table S3. Land-use classes over each sample site of ASDII. Lo: local scale; Re: regional scale, and Ab: broad scale.

	IRA			SO			SQ1			SQ2		
Classes	**Lo**	**Re**	**Ab**	**Lo**	**Re**	**Ab**	**Lo**	**Re**	**Ab**	**Lo**	**Re**	**Ab**
Forest	2	110	1249	4	260	1613	3	202	**866**	7	135	715
Grassland	6	25	55									

Agri and Past	194	1137	**3572**	11	169	795	19	291	1950	30	426	2220
past	792	4772	**9579**	27	311	1860	25	418	2268	56	1408	3918
rocky	2	45	**82**						2			
savanna	252	5074	30105	1201	10409	**40406**	1209	10083	38158	1160	9259	37928
urban	0	5	217			16		7	**338**			56
water	8	137	348	14	154	**519**		301	1626	2	74	371

ANEXO B - SCRIPTS E CÓDIGOS

RUN PROCESS_RADTAGS TO DEMULTIPLEX

*process_radtags -1 SBXYL01_USPD16095370_HWLVHBBXX_L3_1.fq.gz -2 SBXYL01_USPD16095370_HWLVHBBXX_L3_2.fq.gz *

-o clean -e ecoRI -i gzfastq -y fastq -c -q -r --barcode_dist_1 2 -b barcodes.txt

*process_radtags -1 SBXYL01_USPD16095370_HWLVHBBXX_L3_1.fq.gz -2 SBXYL01_USPD16095370_HWLVHBBXX_L3_2.fq.gz *

-o clean2 -e ecoRI -i gzfastq -y fastq -c -q -r --barcode_dist_1 1 -b barcodes2.txt

RUNDENOVOMAP.SH

*denovo_map.pl --samples clean/ --popmap popmap.txt -o stacks -T 48 --paired -X "ustacks: --bound_high 0.05" *

RUN POPULATIONS ON STACKS

populations -P stacks2/ --popmap popmap.txt --min-maf 0.1 -r 075 --vcf

MAKE .PED AND .MAP FILES TO USE IN ADMIXTURE

$ module load vcftools
$ vcftools --gzvcf populations.snps.vcf.gz --plink --out output_plink

MAKE .BED FROM .PED FILES TO USE IN ADMIXTURE

$ module load plink
$ plink --file output_plink --make-bed --out bedfromped

RUNNING FASTSTRUCTURE IN UNIX

Followed: fastStructure/README.md at master · rajanil/fastStructure · GitHub

$ python structure.py -K 2 --input=bedfromped --output=fsct_output $ python structure.py -K 3 --input=bedfromped --output=fsct_output $ python structure.py -K 4 --input=bedfromped --output=fsct_output $ python structure.py -K 5 --input=bedfromped --output=fsct_output $ python structure.py -K 6 --input=bedfromped --output=fsct_output $ python structure.py -K 7 --input=bedfromped --output=fsct_output

$ python structure.py -K 8 --input=bedfromped --output=fsct_output

$ python ~/proj/fastStructure/chooseK.py --input=fsct_output Model complexity that maximizes marginal likelihood = 2 Model components used to explain structure in data = 3

$ python distruct.py -K 3 --input= fsct_output --output= fsct_output_distruct.svg

RUN FST ANALYSIS USING VCFTOOLS

EXAMPLE:

$ vcftools --gzvcf populations.snps.vcf.gz --weir-fst-pop JM_Fst.txt --weir-fst-pop MN_Fst.txt --out JM_MN

R SCRIPTS
Fst analysis

library(reshape2)
read.table("FST_83KSNPS.txt", sep="\t", header = T) -> Fst_abc
Fst_abc
melted_Fst_abc <- melt(Fst_abc)
head(melted_Fst_abc)
fstplot<- ggplot(data= melted_Fst_abc, aes(x=X, y=variable, fill=value))+

geom_raster(data = melted_Fst_abc, aes(x=X, y=variable, fill=value)) +
geom_text(aes(x=X, y=variable, label = value,), color = "white", size

= 6)+
labs(x="Populations", y="Populations", title = "Fst matrix") +

theme_bw() + theme(axis.text.x=element_text(size=15, angle=0, vjust=0.3), axis.text.y=element_text(size=15), plot.title = element_text(size = 11)+ theme(legend.position = "center"))

fstplot

fstplot +
geom_text(aes(x=X, y=variable, label = value,), color = "white", size = 6) + theme(

axis.title.x = element_blank(), axis.title.y = element_blank(), panel.grid.major = element_blank(), panel.border = element_blank(), panel.background = element_blank(), axis.ticks = element_blank(), legend.justification = c(1.5, 0), legend.position = c(.3, .4),

legend.direction = "vertical")+
guides(fill = guide_colorbar(barwidth = 3, barheight = 14,

title.position = "top", title.hjust = 0.5))

PCA analysis

Used Plink2.0
I was getting the following error:

Error: Invalid chromosome code '31' on line 35 of --vcf file

It seems that plink thinks the numbers are real chr, when in fact they are scaffolds. Then used the following command to insert the word scaffold in the beginning of every number:

s

sed -E 's/^([0-9])/scaffold_\1/g' populations.snps.vcf > new.vcf

"s/<regular expression>/<thing to replace it with>/g"
"^" refer to the beginning of the line, and the parentheses and "\1" are used to copy the digit instead of replacing it (we want to replace "25" with "scaffold_25", not "scaffold_5")

then, ran pca analysis

plink2 --vcf new.vcf --pca –aec which generated *eingeval* and *eigenvec* files. Plotted eigenvec on R creating a scatterplot

install.packages("ggplot") install.packages("tidyverse") library(ggplot2) library(tidyverse)

pca1<- read.table("PCA_eigenvec.txt", header = F) plot(data= pca1, V3~V4)

ggplot(pca1, aes(V3, V4, color = V2, size=0.5)) + geom_point() +
theme(axis.text.x=element_text(size=15, angle=0, vjust=0.3),

axis.text.y=element_text(size=15), plot.title = element_text(size = 11)+ theme(legend.position = "center"))

MRMD for IBD analysis

library(ecodist

###creating a geographic matrix distance### geo = read.table("lat_long.txt", header = T)
geo1 <- as.matrix(geo)
geoDist <- fields::rdist.earth(geo1, miles = F) geoDist[upper.tri(geoDist, diag = TRUE)] <- NA geoDist

###creating a genetic matrix distance###
genDist = read.delim("Fst_to.txt", row.names = 1, header = T) genDist1 <- as.matrix(genDist) genDist1[is.na(genDist1)]<-NA
genDist1

 # Does genetic distance is related to geographic distance? # MRM(dist(genDist1) ~ dist(geoDist), nperm = 10000)

###Plotting using ggplot2 ### library(ggplot2)

geogen <- read.table("geo_gen_ggplot.txt", header = T)
ggplot(geogen, aes(geo, gen)) + geom_point(size=4) + geom_smooth(method = lm) +

labs(title = "Isolation by Distance", x= "Geographic distance", y = "Genetic distance (Fst)")

p + theme_bw()

$coef

dist(genDist1) pval Int 3.863776e-03 0.3405

dist(geoDist) -8.500240e-07 0.7021

$r.squared
R2 pval

0.003768091 0.702100000

$F .test
F F .pval

0.07186453 0.70210000

obs: The gap existing in the middle of the graph is due to the absence of genetic data between ~100 and ~200 km distances, as one can see above:

RDA

Call: rda(formula = gen.imp ~ alt + TAR + PS + PWQ + PCQ, data = envfinal, scale = F)

Inertia Proportion Rank Total 3.501e+03 1.000e+00

Constrained 3.013e+02 8.606e-02 5 Unconstrained 3.200e+03 9.139e-01 74 Inertia is variance

Eigenvalues for constrained axes: RDA1 RDA2 RDA3 RDA4 RDA5

66.77 62.02 57.95 57.89 56.64

Eigenvalues for unconstrained axes:
PC1 PC2 PC3 PC4 PC5 PC6 PC7 PC8

66.79 65.65 59.42 55.70 53.02 52.83 52.00 51.57 (Showing 8 of 74 unconstrained eigenvalues)

> RsquareAdj(xylo.rda) $r.squared
[1] 0.08605619

$adj.r.squared [1] 0.02430323

Importance of components:
RDA1 RDA2 RDA3 RDA4 RDA5

Eigenvalue 333.7117 311.7552 281.0259 268.5746 256.2249

Proportion Explained 0.2299 0.2148 0.1936

0.1851 0.8235

alt TAR PS
12.851700 4.758441 9.885232 2.145550 17.177633

ALT PCQ PS PWQ TAR 33 30 133 35 6

0.1765 1.0000

Cumulative Proportion 0.2299 > screeplot(xylo.rda)
> vif.cca(xylo.rda)

0.4448 0.6384

PWQ PCQ

Analysis of Deviance Table
Model: poisson, link: log
Response: NPA
Terms added sequentially (first to last)

Df Deviance Resid. Df Resid. Dev Pr(>Chi)

NULL
P2000 1 0.319 P3000 1 49.724 P3500 1 0.126 P6000 1 7.061 P8000 1 2.780

P10000 1 18.683 P12000 1 0.049

78.743
6 78.424 0.571923

5 28.700 1.770e-12 *** 4 28.575 0.722955
3 21.513 0.007877 **
2 18.733 0.095425 .

1 0.049 1.543e-05 *** 0 0.000 0.824113