

UNIVERSIDADE FEDERAL DE SÃO CARLOS  
CENTRO DE CIÊNCIAS BIOLÓGICAS E DA SAÚDE  
DEPARTAMENTO DE GENÉTICA E EVOLUÇÃO  
Programa de Pós-graduação em Genética Evolutiva e Biologia Molecular

FAGNER MIGUEL DA SILVA

COLONIZAÇÃO DO BRASIL PELA GARÇA-VAQUEIRA (*Bubulcus ibis*):  
distribuição intercontinental da variação genética determinada por dados genômicos

SÃO CARLOS - SP  
2019

FAGNER MIGUEL DA SILVA

COLONIZAÇÃO DO BRASIL PELA GARÇA-VAQUEIRA (*Bubulcus ibis*):  
distribuição intercontinental da variação genética determinada por dados genômicos

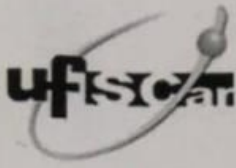
Tese apresentada ao Programa de Pós-Graduação em Genética Evolutiva e Biologia Molecular da Universidade Federal de São Carlos para obtenção do título de Doutor em Ciências (Área de concentração: Genética e Evolução).

Orientação: Profa. Dra. Sílvia Nassif Del Lama

Co-orientação: Profa. Dra. Sônia Cristina da Silva Andrade

Financiamento: Fapesp (2016/01673-7) e CNPq (870195/1997-2)

SÃO CARLOS - SP  
2019



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Biológicas e da Saúde  
Programa de Pós-Graduação em Genética Evolutiva e Biologia Molecular

Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Tese de Doutorado do candidato Fagner Miguel da Silva, realizada em 31/05/2019:

Prof. Dra. Silvia Nassif Del Lama  
UFSCar

Prof. Dra. Sonia Cristina da Silva Andrade  
USP

Prof. Dr. Caio Cesar de Melo Freire  
UFSCar

Prof. Dr. Evandro Marsola de Moraes  
UFSCar

Prof. Dra. Elen Arroyo Peres  
UNICAMP

Prof. Dr. Celso Teixeira Mendes Junior  
USP

## AGRADECIMENTOS

É impossível não olhar para trás, após tantos anos dedicado a um trabalho, e ver que tudo mudou. Felizmente, a maior parte das mudanças foram boas e impulsionaram o meu crescimento pessoal e profissional durante o desenvolvimento desse trabalho. Mas não carrego sozinho o mérito de tê-lo concluído. Por isso, agradeço:

Aos meus pais, Joaquim e Elza, por serem o meu porto seguro e me ensinarem “as coisas da vida”. Ao meu namorado, Vinícius, que tem me acompanhado já a alguns anos, pelo amor e pela lembrança frequente de que a minha vida não se resume a trabalho.

À Professora Dra. Silvia Nassif Del Lama e à Dra. Sônia Cristina da Silva Andrade pela orientação, conselhos, direcionamento acadêmico e amizade. Agradeço, especialmente à Professora Dra. Silvia pela confiança no meu trabalho e pelas inúmeras horas dedicadas à minha formação.

Aos amigos, colegas e ex-colegas dos Laboratórios de Genética de Aves e de Genética Evolutiva de Himenópteros na UFSCar, e do Laboratório de Diversidade Genômica no Instituto de Biociências da USP: Talita, Bruno, Carlos, Cristiana, Camila, Mariana, Paula, Diana, Aline, Leandro, Kátia, Isabel, Regina, Cecili, Gabriel, Thainá e Tammy. Obrigado por compartilharem não só as dificuldades e os conhecimentos, mas os momentos alegres do dia.

Ao Prof. Marco Antonio Del Lama pela amizade e ajuda intelectual.

A todos que contribuíram direta e indiretamente para a obtenção das amostras de material biológico utilizadas nesse trabalho: Silvia Del Lama, Emmanuel Moralez-Silva, Cristiano Dosualdo Rocha, Rá Cavalcanti, Hamilton Amarildo Barbosa Monteiro, Aissa Regalla, Moussa Dióp, Daryl Bosu, David, Dominic Henry, Margarida Fernandes e Vitor Encarnação.

À Universidade Federal de São Carlo, à FAPESP e ao CNPq pelo apoio financeiro para execução desse projeto.

A todos os colegas e amigos do Departamento de Genética e Evolução da UFSCar pelos momentos de descontração na rotina da Pós-graduação.

Aos amigos feitos nestes quatro últimos anos ou que me acompanham já a muito tempo, pelo entusiasmo e apoio, apesar da distância e do número reduzido de encontros.

Não deixo só o meu agradecimento, mas o meu desejo de que continuemos juntos lutando pela ciência no Brasil.

*The cattle egrets known to travel far  
And you always see them near where livestock are  
Insects, beetles, toadlets and such they eat  
That fly or hop clear of big trampling feet.*

*The cattle egrets are whiter than the snow  
Associate with cattle everywhere they go  
They keep insects and harmful slugs at bay  
And in so doing help farmer no small way.*

*As members of the heron family  
They build their nests on higher bush or tree  
On lake or river, sometimes estuary  
On nest of sticks pale blue eggs often three.*

*Oft seen them in Narre Warren in Winter-time  
These cattle egrets favourite birds of mine  
In paddocks with the herds they spend their days  
Eating insects flushed by cattle as they graze.*

*The cattle egrets that you see today  
Tomorrow may be many miles away  
Nomadic birds their journeys take them far  
But you'll always see them near where cattle are.*

Francis Duggan<sup>1</sup>, *The cattle egrets* (2009)

*“... Eu tenho que essa visão oblíqua vem de eu ter sido criança em algum lugar perdido onde havia transfusão da natureza e comunhão com ela. Era o menino e os bichinhos.”*

Manoel de Barros<sup>2</sup>, *Memórias inventadas* (2010)

---

<sup>1</sup> Francis Duggan (1946 – presente) é um poeta contemporâneo, autor de mais de nove mil peças publicadas online em sites de poesia.

<sup>2</sup> Manoel de Barros (1916-2014) foi um advogado e poeta brasileiro mato-grossense, que teve a natureza como tema frequente de seus escritos.

## RESUMO

A garça-vaqueira, *Bubulcus ibis* (Linnaeus, 1758), é uma espécie de ave com alta capacidade de colonização e potencial invasor. Durante o último século, a espécie expandiu consideravelmente a sua distribuição geográfica e o tamanho de suas populações. Atualmente, a espécie ocorre em grande parte do continente africano, sul da Europa e nas Américas. O objetivo principal deste trabalho foi, a partir da distribuição da variação genética, inferir o padrão atual de estruturação genética e uma possível rota de invasão que explicasse o processo de colonização do Brasil por *Bubulcus ibis*. Para alcançá-lo, um rascunho do genoma da espécie foi montado e novos marcadores genéticos, 2.751 SNPs e 22 microssatélites, foram prospectados. Para avaliar a distribuição da variação genética e a conectividade entre as populações estimou-se: a) índices de diversidade genética e medidas de diferenciação genética populacional, b) taxas de migração por geração, padrões de divisão e mistura populacional histórica e superfícies efetivas de migração. Além disso, foi avaliado se um padrão geográfico e/ou adaptativo explicava a distribuição da variação genética, buscando SNPs potencialmente sob seleção e testando a adequação dos dados a um modelo de isolamento por distância. Utilizando a informação de 2.751 SNPs e oito dos *loci* microssatélites prospectados, determinou-se que, embora com níveis similares de diversidade genética, há estruturação genética entre as populações de *Bubulcus ibis* oriundas das áreas nativa e colonizada (um agrupamento genético mais representativo para cada área). As proporções de ancestralidade genética estimadas mostraram uma conexão entre populações africanas (Guiné-Bissau e Senegal) e o nordeste brasileiro (Fernando de Noronha e Pernambuco). Esse resultado é suportado pela rota de migração mais provável inferida pelos padrões de divisão e mistura populacional histórica (de Guiné Bissau para o Pernambuco). Ainda, a modelagem de migração efetiva detectou uma resistência ao fluxo gênico entre as populações brasileiras e o Sul da África. A ausência de SNPs potencialmente sob seleção e a correlação positiva entre as distâncias genética e geográfica indicaram que o fator geográfico é mais determinante do que o fator adaptativo para a estruturação genética encontrada. A rota mais provável de invasão de *Bubulcus ibis* tem como ponto de chegada o Nordeste brasileiro e não o Norte do país (onde o primeiro registro foi feito em 1964), com indícios de que as partidas na área nativa ainda ocorram ou ocorreram no centro-oeste africano.

Palavras-chave: SNPs; microssatélites; Aves; Invasão; genômica.

## ABSTRACT

The cattle egret, *Bubulcus ibis* (Linnaeus, 1758), is a species of bird with high colonization capacity and invading potential. During the last century, this species has considerably expanded its geographical distribution and the size of its populations. Currently, *Bubulcus ibis* occurs in much of the African continent, southern Europe and the Americas. The main objective of this work was, based on the distribution of genetic variation, to infer the current pattern of genetic structure and a possible invasion route that explained the process of colonization of Brazil by *Bubulcus ibis*. To reach it, a draft genome of the species was assembled and new genetic markers, 2,751 SNPs and 22 microsatellites, were prospected. To assess the distribution of genetic variation and connectivity among populations, we estimated: a) indices of genetic diversity and measures of population genetic differentiation, b) migration rates per generation, patterns of population division and historical mix and effective migration surfaces. In addition, it was evaluated whether a geographic and/or adaptive pattern explained the genetic variation distribution, searching for potential SNPs under selection and testing the adequacy of the data to an isolation by distance model. Using information from 2,751 SNPs and eight prospective microsatellite loci, it was determined that, although with similar levels of genetic diversity, there is genetic structuring between the populations of *Bubulcus ibis* from the native and colonized areas (a more representative genetic cluster for each area). The estimated proportions of genetic ancestry showed a connection between African populations (Guinea-Bissau and Senegal) and the Brazilian northeast (Fernando de Noronha and Pernambuco). This result is supported by the most likely migration route inferred by the patterns of division and historical population mix (from Guinea Bissau to Pernambuco). Also, effective migration modelling detected a resistance to gene flow between Brazilian populations and southern Africa. The absence of potentially selective SNPs and the positive correlation between genetic and geographic distances indicated that the geographic factor is more determinant than the adaptive factor for the genetic structuring found. The most likely route of invasion of *Bubulcus ibis* is the Northeast of Brazil and not the North of the country (where the first record was made in 1964), with indications that departures in the native area still occur or occurred in the center-west African.

Keywords: SNPs; microsatellites; Birds; Invasion; genomics.

## LISTA DE FIGURAS

### **CAPÍTULO I**

<b>Figura 1</b> – Subespécie ocidental da garça-vaqueira ( <i>Bubulcus ibis ibis</i> ).....	13
<b>Figura 2</b> - Distribuição geográfica da subespécie ocidental de garça-vaqueira ( <i>Bubulcus ibis ibis</i> ). .....	15
<b>Figura 3</b> - Mapa com a localização das populações de garça-vaqueira ( <i>Bubulcus ibis</i> ) amostradas.....	26
<b>Figura 4</b> - Frequências alélicas para os 22 loci de microssatélites prospectados para a garça-vaqueira ( <i>Bubulcus ibis</i> ). .....	40
<b>Figura 5</b> - Informatividade do novo conjunto de loci de microssatélites isolados para a garça-vaqueira ( <i>Bubulcus ibis</i> ) ao inferir relações pareadas de parentesco genético. ....	41
<b>Figura 6</b> - Evidência de diferenciação genética entre as populações de garça-vaqueira ( <i>Bubulcus ibis</i> ) a partir de dados de microssatélites.....	44
<b>Figura 7</b> - Frequências alélicas dos SNPs identificados nas populações amostradas de garça-vaqueira ( <i>Bubulcus ibis</i> ). .....	47
<b>Figura 8</b> - Relação entre as heterozigosidades esperada e observada estimadas para cada um dos 2.751 SNPs genotipados nas populações ( $n = 174$ ) amostradas de garça-vaqueira ( <i>Bubulcus ibis</i> ). .....	47



## LISTA DE FIGURAS

### CAPÍTULO II

- Figura 9** - Evidência de diferenciação genética entre as populações de garça-vaqueira (*Bubulcus ibis*) amostradas obtida a partir da informação de oito loci de microssatélites. .... 73
- Figura 10** - Taxas de migração ( $m$ ) estimadas entre as populações da garça-vaqueira (*Bubulcus ibis*) amostradas a partir da informação de oito loci de microssatélites. 74
- Figura 11** - Coordenadas dos indivíduos amostrados de garça-vaqueira (*Bubulcus ibis*) ao longo de dois eixos (1 e 2) de variação genética determinados pela análise de componentes principais usando os dados de 2.751 SNPs. .... 79
- Figura 12** - Coordenadas dos indivíduos amostrados de garça-vaqueira (*Bubulcus ibis*) ao longo de dois eixos (1 e 3) de variação genética determinados pela análise de componentes principais usando os dados de 2.751 SNPs. .... 80
- Figura 13** - Coordenadas dos indivíduos amostrados de garça-vaqueira (*Bubulcus ibis*) ao longo de dois eixos (2 e 3) de variação genética determinados pela análise de componentes principais usando os dados de 2.751 SNPs ..... 81
- Figura 14** - Estrutura genética inferida para as populações de garça-vaqueira (*Bubulcus ibis*) a partir de dados genômicos ..... 83
- Figura 15** - Estrutura genética inferida para as populações de garça-vaqueira (*Bubulcus ibis*) a partir de 2.751 SNPs. .... 84
- Figura 16** - Estrutura genética inferida para as populações de garça-vaqueira (*Bubulcus ibis*) a partir de 2.751 SNPs, incorporando a localização geográfica das populações. .... 85
- Figura 17** - Padrão de divisão e mistura histórica estimado entre as populações de garça-vaqueira (*Bubulcus ibis*) a partir da informação de 2.751 SNPs. .... 86
- Figura 18** - Superfícies de migração efetiva estimadas entre as populações de garça-vaqueira (*Bubulcus ibis*) a partir da informação de 2.751 SNPs e a localização geográfica das populações. .... 87
- Figura 19** - Padrão de isolamento por distância encontrado para as populações de garça-vaqueira (*Bubulcus ibis*) a partir da informação de 2.751 SNPs. .... 89
- Figura 20** - SNPs com maior contribuição para o padrão de estruturação genética encontrado entre as populações de garça-vaqueira (*Bubulcus ibis*) amostradas. .... 89
- Figura 21** - SNPs potencialmente correlacionados a variáveis ambientais das localidades onde as populações de garça-vaqueira (*Bubulcus ibis*) foram amostradas. .... 90

## LISTA DE TABELAS

### **CAPÍTULO I**

- Tabela 1** - Informações sobre as populações de garça-vaqueira (*Bubulcus ibis*) amostradas e utilizadas para a prospecção e caracterização de loci microssatélites e análises populacionais subsequentes..... 27
- Tabela 2** - Informações sobre as populações de garça-vaqueira (*Bubulcus ibis*) amostradas e utilizadas para a identificação de polimorfismos de nucleotídeo único (SNPs) por Genotyping-by-sequencing. .... 28
- Tabela 3** - Características dos loci de microssatélites da garça-vaqueira (*Bubulcus ibis*) selecionados para verificar a existência de polimorfismos..... 30
- Tabela 4** - Loci de microssatélites polimórficos prospectados ..... 39
- Tabela 5** - Índices de diversidade genética e informatividade determinados em 25 loci de microssatélites para populações de garça-vaqueira (*Bubulcus ibis*) ..... 43
- Tabela 6** - Taxas de migração ( $m$ ) estimadas entre as populações da garça-vaqueira (*Bubulcus ibis*) a partir de dados de microssatélites..... 44
- Tabela 7** - Alterações no tamanho populacional e na diversidade genética estimadas para a população de garça-vaqueira (*Bubulcus ibis*) no arquipélago de Fernando de Noronha entre os anos de 2008 e 2020. .... 45

## LISTA DE TABELAS

### ***CAPÍTULO II***

- Tabela 8 - Variáveis ambientais extraídas para as localidades das populações de garça-vaqueira (Bubulcus ibis) amostradas..... 71***
- Tabela 9 - Diferenciação genética determinada entre as populações de garça-vaqueira (Bubulcus ibis) amostradas a partir da informação de oito loci de microssatélites. 72***
- Tabela 10 - Diversidade genética determinada para as populações de garça-vaqueira (Bubulcus ibis) a partir da informação de 2.751 SNPs..... 75***
- Tabela 11 - Distribuição da variação genética em grupos de populações inferidos por uma análise espacial de variância molecular (SAMOVA) a partir da informação de 2.751 SNPs..... 76***
- Tabela 12 - Diferenciação genética determinada entre as populações de garça-vaqueira (Bubulcus ibis) amostradas a partir da informação de 2.751 SNPs..... 77***
- Tabela 13 - Atribuição dos indivíduos amostrados de garça-vaqueira (Bubulcus ibis) aos grupos genéticos inferidos pela análise discriminante de componentes principais ..... 82***
- Tabela 14 - Atribuição populacional dos indivíduos de garça-vaqueira (Bubulcus ibis) realizada a partir da informação de 2.751 SNPS..... 88***
- Tabela 15 - SNPs mapeados no genoma de referência da garça-vaqueira (Bubulcus ibis) e na base de dados NCBI..... 91***

## SUMÁRIO

<b>I.</b>	<b>Introdução geral</b> .....	13
1	A garça-vaqueira: <i>Bubulcus ibis</i> .....	13
2	<i>Bubulcus ibis</i> como modelo em estudos genéticos populacionais.....	15
3	A proposta geral do trabalho.....	17
<b>II.</b>	<b>Capítulo I – Novos marcadores genéticos para a garça-vaqueira (<i>Bubulcus ibis</i>): prospecção, caracterização e um estudo de caso com a população de Fernando de Noronha</b> .....	18
1	Introdução .....	19
1.1	A distribuição de marcadores microssatélites e SNPs nos genomas das aves.....	19
1.2	O advento das tecnologias de sequenciamento de próxima geração na identificação de marcadores moleculares e em estudos populacionais ...	20
1.3	Estudo de caso: A garça-vaqueira ( <i>Bubulcus ibis</i> ) no arquipélago de Fernando de Noronha .....	22
1.4	A proposta de trabalho: capítulo I .....	24
2	Objetivos.....	25
2.1	Objetivo geral.....	25
2.2	Objetivos específicos .....	25
3	Material e Métodos .....	26
3.1	Coleta de material biológico e extração do DNA genômico .....	26
3.2	Montagem e anotação gênica do genoma de <i>Bubulcus ibis</i> .....	28
3.3	Prospecção, amplificação e genotipagem de loci de microssatélites no genoma de referência de <i>Bubulcus ibis</i> .....	29
3.4	Caracterização dos loci de microssatélites prospectados em populações naturais de <i>Bubulcus ibis</i> .....	32
3.5	Um estudo de caso com a população de <i>Bubulcus ibis</i> no arquipélago de Fernando de Noronha .....	34
3.6	Identificação de SNPs no genoma de <i>Bubulcus ibis</i> .....	35
3.6.1	Construção de bibliotecas genômicas para Genotyping-by-sequencing (GBS).....	35
3.6.2	Filtragem, montagem, clusterização das sequências e identificação de SNPs .....	36
4	Resultados.....	38
4.1	O genoma de referência de <i>Bubulcus ibis</i> .....	38
4.2	Caracterização dos loci de microssatélites prospectados para <i>Bubulcus ibis</i> .....	38

4.2.1	<i>Construção de bibliotecas genômicas para Genotyping-by-sequencing (GBS) Aplicação dos loci de microssatélites prospectados no estudo genético em populações de Bubulcus ibis</i>	41
4.3	<i>O estudo de caso com a população de Bubulcus ibis no arquipélago de Fernando de Noronha</i>	44
4.4	<i>SNPs identificados no genoma de Bubulcus ibis por Genotyping-by-sequencing</i>	46
5	Discussão	48
5.1	<i>O genoma de referência de Bubulcus ibis</i>	48
5.2	<i>Caracterização dos loci de microssatélites prospectados para Bubulcus ibis</i>	48
5.3	<i>Aplicação dos loci de microssatélites prospectados no estudo genético em populações de Bubulcus ibis</i>	50
5.4	<i>O estudo de caso com a população de Bubulcus ibis no arquipélago de Fernando de Noronha</i>	51
5.5	<i>SNPs identificados no genoma de Bubulcus ibis por Genotyping-by-sequencing</i>	53
6	Conclusões	54
<b>II.</b>	<b>Capítulo II – Um século após a colonização do Brasil pela garça-vaqueira (<i>Bubulcus ibis</i>): distribuição da variação genética e conectividade entre as populações naturais da espécie</b>	56
1	Introdução	57
1.1	<i>O processo de colonização por espécies com dispersão de longa distância</i>	57
1.2	<i>Evidências históricas de migração e da expansão da distribuição geográfica de Bubulcus ibis</i>	60
1.3	<i>A proposta de trabalho: Capítulo II</i>	63
2	Hipóteses e objetivos	64
2.1	<i>Hipóteses</i>	64
2.2	<i>Objetivo geral</i>	64
2.3	<i>Objetivos específicos</i>	65
3	Métodos	65
3.1	<i>Dados de microssatélites no estudo da colonização do Brasil por Bubulcus ibis</i>	65
3.2	<i>Dados de SNPs no estudo da colonização do Brasil por Bubulcus ibis</i>	66
3.3	<i>Padrão geográfico e/ou adaptativo na distribuição da variação genética populacional de Bubulcus ibis</i>	70

4	Resultados.....	72
4.1	<i>A colonização do Brasil por Bubulcus ibis: dados de microssatélites</i> .....	72
4.2	<i>A colonização do Brasil por Bubulcus ibis: dados de SNPs</i> .....	74
4.3	<i>Padrão geográfico e/ou adaptativo na distribuição da variação genética populacional de Bubulcus ibis</i> .....	88
5	Discussão .....	92
5.1	<i>Níveis de diversidade genética das populações amostradas de Bubulcus ibis</i> .....	92
5.2	<i>A rota de invasão mais provável durante o processo de colonização do Brasil por Bubulcus ibis</i> .....	94
5.3	<i>Padrão de estruturação genética entre as populações amostradas de Bubulcus ibis</i> .....	96
5.4	<i>Padrão geográfico e/ou adaptativo na distribuição da variação genética populacional de Bubulcus ibis</i> .....	99
6	Conclusões.....	101
<b>IV.</b>	<b>Referências Bibliográficas</b> .....	102
<b>V.</b>	<b>Anexo I:</b> .....	118
<b>VI.</b>	<b>Anexo I:</b> .....	137

## I INTRODUÇÃO GERAL

### 1 A garça-vaqueira: *Bubulcus ibis*

A garça-vaqueira, *Bubulcus ibis* (Pelecaniformes: Ardeidae, Linnaeus, 1758), é uma espécie de ave com reprodução colonial e alta capacidade de dispersão e colonização (KUSHLAN; HAFNER, 2000). *Bubulcus ibis* é considerada uma espécie invasora pela *International Union for Conservation Nature* (IUCN) e foi listada como espécie potencialmente invasora em diversas localidades como os arquipélagos de Galápagos e de Fernando de Noronha (NUNES et al., 2010; PHILLIPS; WIEDENFELD; SNELL, 2012; GISD, 2018). Três subespécies são atualmente reconhecidas como pertencentes ao gênero monotípico *Bubulcus*: *Bubulcus ibis ibis* (Linnaeus, 1758), *Bubulcus ibis coromandus* (Boddaert, 1783) e *Bubulcus ibis seychellarum* (Salomonsen, 1934) (VAURIE, 1963; KUSHLAN; HAFNER, 2000). *Bubulcus ibis ibis* (Figura 1) e *Bubulcus ibis coromandus* são as duas subespécies de garça-vaqueira mais comuns: *Bubulcus ibis ibis* ocorre em grande parte do continente africano, sul da Europa (principalmente França, Portugal e Espanha) e nas Américas (Figura 2), enquanto *Bubulcus ibis coromandus* ocorre no sul e sudeste da Ásia, norte e sul da China, além do Japão e norte e nordeste da Austrália oriental (AHMED, 2011). A população de garça-vaqueira que ocorre no arquipélago de Seychelles tem sido identificada como uma terceira subespécie, *Bubulcus ibis seychellarum*, mas a validade dessa classificação é questionada por taxonomistas que acreditam não haver diferenças morfológicas em relação a *Bubulcus ibis ibis* (DICKINSON, 2003; RASMUSSEN; ANDERSON, 2005).

**Figura 1** – Subespécie ocidental da garça-vaqueira (*Bubulcus ibis ibis*).



Fonte: Arquivo do laboratório de Genética de Aves (Avegene, UFSCar).

O foco da investigação desse trabalho foi a subespécie *Bubulcus ibis ibis*,

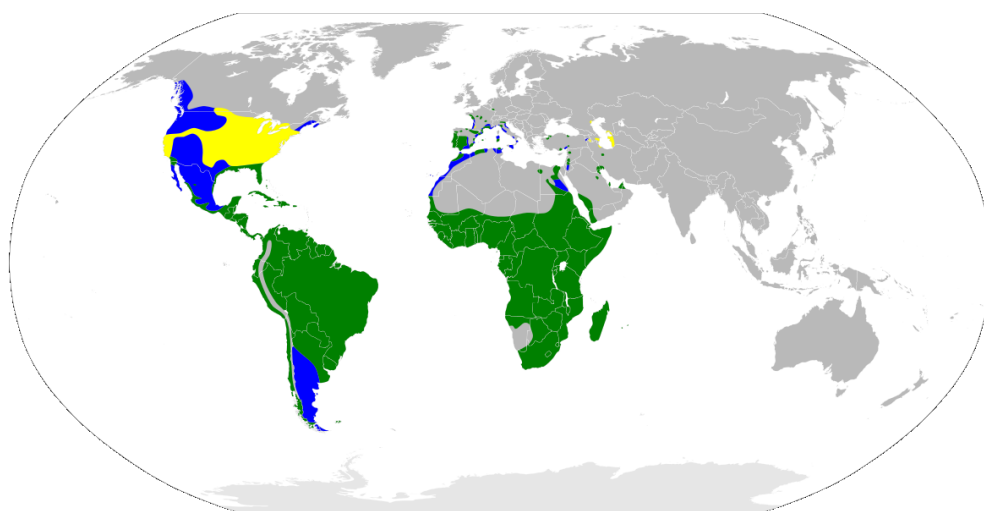
denominada a seguir apenas como *Bubulcus ibis*. As áreas de pouso, dormitório e nidificação dessa espécie são estabelecidas preferencialmente próximas a corpos d'água, embora *Bubulcus ibis* seja considerada a mais resiliente de todas as garças, sobrevivendo em ambientes relativamente secos e altamente antropizados (MARFIM, 2000). A plumagem é branca e o tamanho dos adultos varia entre 46 e 56 cm, com envergadura máxima de 96 cm e peso entre 270 e 512 g (KUSHLAN; HANCOCK, 2005). Durante o período reprodutivo, o bico amarelo e as patas amarelo-acinzentadas tornam-se alaranjados e manchas dessa cor na plumagem do peito, das costas e do alto da cabeça ficam evidentes (KUSHLAN; HANCOCK, 2005). O hábito alimentar geralmente é gregário e associado ao deslocamento de animais pastadores (SIEGFRIED, 2015). A dieta é constituída majoritariamente por artrópodes, mas oportunisticamente inclui anelídeos, miriápodes, peixes, anfíbios, répteis, roedores e ovos e filhotes de outras aves (KUSHLAN; HANCOCK, 2005; SHARAH; ALI; MOHAMMED, 2008). A longevidade máxima observada na natureza foi de 23 anos (CAREY; JUDGE, 2000) e a maturidade sexual é atingida entre o primeiro e o segundo ano de vida (KOLAR, 1966; SIEGFRIED, 1965; WESTERSKOV, 1974). Ambos os sexos incubam em média de dois a cinco ovos por um período de 21 a 26 dias (TELFAIR, 1983; KUSHLAN; HANCOCK, 2005). Os jovens começam a voar com três semanas de vida e tornam-se independentes entre seis e sete semanas de vida (KUSHLAN; HANCOCK, 2005). Os jovens podem dispersar para áreas localizadas até 5000 km de distância a partir das colônias natais (BROWDER, 1973; TELFAIR, 1994). O sucesso reprodutivo é alto, chegando a 94% de tentativas de reprodução bem-sucedidas por ano e com produtividade média de 1,22 jovens por adulto (KUSHLAN; HANCOCK, 2005; TELFAIR, 1994). A mortalidade anual foi reportada variando entre 37,4 e 55,8% para os jovens e entre 20,6 e 27,8% para os adultos (TELFAIR, 1994).

*Bubulcus ibis* expandiu sua distribuição geográfica e os tamanhos de suas populações durante o último século em uma velocidade de expansão estimada de 106,2 km/ano (VAN DEN BOSCH; HENGEVELD; METZ, 2006). Nas estimativas da Birdlife International (2009) a espécie possui cerca de 2.750.000 indivíduos na região Neártica e 3.250.000 na região Neotropical. Historicamente, sua distribuição estava concentrada pela África central tropical (KUSHLAN; HANCOCK, 2005). A expansão populacional na África incluiu a região central e sul do continente, com exceção dos desertos do Saara e da Namíbia (DEL HOYO et al., 1992; KUSHLAN; HANCOCK, 2005). Entre a segunda metade do século XIX e o início do século XX, a espécie colonizou as Américas e



expandiu sua distribuição geográfica também pelo sul da Europa (ARENDRT, 1988; BLAKER, 1971). A capacidade de dispersão, o comportamento gregário e a adaptabilidade na alimentação e na reprodução, especialmente à medida que grandes áreas da paisagem foram convertidas em pastagens, são aparentemente as características responsáveis pelo sucesso da expansão populacional de *Bubulcus ibis* (TELFAIR, 1994). Apesar do conhecimento sobre as características de história de vida de *Bubulcus ibis* e a sua expansão populacional, há questões não respondidas sobre o processo de colonização do Brasil pela espécie.

**Figura 2** - Distribuição geográfica da subespécie ocidental de garça-vaqueira (*Bubulcus ibis ibis*).



Legenda: Áreas de residência permanente (verde), de ocorrência sem reprodução confirmada (azul) e de ocorrência somente durante o período reprodutivo (amarelo).

Fonte: (KUSHLAN; HANCOCK, 2005; SVENSSON; ZETTERSTROM, 2009).

## 2 *Bubulcus ibis* como modelo em estudos genéticos populacionais

Estudos genéticos têm contribuído com informações sobre as relações filogenéticas, as características de história de vida e a dinâmica de populações de *Bubulcus ibis* (CAMPANINI et al., 2012; CONGRAINS et al., 2016; MIÑO et al., 2017; MORALEZ-SILVA; DEL LAMA, 2017; MORALEZ-SILVA; DEL LAMA, 2014; ZHOU et al., 2014). Análises filogenéticas utilizando sequências completas do genoma mitocondrial de *Bubulcus ibis* e de outras 15 espécies de garças questionaram a classificação atual da família Ardeidae (ZHOU et al., 2014). Essas análises com o genoma mitocondrial sustentaram que *Bubulcus ibis* deveria ser classificada no gênero *Ardea* (ZHOU et al., 2014).

O primeiro estudo publicado sobre a genética populacional de *Bubulcus ibis* utilizou a informação da região controladora do DNA mitocondrial (mtDNA) a partir de indivíduos ( $n = 112$ ) de quatro populações brasileiras e quatro populações africanas da espécie (MORALEZ-SILVA; DEL LAMA, 2014). Esses autores encontraram um total de 35 haplótipos: 22 exclusivos para a África, 10 exclusivos para o Brasil e três compartilhados por ambas as amostras. O grau de diversidade genética foi semelhante entre o Brasil e a África e os dados de diferenciação genética sugeriram que a colonização do Brasil pela garça teria começado pela região sul e se expandido para as regiões sudeste e nordeste do país. Além disso, sinais de expansão demográfica foram detectados na amostra africana (MORALEZ-SILVA; DEL LAMA, 2014).

Em outro estudo, Congrains et al. (2016) investigou a estruturação genética de 15 colônias reprodutivas de *Bubulcus ibis* do oeste e sul africano, a partir da informação de sequências da região controladora e das ATPases 6 e 8 do mtDNA, bem como de um íntron do gene nuclear do fator de transformação do crescimento  $\beta 2$ . Os dados genéticos juntamente com uma modelagem de nicho ecológico suportaram a expansão no tamanho populacional e na distribuição geográfica da espécie na África durante o Último Máximo Glacial, assim como a ausência de estruturação genética entre as populações africanas estudadas (CONGRAINS et al., 2016).

Oito marcadores microssatélites polimórficos foram isolados para *Bubulcus ibis* usando bibliotecas genômicas enriquecidas com sondas biotiniladas (CAMPANINI et al., 2012). Uma investigação preliminar, baseada nesses oito *loci* microssatélites descritos por Campanini et al. (2012), revelou níveis de estruturação genética baixos entre as populações brasileiras e africanas de *Bubulcus ibis* (CAMPANINI, 2011). Essa estruturação genética baixa entre as populações foi atribuída à suposição de fluxo gênico recorrente entre as populações das áreas nativas e colonizadas ou a utilização de *loci* pouco informativos no estudo (CAMPANINI, 2011).

Dois estudos sobre a biologia reprodutiva na espécie *Bubulcus ibis* foram desenvolvidos (MIÑO et al., 2017; MORALEZ-SILVA; DEL LAMA, 2017). O primeiro, baseado em seis dos oito *loci* de microssatélites anteriormente prospectados (MORALEZ-SILVA; DEL LAMA, 2017), utilizou esse conjunto de dados para confirmar a presença de múltiplas fêmeas ovipositando em um mesmo ninho como evidência de biginia e parasitismo de ninho intraespecífico. O segundo investigou táticas alternativas de reprodução em *Bubulcus ibis*, com um número maior de microssatélites informativos, inferindo com confiança as relações de parentesco entre indivíduos amostrados em um

mesmo ninho (MIÑO et al., 2017). No estudo de Miño et al. (2017), a informação de 14 *loci* de microssatélites (seis recém prospectados) foi utilizada para encontrar evidências genéticas de paternidade extra-par e parasitismo de ninho intraespecífico, e suporte para um sistema de acasalamento não monogâmico em *Bubulcus ibis*.

Esses estudos anteriormente desenvolvidos pelo nosso grupo de trabalho demonstraram a necessidade de isolamento de um número maior de marcadores microssatélites polimórficos para inferir com maior precisão sobre a genética populacional e familiar de *Bubulcus ibis*. Além disso, a exploração de outros tipos de marcadores genéticos, como polimorfismos de nucleotídeo único (SNPs), se mostrou como uma alternativa para a investigação da estruturação genética das populações de *Bubulcus ibis* após a colonização do Brasil.

### 3 *A proposta geral do trabalho*

Neste trabalho, avaliou-se como a variação genética está distribuída entre populações de *Bubulcus ibis* nas áreas nativas, no continente africano e em Portugal, e nas áreas colonizadas, no Brasil. A meta principal foi inferir o padrão atual de estruturação genética e uma possível rota de invasão a partir da variação genética e da conectividade entre áreas onde a espécie distribuía-se antes da colonização das Américas e áreas do território brasileiro. Um estudo preliminar com enfoque na população de Fernando de Noronha foi desenvolvido para avaliar possíveis mudanças nos tamanhos populacionais e na diversidade genética em anos recentes. Para viabilizar essas investigações e trabalhos futuros, dados gerados por métodos de sequenciamento de próxima geração foram utilizados para a montagem de um rascunho do genoma de *Bubulcus ibis* e a prospecção de marcadores microssatélites e SNPs. As informações neste trabalho foram divididas em dois capítulos:

- a) capítulo I – Novos marcadores genéticos para a garça-vaqueira (*Bubulcus ibis*): prospecção, caracterização e um estudo de caso com a população de Fernando de Noronha
- b) capítulo II – Um século após a colonização do Brasil pela garça-vaqueira (*Bubulcus ibis*): distribuição da variação genética e conectividade entre as populações naturais da espécie

## II - CAPÍTULO I

NOVOS MARCADORES GENÉTICOS PARA A GARÇA-VAQUEIRA (*BUBULCUS IBIS*): PROSPECÇÃO, CARACTERIZAÇÃO E UM ESTUDO DE CASO COM A POPULAÇÃO DE FERNANDO DE NORONHA

## 1 INTRODUÇÃO

### 1.1 *A distribuição de marcadores microssatélites e SNPs nos genomas das aves*

Os genomas das aves são menores e mais simples em comparação aos genomas de outros amniotas em virtude de terem menos elementos repetitivos e DNA não codificante (ICGSC, 2004; ELLEGREN, 2005). O tamanho médio dos genomas das aves é de 1,45 bilhão de bases nitrogenadas (1,45 Gb), correspondendo a aproximadamente um terço do tamanho médio dos genomas dos mamíferos (GREGORY, 2005; WALTARI; EDWARDS, 2002). A adaptação ao voo é apontada como um dos principais fatores que favoreceu a manutenção de genomas menores, uma vez que reduz os custos metabólicos associados ao tamanho do genoma e das células (HUGHES; HUGHES, 1995). De fato, aves que não voam possuem genomas maiores do que as aves que voam, e morcegos possuem genomas menores em relação aos seus grupos irmãos de mamíferos (HUGHES; HUGHES, 1995; VAN DEN BUSSCHE; LONGMIRE; BAKER, 1995). Uma análise comparativa entre os genomas completos de 10 espécies de mamíferos placentários e de 24 espécies de aves indicou que a manutenção do tamanho reduzido do genoma das aves se dá principalmente por eventos de deleção de regiões não codificantes, o que por consequência neutraliza os ganhos de DNA resultantes da expansão de elementos transponíveis (KAPUSTA; SUH; FESCHOTTE, 2017). Em concordância com esse modelo de evolução, o tamanho reduzido do genoma das aves quando comparado ao genoma dos mamíferos é explicado pela redução no número de sequências repetitivas, pseudogenes e duplicações segmentares (BURT et al., 1999; ELLEGREN, 2005). As sequências repetitivas dispersas representam mais de 50% do genoma dos mamíferos, mas menos de 15% do genoma das aves (BURT et al., 1999). A representatividade baixa de sequências repetitivas no genoma das aves também se aplica às sequências repetitivas simples, os microssatélites (PRIMMER et al., 1997). Análises de hibridação dot-blot revelaram que microssatélites com repetições de di-, tri- e tetranucleotídeos estão presentes no genoma das aves com aproximadamente metade da frequência em que são encontrados no genoma dos mamíferos (PRIMMER et al., 1997).

Devido principalmente às suas taxas mutacionais elevadas, os microssatélites têm sido os marcadores moleculares mais utilizados em estudos genéticos populacionais e investigações sobre sistemas de acasalamento e relações de parentesco (ALI et al., 2019;

NIETLISBACH et al., 2017; OJEDA et al., 2017). Entretanto, a prospecção de marcadores microssatélites pode ser desafiadora em espécies que não são organismos-modelos. Pelos métodos tradicionais utilizados para o isolamento de *loci* de microssatélites, as sequências repetitivas são capturadas usando sondas biotiniladas e seletivamente ligadas a esferas magnéticas. Após várias etapas de lavagem, os fragmentos de DNA capturados são eluídos, amplificados e clonados para produzir uma biblioteca de clones enriquecida para as sequências dos microssatélites. Os clones positivos, contendo as sequências-alvo, são selecionados e sequenciados (ZANE; BARGELLONI; PATARNELLO, 2002). Embora relativamente simples, essas abordagens tradicionais podem consumir muito tempo e têm baixa eficiência para espécies com frequências baixas de microssatélites ao longo do genoma (ABDELKRIM et al., 2009). Como exemplo, o percentual médio de recuperação de clones positivos foi reportado como sendo menor para as aves (0,46%) em comparação com mamíferos (1,67%), reptéis (1,4%), peixes (3,1%), moluscos (1,96%), crustáceos (2,4%), artrópodes (2%) e plantas (2,3%) (ZANE; BARGELLONI; PATARNELLO, 2002).

Uma alternativa ao uso de marcadores microssatélites é a aplicação dos SNPs, marcadores presentes em altas frequências e espalhados por todo o genoma (MORIN; LUIKART; WAYNE, 2004). Por exemplo, foram identificados 2,8 milhões de SNPs em uma estimativa baseada em três genomas completos da galinha doméstica (*Gallus gallus*) (ELLEGREN, 2005). Além do grande número, a aplicação dos SNPs é vantajosa pois possibilita a verificação da variação genética seletivamente neutra ou adaptativa no genoma (LUIKART et al., 2003; MORIN; LUIKART; WAYNE, 2004). Em relação aos marcadores microssatélites, os SNPs têm um menor potencial para homoplasias e uma distribuição mais densa e uniforme dentro dos genomas, tornando-se úteis para estudos populacionais e de mapeamento gênico (MORIN; MARTIEN; TAYLOR, 2009; XING et al., 2006). Os benefícios dos SNPs incluem a aplicabilidade universal das sequências de nucleotídeos e a flexibilidade nos protocolos de detecção, gerando genótipos universalmente comparáveis e que não requerem padronização em plataformas de detecção distintas (KIM et al., 2008).

## 1.2 O advento das tecnologias de sequenciamento de próxima geração na identificação de marcadores moleculares e em estudos populacionais

Novas tecnologias têm surgido na última década permitindo a descoberta de marcadores moleculares mais rapidamente e em maior número (ABDELKRIM et al., 2009; MORIN; MARTIEN; TAYLOR, 2009). Atualmente, os principais métodos de isolamento de marcadores microssatélites e SNPs estão baseados nas tecnologias de sequenciamento de próxima geração (NGS). O uso de dados gerados por NGS facilitou o desenho de *primers* para as regiões flanqueadoras de microssatélites e aumentou as possibilidades de se encontrar *loci* polimórficos, isto porque a análise nesse protocolo inclui uma quantidade maior de sequências contendo unidades repetitivas do que a abordagem anterior (COWEN et al., 2013; HORREO et al., 2013). Em comparação aos métodos tradicionais, as metodologias baseadas nos dados de NGS não requerem bibliotecas e sondas individualizadas para cada tipo de repetição porque preveem a detecção de repetições usando ferramentas de bioinformática (ABDELKRIM et al., 2009; YU et al., 2011). Além disso, uma execução mínima na plataforma NGS é suficiente para a produção de dados para a identificação de microssatélites, uma vez que várias amostras podem ser sequenciadas concomitantemente e não se faz necessária uma cobertura alta de profundidade do genoma (YU et al., 2011). Estudos têm demonstrado o sucesso do uso do NGS no isolamento dos microssatélites em organismos não modelos de diversos grupos taxonômicos, incluindo o grupo das aves (BASTÍAS et al., 2016; COWEN et al., 2013; GONZÁLEZ-CASTELLANO et al., 2018; HORREO et al., 2013; SAHU et al., 2014).

As técnicas baseadas no NGS permitem a descoberta e genotipagem de milhares de SNPs em centenas de indivíduos a partir do sequenciamento de bibliotecas de DNA genômico, em um único passo (DAVEY et al., 2011; MASTRETTA-YANES et al., 2014; POOTAKHAM et al., 2015). Mesmo sem o sequenciamento do genoma completo, a cobertura de todo o genoma pode ser aproximada por meio de uma variedade de métodos de NGS, que a cada dia se tornam mais rápidos, de menor custo e com maior precisão (DAVEY et al., 2011). Métodos de NGS utilizando enzimas de restrição e bibliotecas de representação reduzida do genoma, tais como o GBS (*Genotyping-by-sequencing*) (ELSHIRE et al., 2011) e o RAD-Seq (*Restriction site Associated DNA*) (BAIRD et al., 2008), apresentam um rendimento alto durante a descoberta e genotipagem de SNPs em organismos não-modelos (BOEHM et al., 2015; ELLEGREN, 2014; POLAND; RIFE, 2012). A facilidade de se encontrar densidades altas de SNPs por todo o genoma com métodos de NGS tem estimulado o uso dos SNPs para solucionar muitas questões ecológicas e evolutivas (ANDREWS; LUIKART, 2014; HEMMER-HANSEN;

THERKILDSEN; PUJOLAR, 2014; NARUM et al., 2015). SNPs identificados por NGS têm sido utilizados com sucesso, por exemplo, na busca de *loci* de caracteres quantitativos (QTLs), na localização de regiões genômicas com semelhanças ou divergências adaptativas, na identificação de diferenciação genética significativa entre populações, na reconstrução filogenética de espécies, em estudos de filogeografia, na investigação de padrões de expressão alelo-específicos e na identificação de regiões genômicas envolvidas na hibridação (DE WIT; PESPENI; PALUMBI, 2015; GORJANC et al., 2015; HANS et al., 2015; LARSON et al., 2014).

A aplicação de métodos genômicos no grupo das aves cresceu nos últimos anos desde a publicação do artigo “*Prospects for the use of next-generation sequencing methods in ornithology*” de Lerner e Fleischer (2010). As aves são particularmente interessantes à aplicação de métodos genômicos, por apresentarem tamanho pequeno do genoma, a ploidia e a estrutura dos cromossomos conservados (ELLEGBREN, 2014; TOEWS et al., 2015). O tamanho reduzido do genoma e a menor representatividade de elementos repetitivos e rearranjos em comparação a outros grupos taxonômicos, refletem em uma redução dos custos e permitem que genomas existentes de outras aves, mesmo distantes filogeneticamente, possam ser usados como genomas de referência (TOEWS et al., 2015; ZHANG et al., 2014). Toews et al. (2015) reuniram estudos genômicos que investigaram a variação genética e a especiação em aves não modelos. Dos 29 estudos reunidos na revisão, 17 publicações utilizaram com sucesso estratégias de NGS com o uso de bibliotecas de representação reduzida em estudos de estruturação populacional, genômica comparativa ou filogenômica. Segundo os autores, estes métodos são mais adequados para estudos populacionais ou entre espécies proximamente relacionadas, destacando a sua importância em investigações de estrutura populacional e na busca de padrões de divergência genômica em grupos com pouca diferenciação (TOEWS et al., 2015).

### 1.3 Estudo de caso: A garça-vaqueira (*Bubulcus ibis*) no arquipélago de Fernando de Noronha

Ilhas oceânicas estão entre os ambientes mais vulneráveis do planeta devido a ameaças como a elevação do nível do mar, a superpopulação humana, o uso insustentável de recursos naturais e a introdução de espécies exóticas invasoras (BLACKBURN et al.,



2004; COURCHAMP et al., 2014; JONES et al., 2008; MEDINA et al., 2011). Fernando de Noronha é um arquipélago vulcânico composto por 21 ilhas, localizado no oceano Atlântico a 345 km a partir da costa do nordeste do país (IBAMA; FUNATURA, 1990). Desde o século XVIII, o arquipélago tem sido progressivamente alterado por ações humanas que foram responsáveis, por exemplo, pela exploração madeireira, criação de pastagens e plantações, mineração e introdução de espécies exóticas (VITALI, 2009). A introdução nas ilhas de espécies exóticas pode se tornar uma ameaça às espécies nativas pela competição e predação, modificação do habitat, transmissão de doenças e mudanças na dinâmica trófica (SIMBERLOFF et al., 2013). Duas espécies de aves exóticas continentais estão presentes no arquipélago de Fernando de Noronha desde a década de 1980 e são consideradas potencialmente invasoras: o pardal-doméstico, *Passer domesticus* (Linnaeus, 1758), e a garça-vaqueira, *Bubulcus ibis* (MESTRE; ROOS; RECHETELO, 2016; NUNES et al., 2010). Embora, aparentemente, *Bubulcus ibis* não tenha sido introduzida diretamente por humanos, elas são consideradas espécies exóticas porque sua colonização provavelmente foi facilitada pela antropização dos habitats e a introdução de animais pastadores (PHILLIPS; WIEDENFELD; SNELL, 2012; TELFAIR, 1994).

No continente, *Bubulcus ibis* pode até ter um impacto economicamente positivo na agropecuária devido a sua dieta alimentar que pode incluir majoritariamente organismos considerados pragas agrícolas ou vetores de doenças (KUSHLAN; HANCOCK, 2005; TELFAIR, 1994). Por outro lado, a alta competitividade interespecífica e o comportamento predatório de *Bubulcus ibis* podem resultar em impactos negativos à biodiversidade nativa em áreas recém-colonizadas (AVERY; TILLMAN, 2005; BURGER, 1978; MEESE, 2012; NUNES et al., 2010). Por exemplo, no arquipélago de Fernando de Noronha, há registros de *Bubulcus ibis* predando uma espécie endêmica de lagarto, *Trachylepis atlantica* (Schmidt, 1945), e interagindo agonisticamente com *Sula sula* (atobá-de-patas-vermelhas; Linnaeus, 1766) em sítios de nidificação (SILVA-JR.; PÉRES-JR.; SAZIMA, 2005; BARBOSA-FILHO et al., 2009). Além disso, *Bubulcus ibis* mostrou-se um risco para o tráfego de aviões devido a sua presença nas proximidades do aeroporto, onde colisões entre indivíduos da espécie e aviões já foram registradas (NUNES et al., 2010). Para minimizar esses impactos, o manejo ecológico da espécie com eventos de captura e eutanásia foi realizado entre 2005 e 2015, visando o controle do tamanho populacional no arquipélago (ADEFN, 2012; 2015).

A presença de *Bubulcus ibis* no arquipélago de Fernando de Noronha foi observada pela primeira vez em 1986 com o avistamento de dois indivíduos (NACINOVIC; TEIXEIRA, 1989). Durante as décadas de 1980 e 1990, a espécie foi classificada como visitante ocasional devido a sua ocorrência esporádica no arquipélago (NACINOVIC; TEIXEIRA, 1989; SCHULZ-NETO, 2004). Entre 2001 e 2003 o tamanho populacional aumentou de 54 para aproximadamente 400 indivíduos e as primeiras colônias reprodutivas foram localizadas (NUNES et al., 2010; SILVA; SILVA, 2003). Em 2005, antes do início do manejo ecológico, a população de *Bubulcus ibis* foi estimada em 655 indivíduos distribuídos em uma área de aproximadamente 20 km<sup>2</sup> (NUNES et al., 2010). Após o início das atividades de manejo, o tamanho populacional reduziu gradativamente, chegando a uma estimativa média de 298 indivíduos em novembro de 2008 durante o último censo populacional reportado (ADFN, 2012; NUNES et al., 2010). Após 2008, mesmo sem estimativas oficiais do tamanho populacional, indivíduos foram retirados da população e eutanasiados: 274 em 2009, 241 em 2010, 168 em 2012 e 88 em 2015 (ADFN, 2012; 2015). Atualmente, o tamanho populacional de *Bubulcus ibis* em Fernando de Noronha é desconhecido, mas sabe-se que é uma das espécies de aves mais facilmente avistadas em todo o arquipélago.

#### 1.4 A proposta de trabalho: capítulo I

Atendendo à demanda de prospecção de um número maior de marcadores genéticos polimórficos, propôs-se utilizar dados gerados por métodos de NGS para o isolamento desses marcadores de forma mais rápida e eficiente. A meta principal dessa proposta foi de isolar pelo menos 15 marcadores microssatélites polimórficos e identificar de um painel com milhares SNPs ao longo do genoma de *Bubulcus ibis*. Para viabilizar o isolamento dos *loci* de microssatélites, foi feita a montagem do primeiro genoma de referência para a espécie (Anexos I e II). Os marcadores microssatélites isolados foram caracterizados em populações naturais de *Bubulcus ibis* para identificar polimorfismos e determinar as frequências alélicas (Anexos I e II). Neste capítulo, ainda foram descritas as etapas para a identificação dos SNPs no genoma de *Bubulcus ibis* e as filtragens necessárias para a formação do conjunto de dados utilizado no Capítulo II na investigação da colonização do Brasil pela espécie.

Ainda nessa proposta, avaliou-se a informatividade dos *loci* de microssatélites na discriminação entre relações de parentesco e a utilidade desses marcadores na determinação dos níveis de diversidade e estruturação genética populacional. Os marcadores microssatélites prospectados foram aplicados em um estudo de caso com a população que colonizou recentemente o arquipélago de Fernando de Noronha com o intuito de prever as alterações na diversidade genética e no tamanho populacional após o encerramento das atividades de manejo ecológico da espécie na região.

## 2 OBJETIVOS

### 2.1 *Objetivo geral*

Ampliar o número de marcadores genéticos polimórficos disponíveis para a garça-vaqueira, *Bubulcus ibis*, isolando microssatélites e identificando SNPs no genoma da espécie. Ainda, aplicando os microssatélites prospectados em um estudo de caso com a população da espécie presente no arquipélago de Fernando de Noronha.

### 2.2 *Objetivos específicos*

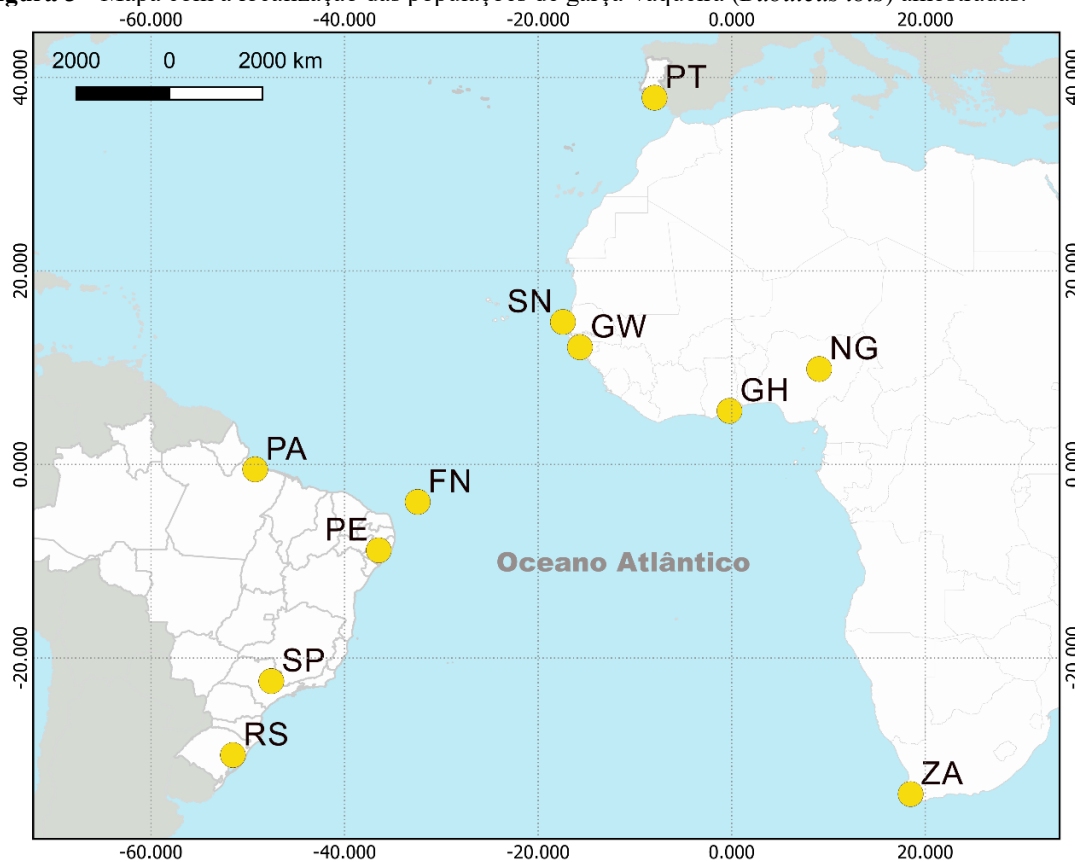
- a. Montar e anotar um genoma de referência para *Bubulcus ibis*, a partir de dados de sequenciamento de próxima geração;
- b. Isolar *loci* de microssatélites polimórficos no genoma de referência de *Bubulcus ibis*;
- c. Caracterizar os *loci* de microssatélites isolados em populações naturais de *Bubulcus ibis*;
- d. Aplicar os *loci* de microssatélites recém-prospectados em um estudo com a população de *Bubulcus ibis* do arquipélago de Fernando de Noronha;
- e. Identificar um painel de SNPs no genoma de *Bubulcus ibis* a partir de dados gerados por *Genotyping-by-sequencing*.

### 3 MATERIAL E MÉTODOS

#### 3.1 Coleta de material biológico e extração do DNA genômico

Amostras de sangue de *Bubulcus ibis* foram coletadas entre 2009 e 2017 durante incursões coordenadas pela equipe do Laboratório de Genética de Aves (DGE, UFSCar) e por colaboradores da Nigéria e de Portugal (Tabelas 1 e 2; Figura 3). No geral, somente ninhegos foram amostrados, com a exceção das populações de Pernambuco (PE) e Fernando de Noronha (FN) que são representadas por sangue de adultos. Coletou-se entre 0,01 a 0,05 ml de sangue a partir da veia jugular dos indivíduos usando uma seringa abluída em EDTA (0,3%) e uma agulha hipodérmica de parede fina trifacetada estéril. O sangue coletado foi mantido em etanol absoluto e refrigerado até ser armazenado a  $-20^{\circ}\text{C}$  no laboratório. O DNA genômico de cada uma dessas amostras foi extraído com o kit *DNeasy Blood & Tissue* (QIAGEN<sup>®</sup>), seguindo o protocolo do fabricante.

**Figura 3** - Mapa com a localização das populações de garça-vaqueira (*Bubulcus ibis*) amostradas.



Legenda: Os círculos amarelos representam as populações amostradas no Brasil no Rio Grande do Sul (RS), São Paulo (SP), Pernambuco (PE), Fernando de Noronha (FN) e Pará (PA), e nos países África do Sul (ZA), Nigéria (NG), Gana (GH), Guiné Bissau (GW), Senegal (SN) e Portugal (PT).

Fonte: Produção do próprio autor.

**Tabela 1** - Informações sobre as populações de garça-vaqueira (*Bubulcus ibis*) amostradas e utilizadas para a prospecção e caracterização de *loci* microssatélites e análises populacionais subsequentes.

País (Região)	Colônia (Abreviação)	Coordenadas Geográficas	<i>n</i>	Ano de coleta	Etapa	
Senegal (Thiès)	Reserva Somone	N 14° 29' 51.66" O 17° 4' 55.86"	1	2011	Bibliotecas genômicas <i>paired-end</i> Illumina	
África do Sul (Western Cape)	Paarl	S 33° 41' 3.06" L 18° 59' 1.86"	1	2011		
Brasil (Pará)	Ilha do Marajó	S 0° 30' 19.5" O 49° 13' 46.08"	1	2012		
Brasil (Rio Grande do Sul)	Eldorado do Sul	S 30° 1' 50.84" O 51° 31' 47.78"	1	2009		
Brasil (São Paulo)	Parque Lago Azul (SP)	S 22° 23' 30.83" O 47° 33' 51.02"	4	2014	Testes de polimorfismos dos 22 <i>loci</i> de microssatélites prospectados	
Brasil (Pernambuco)	Garanhuns (PE)	S 8° 52' 0" O 36° 28' 0"	4	2010		
Brasil (Rio Grande do Sul)	Eldorado do Sul (RS)	S 30° 1' 50.84" O 51° 31' 47.78"	3	2009		
Brasil (Pará)	Ilha do Marajó (PA)	S 0° 30' 19.5" O 49° 13' 46.08"	4	2012		
Senegal (Dakar)	Parque florestal de Hann (SN)	N 14° 43' 39.18" O 17° 25' 53.70"	4	2011		
Gana (Accra)	Lagoa Korle (GH)	N 5° 32' 17.51" O 0° 13' 12.14"	4	2011		
África do Sul (Western Cape)	Reserva Natural Rondevlei (ZA)	S 34°3' 45.54" L 18° 29' 43.5"	4	2011		
Nigeria (Plateau)	Rodovia Fobur (NG)	N 9° 51' 51.62" L 9° 2' 44.95"	4	2011		
África do Sul (Western Cape)	Reserva Natural Rondevlei (ZA)	S 34°3' 45.54" L 18° 29' 43.5"	29	2011		Caracterização dos 22 <i>loci</i> de microssatélites prospectados
Brasil (São Paulo)	Parque Lago Azul (SP)	S 22° 23' 30.83" O 47° 33' 51.02"	32	2014		
África do Sul (Western Cape)	Reserva Natural Rondevlei (ZA)	S 34°3' 45.54" L 18° 29' 43.5"	29	2011	Análises populacionais com 25 <i>loci</i> de microssatélites	
Brasil (São Paulo)	Parque Lago Azul (SP)	S 22° 23' 30.83" O 47° 33' 51.02"	32	2014		
Brasil (Fernando de Noronha)	Praia do Leão (FN)	S 03°52'00.00" O 032°26'00.00"	31	2010		

Fonte: Produção do próprio autor.

**Tabela 2** - Informações sobre as populações de garça-vaqueira (*Bubulcus ibis*) amostradas e utilizadas para a identificação de polimorfismos de nucleotídeo único (SNPs) por *Genotyping-by-sequencing*. A localização, os anos de coleta e os números de indivíduos (*n*) iniciais e (finais) de garça-vaqueira utilizados na identificação dos SNPs e nas análises populacionais subsequentes são exibidos.

País (População)	Colônia	Abreviação	Ano de coleta	<i>n</i>
Portugal (Beja)	Srª da Graça de Padrões	PT	2017	17 (9)
Guiné Bissau (Bolama)	Ilha de Bandim	GW	2011	18 (17)
Senegal (Dakar)	Parque Florestal de Hann	SN	2011	17 (16)
Gana (Grande Acra)	Lagoa Korle	GH	2011	21 (16)
Nigéria (Plateau)	Rodovia Fobur	NG	2011	20 (15)
África do Sul (Cabo Ocidental)	Reserva Natural Rondevlei	ZA	2011	21 (17)
Brasil (Pará)	Ilha do Marajó	PA	2012	20 (17)
Brasil (Fernando de Noronha)	Praia do Leão	FN	2010	17 (17)
Brasil (Pernambuco)	Garanhuns	PE	2010	17 (17)
Brasil (São Paulo)	Parque Lago Azul	SP	2014	19 (17)
Brasil (Rio Grande do Sul)	Eldorado do Sul	RS	2009	20 (16)
Total				207 (174)

Fonte: Produção do próprio autor.

### 3.2 Montagem e anotação gênica do genoma de *Bubulcus ibis*

Amostras de DNA genômico de quatro indivíduos de *Bubulcus ibis* (Tabela 1) foram utilizadas na preparação de bibliotecas genômicas *paired-end* usando o kit *Nextera® DNA* (Illumina) e seguindo as instruções do fabricante. As bibliotecas com os fragmentos de DNA foram corridas em duas *lanes* em uma plataforma HiScan-SQ (Illumina, Inc., San Diego, CA, USA). O *pipeline* CASAVA v1.8.2 (Illumina Inc., San Diego, CA, USA) foi utilizado para realizar as chamadas de bases dos dados brutos, transformando-os em sequências de nucleotídeos (*reads*) no formato FastQC, sendo acompanhadas de pontuações de qualidade *Phred*. As pontuações de qualidade de *Phred* são definidas como uma propriedade que é logaritmicamente relacionada às probabilidades de erro da chamada de base (EWING; HILLIER; WENDL, 1998). A qualidade dos fragmentos de DNA foi avaliada usando o programa FastQC v0.11.5 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Usando o programa SeqyClean v1.10.06 (ZHBANNIKOV et al., 2017), fragmentos foram excluídos da análise quando apresentavam qualidade  $\leq 24$  numa escala *Phred* em mais da metade das chamadas de bases, com sequências de adaptadores e vetores detectadas usando o algoritmo BLAST na base de dados NCBI UniVec

(<http://www.ncbi.nlm.nih.gov/VecScreen/UniVec.html>) e que apresentavam tamanho menor do que 65 pb.

Montagens *de novo* dos genomas de *Bubulcus ibis* foram realizadas com o programa Velvet v1.2.0.9 usando gráficos de Bruijn (SCHULZ et al., 2012; ZERBINO; BIRNEY, 2008). Um gráfico de Bruijn é uma representação baseada em sequências curtas (*k-mers*), ideal para conjuntos de dados com alta cobertura e sequências de nucleotídeos entre 25-50 pb (ZERBINO; BIRNEY, 2008). Nas montagens, as bibliotecas sequenciadas foram consideradas tanto separadamente quanto em sua totalidade de sequências. Nas montagens, testou-se primeiro comprimentos de *k-mers* entre 31 e 71 pb em intervalos de 10 pb e avaliou-se a distribuição dos números de sequências consenso por cobertura (ou profundidade), usando um *script* customizado do projeto Bioconductor (GENTLEMAN et al., 2004). Com essa abordagem, determinou-se qual a cobertura que apresentava o maior número de sequências consenso, no caso, três vezes. Depois, um novo conjunto de montagens foi realizado definindo um limite mínimo de cobertura de três vezes. A qualidade das montagens foi avaliada comparando-se os valores de N50 (comprimento médio das sequências consenso), o número de sequências consenso maiores do que 1000 pb e o tamanho da sequência consenso mais longa em cada montagem. A melhor montagem foi definida como o genoma de referência de *Bubulcus ibis*. Foi nessa base de informação foi feita uma anotação preliminar dos genes, comparando-se o genoma de referência com a base de dados de proteínas Swiss-Prot (BOECKMANN et al., 2003).

### 3.3 *Prospecção, amplificação e genotipagem de loci de microssatélites no genoma de referência de Bubulcus ibis*

Microssatélites ao longo do genoma de referência de *Bubulcus ibis* foram identificados usando o *pipeline* QDD v3.0 (MEGLÉCZ et al., 2009). O *pipeline* QDD remove sequências contaminantes de adaptadores e vetores, elimina as sequências redundantes, removendo possíveis *loci* duplicados e elementos de transposição, e identifica *loci* de microssatélites com no mínimo cinco repetições (MEGLÉCZ et al., 2009). Pares de sequências de oligonucleotídeos flanqueando os microssatélites identificados foram desenhadas pelo programa PRIMER3 v0.4.0 (KORESSAAR; REMM, 2007; UNTERGASSER et al., 2012). Critérios foram estabelecidos para selecionar sequências com tamanho máximo de 250 pb, contendo microssatélites perfeitos e com no mínimo dez repetições de dinucleotídeos. Para o desenho das sequências de

oligonucleotídeos, estabeleceu-se um limite máximo de 60% de bases GC, comprimento entre 18 pb e 25 pb, probabilidade máxima de 2% para *hairpin loop*, temperaturas de anelamento mínimas de 53 °C e máximas de 62°C, e diferença máxima de 2 °C entre as temperaturas de anelamento das sequências *forward* e *reverse*. No total, 22 *loci* foram selecionados para os testes de amplificação e de polimorfismos (Tabela 3).

**Tabela 3** - Características dos *loci* de microssatélites da garça-vaqueira (*Bubulcus ibis*) selecionados para verificar a existência de polimorfismos. As sequências de oligonucleotídeos, *forward* (F) e *reverse* (R), os motivos de repetições e as temperaturas de anelamento são mostradas para os 22 *loci* selecionados.

<i>Locus</i>	Sequência de oligonucleotídeos (5'-3')	Repetição	TA (°C)
Bi31	F: CCTTTCAGTGTAATACCAACCG R: AAAGCAGCATCCGCAGTAAT	(AC) <sub>10</sub>	55
Bi37	F: TGAATTCAGGAGGAAAGCAGG R: GGTGTGGCCTGCATGTTTAC	(GT) <sub>10</sub>	57
Bi39	F: ACCACCTGATTCCACTTCCC R: GCGGCCCTGAACAGATTTAC	(TC) <sub>10</sub>	56
Bi40	F: GCTGAATCCTCGTGTGGT R: CCTCCTTACCTTGCGATTT	(AG) <sub>10</sub>	56
Bi42	F: CCTGGAAGCCAAGGTGATTT R: TCCTTTGCACTAACAAGCCC	(TG) <sub>11</sub>	58
Bi44	F: GAGCATTTCCACCTCCAGC R: GTGGATGGAAAGCATCGAAA	(TA) <sub>10</sub>	55
Bi45	F: GGTTTAATTTGCAGAGCAATGAG R: TTCGTATTTGCTGCCCAAAC	(CA) <sub>12</sub>	57
Bi46	F: GGAGACTCCGGGAAGCCT R: CCGTCGGGGTTGTGCAAT	(TG) <sub>10</sub>	60
Bi47	F: ATCCCGCCTAGAGATGCCTT R: TCCAGGCTAACTTCTGCATGA	(GT) <sub>11</sub>	60
Bi48	F: ATGGAAAGTGCTTAGCCTGC R: TGACAGAGGAACCAAGAGCT	(AG) <sub>11</sub>	60
Bi49	F: GACCAATAATCGCCTTTGGA R: GCACTGAGAGCGATTCACAA	(GA) <sub>14</sub>	58
Bi50	F: AGCGCTAACCCACGGATG R: CTGCAAGCGTTGGGTTCTC	(CA) <sub>11</sub>	61-60
Bi51	F: TTGCATACACGCTACTGTGC R: ACATAGGAGCAGGGCGTAAA	(TG) <sub>10</sub>	60
Bi52	F: AGCTTGTGAGATATTTGTGGCA R: ATTTGCTGGGTGTGGAGGTT	(TG) <sub>10</sub>	60
Bi53	F: GCAAAGTTGTAAATACGAAGCG R: ACGCTCACCCACCTACAAG	(GT) <sub>12</sub>	58
Bi54	F: CGCGCCACTGATTTGATACA R: ACTTACAGTGAACGCCTAGC	(TG) <sub>13</sub>	60
Bi55	F: AGTCTGAATCTGGGTGTGGG R: ATCTTCTGGCTGAGCGTGTT	(AC) <sub>13</sub>	61-60
Bi56	F: CCCAGCGGCAGATATCTACA R: ACAAGTGCCCCTTCTCCAC	(TC) <sub>15</sub>	61-60
Bi57	F: GGTGACCCAAGATGTTAGCAC R: ACAAACCTCTACTTCTTCTGTG	(AG) <sub>15</sub>	60
Bi58	F: TTGTCCCCTTAGCCTGTCTG R: GGTACCGTGAAATGATGGAG	(TG) <sub>10</sub>	60
Bi59	F: GAGCATCGATCCATGGGG R: CAGAGTTTGTGTCCCCTCCA	(CA) <sub>10</sub>	61-60
Bi60	F: ACACCTCACACTGCATTAAGT R: CCACAGCGGGTAAATTAGGATT	(AC) <sub>13</sub>	57-56

Fonte: Produção do próprio autor.



Para verificar se os *loci* selecionados eram polimórficos ou não, utilizou-se um conjunto de amostras ( $n = 31$ ) composto por indivíduos de *Bubulcus ibis* coletados tanto na área nativa africana quanto na área colonizada brasileira (Tabela 1). Os *loci* de microssatélites foram amplificados por reação em cadeia da polimerase (PCR), seguindo protocolos específicos desenvolvidos. As PCRs foram padronizadas com um volume total de 10  $\mu\text{L}$  contendo entre 36 e 40 ng de DNA, 0,2  $\mu\text{M}$  de sequências de oligonucleotídeos *forward* sintetizados com uma cauda 5'-M13 (TGTAACGACGGCCAGT) (SCHUELKE, 2000), 0,8  $\mu\text{M}$  de sequências de oligonucleotídeos universais M13 (-21) marcadas com o fluoróforo 6-FAM, 0,8  $\mu\text{M}$  de sequências de oligonucleotídeos *reverse*, 2,5 mM de  $\text{MgCl}_2$ , 0,37 mM de cada dNTP, 1x tampão de BD2 (Solis BioDyne) e 0,5 U da enzima DNA HOT FIREPol® Taq polimerase (Solis BioDyne). As PCRs para a amplificação do *locus* Bi50 foram otimizadas usando 40 ng de DNA e as amplificações nos *locus* Bi39 e Bi40 foram otimizadas usando 2,75 mM e 3,0 mM de  $\text{MgCl}_2$ , respectivamente.

As amplificações foram realizadas em termocicladores Veriti™ (Applied Biosystems®, Foster City, CA, EUA), inicialmente usando um perfil de ciclagem padrão: pré-aquecimento a 95 °C por 15 min, trinta ciclos iniciais (95 °C por 30 s, temperaturas de anelamento durante 45 s e 72 °C durante 30 s), dez ciclos finais (95 °C durante 30 s, 53 °C durante 45 s e 72 °C durante 30 s) e polimerização final a 72 °C durante 10 min. Os perfis de ciclagem foram modificados para aumentar o sucesso de amplificação em alguns *loci*. Para os *loci* Bi31, Bi37, Bi40, Bi42 e Bi45 o número de ciclos iniciais foi alterado de 30 para 37. Para os *loci* Bi50, Bi55, Bi56 e Bi59 o número de ciclos iniciais foi alterado de 30 para 20, e dez ciclos foram adicionados com temperaturas de anelamento decrescendo de 61 °C a 60 °C por uma abordagem *touchdown*. Para o *locus* Bi60 o número de ciclos iniciais foi alterado de 30 para 20, e dez ciclos foram adicionados com temperaturas de anelamento decrescendo de 67 °C a 56°C por uma abordagem *touchdown*. O sucesso de amplificação foi examinado usando eletroforese em gel de agarose a 2%. Os produtos das PCRs foram genotipados em um sequenciador automático MegaBACE®1000 (GE Healthcare, Bucks, UK) e os genótipos foram determinados usando o programa FRAGMENT PROFILER® v1.2 (GE Healthcare, Bucks, UK). Para a genotipagem, os produtos de amplificação foram diluídos com água deionizada estéril de 10 a 15 vezes, dependendo do seu rendimento, e combinados entre si em uma microplaca. Em cada poço da microplaca misturaram-se 2  $\mu\text{L}$  dos produtos de PCR diluídos e 8  $\mu\text{L}$  de uma solução composta por um marcador de peso molecular interno (500 pb Rox, The

Gel Company, São Francisco, CA, EUA) e Tween a 0,1%. %. Amplificações independentes em duplicatas para 10% dos indivíduos foram realizadas em todos os *loci* de microssatélites para minimizar os erros de genotipagem e de leitura e se confirmar os genótipos avaliados.

### 3.4 Caracterização dos *loci* de microssatélites prospectados em populações naturais de *Bubulcus ibis*

Para a caracterização das frequências alélicas nos *loci* de microssatélites prospectados, usando o programa Cervus v3.0.7 (KALINOWSKI; TAPER; MARSHALL, 2007), foi utilizado o conjunto de amostras empregadas, anteriormente, nos testes de polimorfismos e em mais duas populações naturais de *Bubulcus ibis* amostradas em Rio Claro (SP, Brasil,  $n = 32$ ) e Cape Town (WC, África do Sul,  $n = 29$ ) (Tabela 1). Para estimar o poder *multilocus* (PWR) dos *loci* recém-prospectados na inferência de parentesco, as frequências alélicas caracterizadas na população de SP foram utilizadas no programa KinInfor v2 (WANG, 2006), simulando 103 pares de genótipos para cada relação de parentesco testada ( $\alpha = 0,05$ ). Avaliou-se o PWR desses *loci* em distinguir irmãos completos e meios-irmãos de indivíduos não relacionados, bem como irmãos completos de meios-irmãos.

Para as análises populacionais, outros três *loci* espécie-específicos (Bi01, Bi20 e Bi30, Campanini et al., 2012) foram incluídos na análise, com o uso dos protocolos de laboratório descritos por Miño et al., (2017). Nessa etapa, uma amostra oriunda do arquipélago de Fernando de Noronha (Pernambuco, Brasil,  $n = 31$ ) foi incluída para estimar possíveis mudanças no tamanho populacional e na diversidade genética após o término das atividades de manejo ecológico (Tabela 1). Os genótipos foram checados para a evidência de alelos nulos, erros de leitura de picos inespecíficos e amplificação preferencial de alelos usando o programa MICRO-CHECKER v2.2.3 (VAN OOSTERHOUT et al., 2004). Avaliou-se as probabilidades de desequilíbrio de ligação ( $10^4$  dememorizações de MCMC,  $10^3$  *batches* e  $10^4$  iterações por *batch*) e de desvios do equilíbrio de Hardy-Weinberg (HW) ( $10^5$  dememorizações de MCMC,  $10^3$  *batches* e  $10^3$  iterações por *batch*) usando o pacote Genepop v4.7 do programa R v3.4.3 (ROUSSET, 2008; R DEVELOPMENT CORE TEAM, 2018). A neutralidade seletiva dos *loci* foi verificada pelo método de  $F_{ST}$ -outlier ( $10^5$  iterações, burn-in de  $5 \times 10^3$  e 20 corridas pilotos com  $5 \times 10^3$  de comprimento) no programa BayeScan v2.1 (FOLL; GAGGIOTTI,

2008). O método de correção Holm-Bonferroni (HOLM, 1979) foi usado no ajuste dos valores críticos de  $p$  ( $\alpha = 0,05$ ) para comparações múltiplas. O conteúdo informativo polimórfico (PIC) nos *loci* de microssatélites polimórficos foi computado usando o programa Cervus v3.0.7 (KALINOWSKI; TAPER; MARSHALL, 2007). O número de alelos exclusivos a cada população foi determinado pelo programa GENALEX v6.503 (PEAKALL; SMOUSE, 2006). Para se determinar a diversidade genética para cada *locus* e população, as heterozigosidades observada e esperada e o coeficiente de endogamia foram determinados usando o pacote diveRcity do programa R v3.4.3 (KEENAN et al., 2013). A riqueza alélica foi calculada pelo método de rarefação de alelos no programa HP-RARE v1.0 (KALINOWSKI, 2005). Comparações entre os índices de diversidade genética determinados para as populações amostradas foram realizadas usando o teste de Wilcoxon ranqueado no ambiente R (R DEVELOPMENT CORE TEAM, 2018).

A diferenciação genética entre as populações amostradas de *Bubulcus ibis* foi primeiramente investigada pelo cálculo dos valores de  $F_{ST}$  (global e par-a-par) (WEIR; COCKERHAM, 1984) usando o pacote Genepop v4.7 do programa R v3.4.3 (ROUSSET, 2008) e, em seguida por uma análise de variância molecular (AMOVA,  $10^5$  passos de memorização e MCMC) no programa ARLEQUIN v3.5.2.2 (EXCOFFIER; LISCHER, 2010). A existência de estruturação genética entre as populações foi investigada usando o algoritmo Bayesiano de agrupamento do programa STRUCTURE v2.3.4 (PRITCHARD; STEPHENS; DONNELLY, 2000) e uma análise complementar de agrupamento espacial Bayesiano no programa TESS v3 (CAYE et al., 2015). No programa STRUCTURE, utilizou-se os locais de amostragem como informação *a priori* e assumiu-se um modelo de mistura com frequências alélicas correlacionadas, sendo que o grau de mistura foi inferido a partir dos dados. Foram realizadas 20 corridas para cada valor de  $K$  (1 a 5), com  $10^6$  iterações de MCMC e *burn-in* de  $5 \times 10^4$ . O número mais provável de grupos genéticos foi determinado avaliando-se os valores de Delta  $K$  no programa POPHELPER v1 (EVANNO; REGNAUT; GOUDET, 2005; FRANCIS, 2017). No programa TESS, foram considerados tanto um modelo sem mistura quanto os modelos CAR e BYM. O modelo com o menor valor para o critério de desvio de informação (DIC) e estabilização no menor número de grupos genéticos inferidos foi escolhido como o que explicava melhor a variação genética nos dados. Cinco corridas exploratórias foram realizadas testando valores de  $K$  entre 1 e 8 para se determinar o melhor modelo, usando  $10^5$  sweeps e *burn-in* de  $2 \times 10^4$ . Na análise definitiva no programa TESS, foram feitas 20 corridas por valor de  $K$  (entre 1 e 11) no modelo BYM, usando  $7 \times 10^4$  sweeps e *burn-in* de  $10^4$ .

Padrões de migração foram inferidos ( $\alpha = 0,01$ , re-amostragem com  $10^3$  indivíduos simulados) usando o método Bayesiano descrito por Rannala e Mountain (1997) implementado no programa GENECLASS v2.0 (PIRY et al., 2004). As taxas de migração entre as populações foram estimadas ( $2 \times 10^7$  iterações de MCMC e *burn-in* de  $5 \times 10^5$ ) pelo método Bayesiano de Wilson e Rannala (2003) no programa BayesAss v3.04 (RANNALA, 2007). Para as taxas de migração, a convergência da análise foi avaliada no programa Tracer v1.7 (RAMBAUT et al., 2018), a partir da visualização dos dados gerados por MCMC.

### 3.5 *Um estudo de caso com a população de Bubulcus ibis no arquipélago de Fernando de Noronha*

Com foco na população de *Bubulcus ibis* no arquipélago de Fernando de Noronha, cenários foram simulados no programa VORTEX v10.3.3 (LACY, 1993) para prever mudanças no tamanho populacional e na diversidade genética no período entre 2008 e 2020. As simulações se basearam nas características de história de vida da espécie e no conjunto de genótipos determinados em 25 *loci* de microssatélites ( $2 \times 10^3$  iterações e uma geração de indivíduos por ano). O planejamento das simulações considerou dois momentos: o primeiro entre os anos de 2008 e 2011 e o segundo entre os anos de 2011 e 2020. No primeiro período, considerou-se apenas as características de história de vida da espécie para inferir mudanças no tamanho populacional, iniciando no ano em que o último censo populacional foi realizado no arquipélago e finalizando no ano em que a população foi amostrada para a realização do estudo genético. No período entre 2011 e 2020, os dados genéticos foram incorporados na análise e considerou-se a atividade de manejo ecológico realizada no arquipélago até 2015 para estimar, além das oscilações no tamanho populacional, as alterações na diversidade genética. Os genótipos das populações de *Bubulcus ibis* da África do Sul (ZA) e de Rio Claro (SP) foram utilizados como fonte de variação genética cedendo indivíduos durante eventos de dispersão. O programa VORTEX modela a dinâmica populacional baseado em eventos sequenciais discretos que ocorrem de acordo com probabilidades que são variáveis aleatórias, seguindo distribuições especificadas pelo usuário (LACY, 1993). A modelagem descreve ciclos anuais de estocasticidade demográfica a partir da ocorrência de eventos probabilísticos como reprodução, mortalidade, migração entre populações e atividades de manejo com a remoção ou a suplementação de indivíduos (LACY, 1993). Nas simulações com os dados

genético, o modelo usa as frequências alélicas informadas e simula a transmissão de alelos por deriva genética em *loci* hipotéticos, considerando os coeficientes de endogamia da população, as taxas de migração e as características de história de vida informadas (LACY, 1993).

Nos cenários simulados assumiu-se: a) que a variação aleatória nas probabilidades de reprodução e sobrevivência estão correlacionadas (SAMRAOUI; MENAI; SAMRAOUI, 2007); b) que não há dispersão preferencial entre os sexos, que os indivíduos nos primeiros anos de vida têm o dobro da probabilidade de dispersarem e que a dispersão ocorre nas taxas de migração estimadas no programa BayesAss (KUSHLAN; HANCOCK, 2005; TELFAIR, 1994); c) que a monogamia é o sistema de acasalamento predominante e que, aproximadamente, 90% dos indivíduos se reproduzem após atingirem a maturidade sexual (KUSHLAN; HANCOCK, 2005; MIÑO et al., 2017; TELFAIR, 1994); d) que os indivíduos podem viver até os 23 anos e que o tempo de vida reprodutiva se estende de 1 a 17 anos, com cada parental produzindo de 1 a 3 filhotes por geração (KUSHLAN; HANCOCK, 2005; TELFAIR, 1994); e e) as taxas de mortalidade variam entre 37,4 e 55,8% no primeiro ano de vida e entre 20,6 e 27,8% para os adultos (TELFAIR, 1994). Os dados do último censo populacional realizado em 2008 foram utilizados para configurar o tamanho inicial da população ( $n = 298$ ), assumindo uma distribuição etária estável para ambos os sexos (NUNES et al., 2010). Considerando os dados disponíveis do manejo ecológico para o período entre 2009 e 2015 (ADFN, 2012; 2015), incorporou-se nas simulações um manejo com a retirada média anual de 110 indivíduos da população. As simulações foram utilizadas para estimar as taxas de crescimento populacional para cada período, além do tamanho médio e a probabilidade de sobrevivência para cada ano. Para o segundo período simulado, também foram estimadas as heterozigosidades observada e esperada na população a cada geração.

### 3.6 Identificação de SNPs no genoma de *Bubulcus ibis*

#### 3.6.1 Construção de bibliotecas genômicas para Genotyping-by-sequencing (GBS)

O DNA genômico de 207 indivíduos coletados em 11 populações amostradas na área nativa (africana e sul-europeia) e na área colonizada brasileira de distribuição de *Bubulcus*

*ibis* foi utilizado para a construção de seis bibliotecas genômicas de representação reduzida (Tabela 2). A construção das bibliotecas seguiu um protocolo modificado a partir do descrito por Elshire et al. (2011). As amostras foram quantificadas em um fluorímetro Qubit v3 utilizando o kit *dsDNA BR Assay* (Invitrogen™). O DNA genômico de cada indivíduo (100-120 ng) foi digerido em reações de 30 µL contendo 20 U da enzima de restrição PSTI, 1x do tampão NEB 3.1 (100 mM de NaCl, 50 mM de Tris-HCl, 10 mM de MgCl<sub>2</sub>, 10 µg/mL de BSA, pH 7,9 a 25 °C) e água ultrapura. O perfil de ciclagem no termociclador para a digestão enzimática foi: 37° C por 2 h, 80° C por 20 min e resfriamento a 4 °C. As amostras digeridas foram primeiro liofilizadas e depois hidratadas com 4,5 µL de uma solução de adaptadores diluídos a 0,6 ng/µL em tampão TE (10 mM de Tris-HCl e 1M de EDTA, pH 8,0 a 25 °C). As reações de ligação dos adaptadores foram preparadas individualmente contendo a solução da amostra digerida e recém-hidratada, 1x do tampão T4 DNA ligase, 400 U de T4 DNA ligase e água ultrapura para completar o volume de 30 µL. O programa utilizado no termociclador para as reações de ligação dos adaptadores foi: 22 °C por 2h, 65 °C por 30 min e resfriamento a 4 °C. As amostras foram purificadas usando o kit *QIAquick® PCR purification* (Qiagen®) para remover o excesso de adaptadores. Os produtos das purificações foram agrupados em conjuntos de 40 a 53 amostras. Os conjuntos de amostras foram amplificados por PCRs preparadas a 50 µL contendo 1x de tampão, 2 mM de MgCl<sub>2</sub>, 0,75 mM de cada dNTP, 1,25 U de *Taq Platinum* (Invitrogen™), 0,44 µM de cada oligonucleotídeo, 15 µL de cada grupo de amostras e água ultrapura. O programa das PCRs utilizado no termociclador foi: 72 °C por 5 min, 98 °C por 30 s, 18 ciclos (98 °C por 10 s, 65 °C por 30 s e 72 °C por 30 s), extensão final a 72 °C por 5 min e resfriamento a 4 °C. A qualidade do DNA foi avaliada em géis de agarose a 1,5 % submetidos a eletroforese após cada etapa do protocolo. O tamanho dos fragmentos de DNA e a presença de adaptadores sobressalentes foram avaliados pela quantificação das bibliotecas no Agilent 2100 Biolalyzer (Agilent Technologies), usando o kit *Agilent DNA 1000* e por qPCR no Light Cycler 480II (Roche), usando o kit Kapa Biosystems. As bibliotecas foram sequenciadas em três *lanes* na plataforma HiSeq 2500 (Illumina®) do Centro de Genômica Funcional Aplicada a Agropecuária e Agroenergia (Laboratório de Biotecnologia Animal, LZT/ESALQ/USP, Piracicaba/SP).

### 3.6.2 Filtragem, montagem, clusterização das sequências e identificação de SNPs

As sequências brutas foram submetidas ao primeiro controle de qualidade usando a ferramenta FastQC v011.5 disponibilizada na plataforma BaseSpace (Illumina®). Com base no relatório FastQC, estabeleceu-se os critérios de filtragem empregados no *pipeline* SeqyClean v1.10.07 (ZHBANNIKOV et al., 2017) para remoção de sequências menores do que 60 pb e com valores de qualidade abaixo de 20 numa escala *Phred*, assim como sequências de adaptadores, vetores e oligonucleotídeos detectadas após busca na base de dados UniVec (NCBI). O *pipeline* do programa ipyrad (EATON, 2014) foi utilizado para: a) a demultiplexação dos dados e atribuição das sequências (*reads*) às amostras; b) para a filtragem, edição e clusterização dos *reads* dentro das amostras para a formação das sequências consenso; e c) para a clusterização e filtragem das sequências consenso de cada *locus*, nas quais os SNPs foram identificados. No *pipeline* do ipyrad foram filtradas as sequências brutas com mais de quatro bases ambíguas, com valores de qualidade abaixo de 33 numa escala *Phred*. Também para as sequências brutas, foram cortadas as quatro últimas bases do alinhamento final de cada sequência e as sequências residuais de adaptadores, mantendo apenas as sequências alinhadas com um comprimento mínimo de 35 pb após os cortes. Na filtragem das sequências consenso, foram filtradas as sequências com mais de duas bases ambíguas, mais de dois alelos por posição de base, mais de oito bases heterozigóticas e profundidade abaixo de dez nas chamadas estatísticas de bases e abaixo de nove nas chamadas de bases por regra de maioria. Para a clusterização tanto das sequências brutas quanto das sequências consenso, as sequências com mais de 80% de similaridade foram identificadas como homólogas e agrupadas. Na formatação dos arquivos de saída, para as sequências alinhadas de cada *locus*, configurou-se um número mínimo de representatividade em 140 amostras e um número máximo de 20 SNPs, 5 inserções e/ou deleções e 50% de sítios heterozigóticos. O arquivo de saída do ipyrad no formato \*.vcf, reunindo todos os SNPs identificados no alinhamento final das sequências de cada *locus*, foi utilizado para as filtrações subsequentes. O programa PLINK v1.07 (PURCELL et al., 2007) foi utilizado para filtrar SNPs com frequências alélicas menores do que 0,05 e dados faltantes em mais de 3% das amostras, bem como para identificar e excluir SNPs com evidências de desequilíbrio de ligação. Quando um par de SNPs exibiu  $r^2 > 0,5$ , o SNP com menor representação na amostra foi removido. Usando o pacote dartR do programa R (GRUBER et al., 2018; R Development Core Team, 2018) foram excluídas amostras com mais de 15% de dados faltantes em todos os marcadores. As frequências alélicas por população, o percentual global de dados faltantes e as heterozigosidades esperada e observada por SNP foram determinados no pacote hierfstat

(GOUDET, 2005) do programa R (R DEVELOPMENT CORE TEAM, 2018).

## 4 RESULTADOS

### 4.1 *O genoma de referência de Bubulcus ibis*

Foram obtidas 1.457.233 sequências consenso de aproximadamente 30 Gb de dados produzidos durante o sequenciamento das bibliotecas *paired-end*. Após a filtragem por cobertura ( $> 3x$ ) das sequências consenso: 29.422 sequências (9,7 Mb) com cobertura de 3 a 98 x foram mantidas de um total de 134.049 obtidas para a amostra AF189 (SN), 83.336 sequências (27,5 Mb) com cobertura de 3 a 34 x foram mantidas de um total de 242.025 obtidas para a amostra AF326 (ZA), 66.372 sequências (21,9 Mb) com cobertura de 3 a 361 x foram mantidas de um total de 231.848 obtidas para a amostra PA016 (PA) e 398.585 sequências (131,5 Mb) com cobertura de 3 a 15 x foram mantidas de um total de 849.311 obtidas para a amostra RS025 (RS). A melhor montagem, assumida como o genoma de referência de *Bubulcus ibis*, utilizou a totalidade de sequências consenso obtidas a partir dos quatro indivíduos sequenciados: 462.510 sequências (208,1 Mb) com comprimento variando de 140 a 11.920 pb e cobertura de até 13 vezes. No genoma de referência, 32.479 genes homólogos foram anotados e 3.685 sequências contendo 242 *loci* de microssatélites potencialmente amplificáveis foram identificadas.

### 4.2 *Caracterização dos loci de microssatélites prospectados para Bubulcus ibis*

Utilizando os critérios já descritos foram selecionados 22 *loci* de microssatélites e quando testados todos se mostraram polimórficos (Tabela 4). (Tabela 4). No total, 138 alelos diferentes foram detectados nesses *loci* e o número de alelos por *locus* variou entre dois (Bi40) e quatorze (Bi56). As frequências alélicas calculadas mostraram que em seis *loci* (Bi37, Bi40, Bi42, Bi48, Bi51 e Bi59) ocorreu um alelo com frequência maior do que 0,7, tanto na população brasileira quanto na africana (Figura 4).



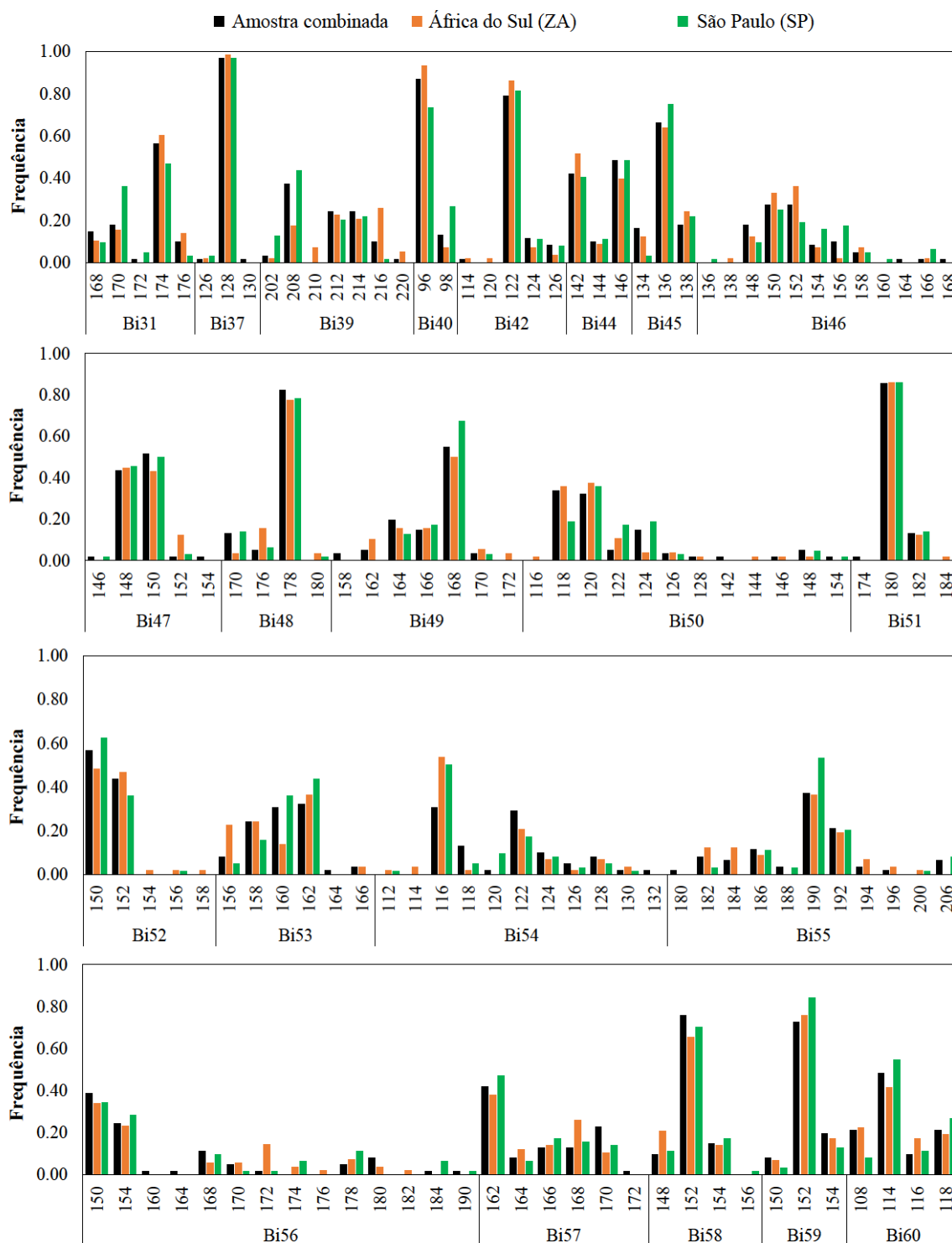
**Tabela 4** - *Loci* de microssatélites polimórficos prospectados. O número de alelos ( $N_a$ ) e os tamanhos dos fragmentos (Frag.) amplificados (menor e maior) são mostrados para cada *locus*.

<i>Locus</i>	$N_a$	Frag. (bp)
Bi31	6	166-176
Bi37	3	126-130
Bi39	7	202-220
Bi40	2	96-98
Bi42	5	114-126
Bi44	3	142-146
Bi45	3	134-138
Bi46	12	136-168
Bi47	5	146-154
Bi48	4	170-180
Bi49	8	158-172
Bi50	12	116-154
Bi51	4	174-184
Bi52	5	150-158
Bi53	6	156-166
Bi54	11	112-132
Bi55	11	180-206
Bi56	14	150-190
Bi57	6	162-172
Bi58	4	148-156
Bi59	3	150-154
Bi60	4	108-118

Fonte: Produzido pelo autor.

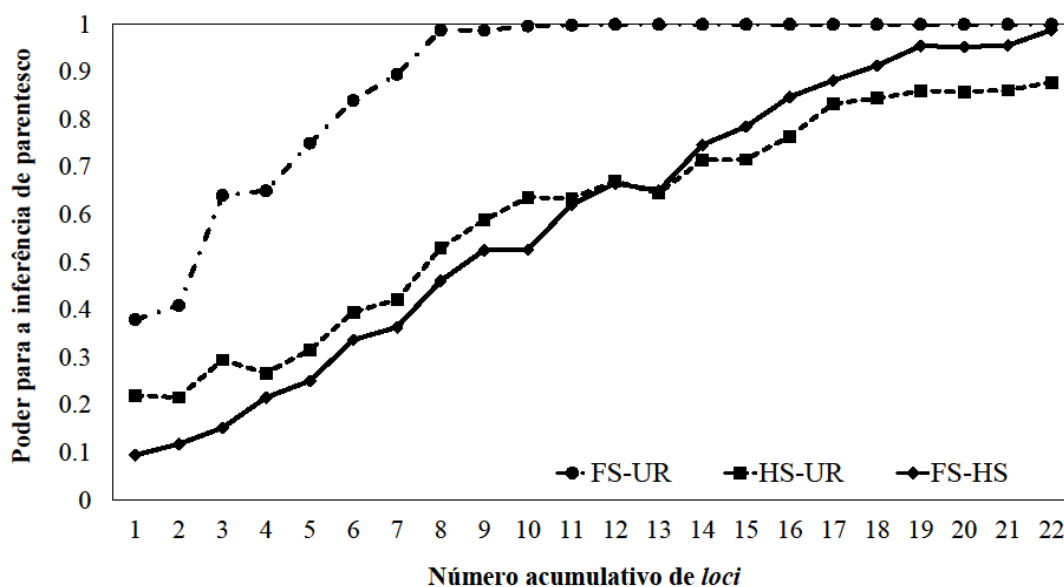
Em relação ao poder *multilocus* (PWR) dos 22 *loci* recém-prospectados na inferência de parentesco genético, obteve-se que com oito desses *loci* é alcançado o poder total de discriminação entre irmãos completos (FS) e indivíduos não aparentados (UR) (Figura 5). Para discriminar meios-irmãos (HS) de irmãos completos, um poder de inferência  $> 0,95$  só foi alcançado empregando-se 19 dos 22 *loci* e o maior poder de inferência para as comparações entre meios-irmãos e indivíduos não relacionados foi de 0,88 usando o conjunto total de 22 *loci* (Figura 5).

**Figura 4** - Frequências alélicas para os 22 *loci* de microssatélites prospectados para a garça-vaqueira (*Bubulcus ibis*). As frequências dos alelos são mostradas para os *loci* caracterizados no grupo de amostras combinadas ( $n = 31$ ) usadas nos testes de polimorfismos e nas populações amostradas na África do Sul ( $n = 29$ ) e em São Paulo ( $n = 32$ ).



Fonte: Produção do próprio autor.

**Figura 5** - Informatividade do novo conjunto de *loci* de microssatélites isolados para a garça-vaqueira (*Bubulcus ibis*) ao inferir relações pareadas de parentesco genético. O poder *multilocus* para a inferência de parentesco genético em função da informação acumulativa dos 22 *loci* de microssatélites ao discriminar entre irmãos completos (FS), meios-irmãos (HS) e indivíduos não relacionados (UR) em simulações baseadas nas frequências alélicas da população amostrada em São Paulo ( $n = 32$ ).



Legenda: Correspondências para os *loci* combinados acima: 1 (Bi31), 2 (Bi37), 3 (Bi39), 4 (Bi40), 5 (Bi42), 6 (Bi44), 7 (Bi45), 8 (Bi46), 9 (Bi47), 10 (Bi48), 11 (Bi49), 12 (Bi50), 13 (Bi51), 14 (Bi52), 15 (Bi53), 16 (Bi54), 17 (Bi55), 18 (Bi56), 19 (Bi57), 20 (Bi58), 21 (Bi59) e 22 (Bi60).

Fonte: Produção do próprio autor.

#### 4.2.1 Aplicação dos *loci* de microssatélites prospectados no estudo genético em populações de *Bubulcus ibis*

Não foram encontradas evidências de desequilíbrio de ligação, alelos nulos, erros de leitura de picos inespecíficos, amplificação preferencial e nem de desvios do equilíbrio de HW ou da neutralidade seletiva ( $\alpha = 0,05$ ). De acordo com o conteúdo informativo polimórfico (PIC), do conjunto usado de *loci* microssatélites, 13 *loci* se mostraram altamente informativos, exibindo valores de  $PIC \geq 0,5$  em pelo menos uma das populações amostradas (Bi31, Bi39, Bi44, Bi46, Bi47, Bi49, Bi50, Bi53, Bi54, Bi55, Bi56, Bi57 e Bi60) (Tabela 5). Os índices de diversidade genética determinados não diferiram significativamente entre as três populações de *Bubulcus ibis* analisadas: as  $H_O$  e  $H_E$  variaram entre 0,52 e 0,55, a  $A_R$  variou entre 4,32 e 4,78 e o  $F_{IS}$  variou entre -0,03 e 0,01 (Tabela 5, valores de  $p < 0,05$  para todas as comparações usando o teste de Wilcoxon ranqueado). No total, 13 alelos privados foram encontrados em 9 *loci* (Bi30, Bi31, Bi37, Bi39, Bi42, Bi46, Bi47, Bi49, Bi50, Bi51, Bi52, Bi53, Bi55, Bi56, Bi57 e Bi58): oito alelos privados na população da África do Sul (ZA), três alelos privados na população de

São Paulo (SP) e dois alelos privados na população de Fernando de Noronha (FN).

Os valores de  $F_{ST}$  global (0,014,  $p < 0,001$ ) e  $F_{ST}$  par a par (ZA-FN = 0,011, ZA-SP = 0,018 e FN-SP = 0,014,  $p < 0,001$ ) foram significativos e evidenciaram uma estruturação genética baixa entre as populações de *Bubulcus ibis* investigadas. Evidências essas que foram reforçadas pela análise de variância molecular (AMOVA) que revelou que 1,43% da variação genética foi explicada pelas diferenças entre as populações ( $F_{ST} = 0,014$ ,  $p = 0,006$ ). A análise complementar de agrupamento Bayesiano espacial no programa TESS v3 mostrou, de forma pouco expressiva, a divisão de oito grupos genéticos supostos como mais prováveis, com um grupo genético mais representativo sendo compartilhado por todas as populações (Figura 6A). Na análise Bayesiana de clusterização realizada no programa STRUCTURE v2.3.4, os valores mais altos de Delta  $K$  foram obtidos quando foram assumidos três agrupamentos genéticos. As três populações de *Bubulcus ibis* investigadas compartilharam todos os grupos genéticos supostos, mas em proporções diferentes (Figura 6B). Cada uma das três populações apresentou um agrupamento genético de maior representatividade, porém a população de Fernando de Noronha apresentou-se com composição intermediária entre as outras duas populações amostradas.

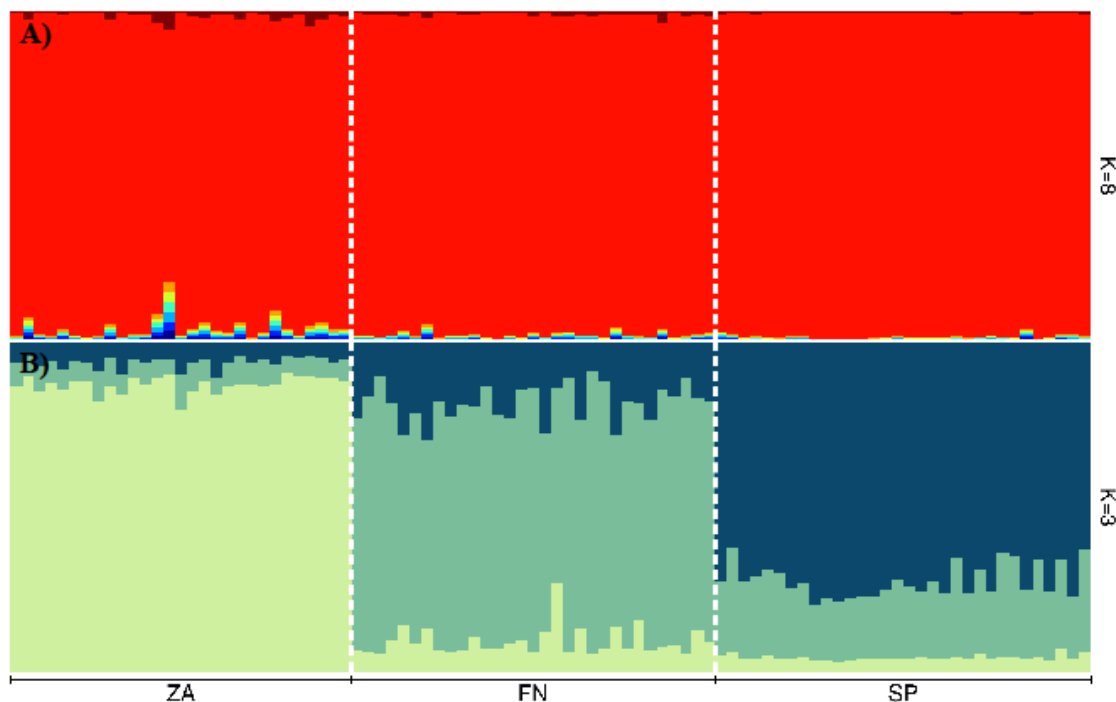
Evidências de migração foram encontradas entre as populações de *Bubulcus ibis* amostradas a partir da informação do conjunto de genótipos com 25 *loci* de microssatélites (Tabela 6). As taxas de migração calculadas com a informação de 25 *loci* evidenciam a população de SP como receptora de indivíduos oriundos das populações de ZA e FN, mas indicaram uma conectividade maior entre a população africana ZA com o arquipélago de FN do que entre as duas populações brasileiras (Tabela 6). Nas análises realizadas no programa GENECLASS v2.0 ( $\alpha = 0,01$ ), todos os indivíduos amostrados na população da ZA foram atribuídos corretamente à própria população. Na população de SP, dois indivíduos foram atribuídos à população de FN, dois indivíduos foram atribuídos à população da ZA e dois indivíduos apresentaram maior probabilidade de pertencerem a qualquer uma das outras duas populações (FN ou ZA). Em FN, cinco indivíduos foram atribuídos à população amostrada na ZA.

**Tabela 5** - Índices de diversidade genética e informatividade determinados em 25 *loci* de microssatélites para populações de garça-vaqueira (*Bubulcus ibis*). O conteúdo de informação polimórfica (PIC), as heterozigosidades observada ( $H_O$ ) e esperada ( $H_E$ ), a riqueza alélica ( $A_R$ ) e o coeficiente de endogamia ( $F_{IS}$ ) foram determinados por *locus* para as populações de *Bubulcus ibis* amostradas na África (ZA,  $n = 29$ ), e no Brasil, em São Paulo (SP,  $n = 32$ ) e em Fernando de Noronha (FN,  $n = 31$ ).

<i>Locus</i>	PIC			$H_O$			$H_E$			$A_R$			$F_{IS}$		
	ZA	BR	FN	ZA	BR	FN	ZA	BR	FN	ZA	BR	FN	ZA	BR	FN
Bi31	0,54	0,57	0,62	0,52	0,81	0,68	0,59	0,65	0,67	4,00	5,00	6,00	0,11	-0,27	-0,03
Bi37	0,03	0,06	0,06	0,03	0,06	0,07	0,03	0,06	0,06	1,97	2,00	2,00	-0,02	-0,03	-0,03
Bi39	0,77	0,66	0,72	0,79	0,78	0,84	0,82	0,72	0,78	6,97	5,00	5,00	0,01	-0,11	-0,10
Bi40	0,12	0,31	0,18	0,14	0,34	0,23	0,13	0,40	0,20	2,00	2,00	2,00	-0,07	0,12	-0,13
Bi42	0,24	0,30	0,28	0,28	0,25	0,32	0,26	0,33	0,31	4,93	3,00	3,00	-0,10	0,22	-0,05
Bi44	0,48	0,50	0,48	0,45	0,56	0,52	0,58	0,60	0,58	3,00	3,00	3,00	0,21	0,04	0,09
Bi45	0,46	0,33	0,42	0,45	0,38	0,42	0,53	0,40	0,50	3,00	3,00	3,00	0,14	0,04	0,14
Bi46	0,70	0,81	0,66	0,79	0,84	0,74	0,75	0,85	0,70	7,90	9,00	7,00	-0,08	-0,01	-0,08
Bi47	0,51	0,44	0,44	0,72	0,47	0,39	0,61	0,55	0,55	3,00	4,00	3,00	-0,21	0,14	0,29
Bi48	0,34	0,34	0,47	0,38	0,44	0,48	0,38	0,37	0,53	4,00	4,00	3,00	-0,02	-0,20	0,07
Bi49	0,65	0,46	0,48	0,79	0,50	0,55	0,70	0,51	0,55	6,00	4,00	5,00	-0,15	0,01	-0,02
Bi50	0,67	0,73	0,76	0,79	0,91	0,65	0,73	0,78	0,81	10,00	7,00	7,00	-0,10	-0,18	0,19
Bi51	0,22	0,21	0,19	0,28	0,22	0,16	0,25	0,25	0,21	2,97	2,00	3,00	-0,14	0,09	0,21
Bi52	0,45	0,38	0,46	0,59	0,56	0,48	0,56	0,49	0,55	4,90	3,00	4,00	-0,07	-0,17	0,11
Bi53	0,70	0,59	0,68	0,76	0,84	0,74	0,75	0,66	0,74	5,00	4,00	4,00	-0,02	-0,29	-0,02
Bi54	0,63	0,67	0,76	0,59	0,72	0,81	0,67	0,71	0,80	8,90	9,00	9,00	0,11	-0,03	-0,02
Bi55	0,77	0,62	0,63	0,86	0,69	0,61	0,80	0,67	0,67	7,96	7,00	8,00	-0,09	-0,05	0,06
Bi56	0,77	0,74	0,68	0,79	0,81	0,65	0,81	0,79	0,73	10,00	9,00	7,00	0,01	-0,05	0,10
Bi57	0,71	0,66	0,70	0,72	0,75	0,77	0,76	0,71	0,75	5,00	5,00	5,00	0,03	-0,07	-0,05
Bi58	0,45	0,42	0,39	0,55	0,47	0,52	0,52	0,47	0,45	3,00	4,00	3,00	-0,08	-0,01	-0,16
Bi59	0,35	0,25	0,34	0,48	0,31	0,45	0,40	0,28	0,38	3,00	3,00	3,00	-0,24	-0,15	-0,21
Bi60	0,66	0,56	0,55	0,66	0,56	0,61	0,73	0,62	0,62	4,00	4,00	4,00	0,08	0,08	-0,01
Bi01	0,42	0,44	0,50	0,62	0,44	0,61	0,53	0,55	0,60	2,99	4,00	4,00	-0,19	0,19	-0,04
Bi20	0,44	0,43	0,49	0,46	0,56	0,58	0,55	0,54	0,57	3,00	3,00	3,00	0,13	-0,05	-0,03
Bi30	0,17	0,06	0,14	0,21	0,06	0,16	0,20	0,06	0,15	2,00	2,00	2,00	-0,12	-0,03	-0,09
Média	0,49	0,46	0,48	0,55	0,53	0,52	0,55	0,52	0,54	4,78	4,40	4,32	-0,03	-0,03	0,01

Fonte: Produção do próprio autor.

**Figura 6** - Evidência de diferenciação genética entre as populações de garça-vaqueira (*Bubulcus ibis*) a partir de dados de microssatélites. Os gráficos mostram as proporções de ancestralidade genética estimadas usando a informação de 25 *loci* de microssatélites nos programas A) STRUCTURE (K = 3) e B) TESS (K = 8) a partir dos dados referentes às populações da África do Sul (ZA,  $n = 29$ ), São Paulo (SP,  $n = 32$ ) e Fernando de Noronha (FN,  $n = 31$ ).



Legenda: A divisão entre as populações é indicada por uma linha branca tracejada e cada barra vertical corresponde às proporções de ancestralidade dos grupos genéticos, caracterizados por cores diferentes, estimados para cada indivíduo amostrado.

Fonte: Produção do próprio autor.

**Tabela 6** - Taxas de migração ( $m$ ) estimadas entre as populações da garça-vaqueira (*Bubulcus ibis*) a partir de dados de microssatélites. Os valores, em frequência de indivíduos por geração, determinados usando o programa BayesAss v3.04 a partir da informação de 25 *loci* de microssatélites são mostrados para as populações oriundas da África do Sul (ZA,  $n = 29$ ), e do Brasil, no arquipélago Fernando de Noronha (FN,  $n = 31$ ) e em São Paulo (SP,  $n = 32$ ).

$m[ZA][ZA]:$	0,6795	$m[ZA][SP]:$	0,2836	$m[ZA][FN]:$	0,0369
$m[SP][ZA]:$	0,0103	$m[SP][SP]:$	0,9654	$m[SP][FN]:$	0,0243
$m[FN][ZA]:$	0,0170	$m[FN][SP]:$	0,2837	$m[FN][FN]:$	0,6993

Fonte: Produção do próprio autor.

#### 4.3 O estudo de caso com a população de *Bubulcus ibis* no arquipélago de Fernando de Noronha

Segundo os dados simulados, a população de *Bubulcus ibis* no arquipélago de Fernando de Noronha reduziu o seu tamanho populacional entre 2008 e 2011 em uma taxa média de crescimento negativo igual a -0,3344. Entre 2011 e 2020, estimou-se um

crescimento da população em uma taxa média de 0,0832. No primeiro período, o tamanho populacional médio foi de 298 indivíduos no final de 2008 para 174 indivíduos no início de 2011, enquanto que a probabilidade de sobrevivência da população oscilou entre 0,99 e 1,00 (Tabela 7). No segundo período, o tamanho populacional médio estimado foi de 245 indivíduos no final de 2011 para 355 indivíduos no início de 2020, apresentando uma tendência negativa entre 2011 até o término das atividades do manejo ecológico em 2015 (Tabela 7). Ainda para o período entre 2011 e 2020, 33,35% das populações simuladas foram extintas no ano de 2015 e recolonizadas devido à dispersão de indivíduos oriundos de outras populações. Nesse período, uma redução significativa nas probabilidades de sobrevivência foi observada entre os anos de 2014 e 2017 (Tabela 7). Apesar de detectada uma tendência negativa nos valores de  $H_E$  e  $H_O$  estimados para a população entre o período de 2011 e 2020, não foi encontrado uma perda significativa de diversidade genética na população (Tabela 7,  $p < 0,05$  no teste de Kolmogorov-Smirnov).

**Tabela 7** - Alterações no tamanho populacional e na diversidade genética estimadas para a população de garça-vaqueira (*Bubulcus ibis*) no arquipélago de Fernando de Noronha entre os anos de 2008 e 2020. Anualmente, o tamanho populacional médio, acompanhado do erro padrão (SE) e do desvio padrão (SD), a probabilidade de sobrevivência da população e os valores das heterozigosidades esperada ( $H_E$ ) e observada ( $H_O$ ) foram estimados a partir de simulações baseadas nas características de história de vida da espécie e na informação de 25 *loci* de microsatélites.

Ano	Tamanho populacional	Sobrevivência	$H_E$	$H_O$
2008	298 (0 SE e 0 SD)	1,00	-	-
2009	234 (1,44 SE e 45,55 SD)	1,00	-	-
2010	174 (2,39 SE e 75,27 SD)	0,99	-	-
2011	245 (0,88 SE e 27,84 SD)	1,00	0,4972	0,4988
2012	224 (1,74 SE e 54,87 SD)	1,00	0,4965	0,4985
2013	194 (2,67 SE e 83,95 SD)	0,99	0,4959	0,4982
2014	177 (3,70 SE e 108,78 SD)	0,86	0,4944	0,4974
2015	189 (5,39 SE e 137,08 SD)	0,65	0,4940	0,4965
2016	239 (8,44 SE e 177,16 SD)	0,44	0,4945	0,4963
2017	303 (12,93 SE e 236,02 SD)	0,33	0,4959	0,4970
2018	159 (8,82 SE e 279,07 SD)	1,00	0,4855	0,4948
2019	240 (12,70 SE e 401,76 SD)	1,00	0,4894	0,4952
2020	355 (18,33 SE e 579,73 SD)	1,00	0,4902	0,4944

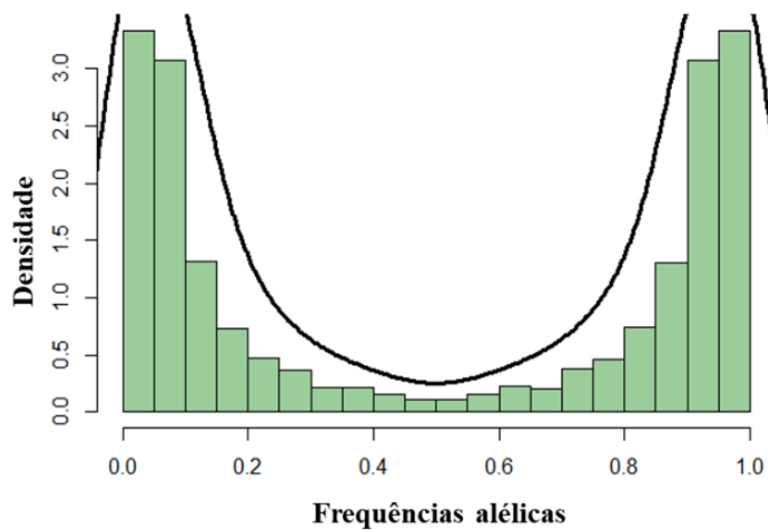
Legenda: Os traços (-) indicam a ausência de informação genética para o período entre 2008 e 2010.  
Fonte: Produção do próprio autor.

#### 4.4 SNPs identificados no genoma de *Bubulcus ibis* por Genotyping-by-sequencing

O sequenciamento das bibliotecas GBS resultou em 613.301.786 sequências brutas com tamanho médio de 101 pb e 54% de bases GC. Em uma escala de qualidade *Phred*, todos os valores individuais por base estiveram acima de 27 e 88,87% das sequências e apresentaram índices acima de 37. O primeiro relatório FastQC mostrou que 65,38% das sequências apresentaram mais de 10 cópias em decorrência da super-representação de *K-mers*. No *pipeline* SeqyClean foram filtradas 16,34% das sequências brutas, restando 513.064.528 sequências com comprimento entre 60 pb e 100 pb. Após a primeira filtragem, o relatório FastQC indicou que todos os valores estiveram acima de 32 em uma escala *Phred*, sendo que 93,45% das sequências apresentaram índices acima de 36. Com a filtragem, a quantidade de sequências com mais de 10 cópias foi reduzida a menos de 4% e a super-representação de *K-mers* foi eliminada. No *pipeline* ipyrad, 26 amostras foram excluídas (8 de PT, 1 de GW, 3 de GH, 3 de NG, 3 de ZA, 3 de PA, 2 de SP e 3 de RS), por estarem sub-representadas nas bibliotecas (menos de 200 sequências brutas recuperadas na demultiplexação) ou na etapa de clusterização das sequências brutas (menos de 70.000 *clusters*). O número total de *clusters* mantidos para cada amostra, após a clusterização das sequências consenso, variou entre 75.923 e 366.738. Na saída do programa ipyrad, de um total de 209.810 sequências consenso alinhadas, 6.043 delas foram mantidas por apresentarem SNPs. Um total de 19.385 SNPs foram identificados nos 181 indivíduos restantes. No programa PLINK, 16.634 SNPs foram filtrados: 16.264 por apresentarem alelos com frequências menores do que 0,05, 176 por apresentarem dados faltantes em mais do que 3% das amostras e 194 por apresentarem probabilidade de desequilíbrio de ligação. Outras sete amostras foram excluídas por apresentarem dados faltantes em mais de 15% dos SNPs mantidos (1 de SN, 1 de GH, 2 de NG, 1 ZA, 1 SP e 1 de RS). Após todas as filtrações, restaram 2.751 SNPs e 174 indivíduos que compuseram o conjunto de dados analisado posteriormente no Capítulo II. Nesse conjunto, foi identificado um percentual de 0,59% de dados faltantes e densidades altas de SNPs com frequências alélicas entre 0,05 e 0,1 (Figura 7). A maioria dos SNPs apresentou índices de heterozigiosidade observada acima de 0,1 (Figura 8).

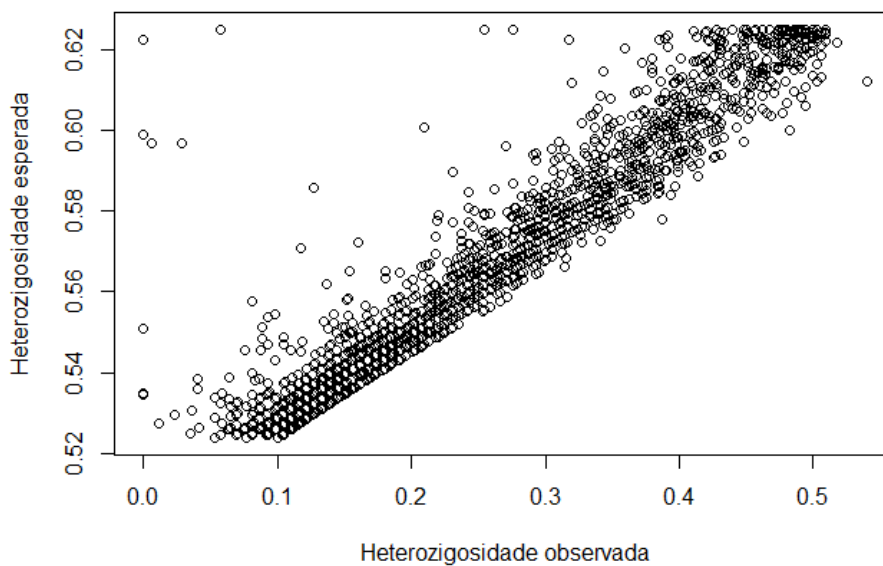


**Figura 7** - Frequências alélicas dos SNPs identificados nas populações amostradas de garça-vaqueira (*Bubulcus ibis*). As frequências alélicas foram determinadas para o conjunto de 2.751 SNPs identificados em 174 indivíduos de 11 populações de *Bubulcus ibis* amostradas ( $n = 174$ ).



Fonte: Produção do próprio autor.

**Figura 8** - Relação entre as heterozigosidades esperada e observada estimadas para cada um dos 2.751 SNPs genotipados nas populações ( $n = 174$ ) amostradas de garça-vaqueira (*Bubulcus ibis*).



Legenda: Cada círculo representa a diversidade genética global determinada para cada SNP como medida de informatividade.

Fonte: Produção do próprio autor.

## 5 DISCUSSÃO

### 5.1 *O genoma de referência de Bubulcus ibis*

Os dados obtidos pelo sequenciamento das bibliotecas *paired-end* permitiram a montagem de um genoma de referência para *Bubulcus ibis* com cobertura suficiente para isolar 242 microssatélites potencialmente amplificáveis e anotar 32.479 genes. Apesar do tamanho do genoma de *Bubulcus ibis* ser desconhecido, a cobertura do genoma de referência em relação ao genoma completo pôde ser estimada, pela comparação com o tamanho dos genomas de espécies relacionadas. Montagens *de novo* dos genomas de 48 espécies de aves, representando 34 das 37 ordens não-extintas do grupo, variaram entre 1,05 Gb e 1,26 Gb em tamanho (ZHANG et al., 2014). Dentre essas espécies de aves, a garça-branca-pequena, *Egretta garzeta* (Linnaeus, 1766), é a única espécie da família Ardeidae cujo o genoma foi sequenciado com uma alta cobertura, apresentando um tamanho de 1,2 Gb (ZHANG et al., 2014). Além de *Bubulcus ibis* e *Egretta garzeta* pertencerem à mesma família, elas também compartilham o mesmo número de macrocromossomos (20 autossomos e dois sexuais) (MOHANTY; BHUNYA, 1990). Assumindo que o genoma de *Bubulcus ibis* e o de *Egretta garzeta* tenham tamanhos similares, pode se afirmar que o genoma de referência de *Bubulcus ibis* (com ~208,1 Mb) representaria, aproximadamente, 17,34% do genoma completo (com 1,2 Gb). Esse genoma de referência, mesmo com baixa cobertura, permitiu o isolamento de microssatélites e poderá ser usado futuramente para outros propósitos, como no estudo de elementos transponíveis e de outros segmentos de DNA altamente repetitivos (CASTOE et al., 2012).

### 5.2 *Caracterização dos loci de microssatélites prospectados para Bubulcus ibis*

O número de sequências identificadas para *Bubulcus ibis* com microssatélites de dinucleotídeos e potencialmente amplificáveis (3.685 sequências) foi também reportado para outras seis espécies de aves (de 696 a 3.804 sequências), utilizando uma metodologia semelhante à descrita neste trabalho (LANCE et al., 2013). Após o trabalho publicado que comprovou a utilidade de dados de sequenciamento *paired-end* no isolamento de microssatélites (CASTOE et al., 2012), Lance et al. (2013) validou a metodologia em 32 espécies animais e vegetais, incluindo 17 espécies do filo Chordata. O número de

sequências com microssatélites de dinucleotídeos obtidas para *Bubulcus ibis* foi inferior ao reportado por Lance et al. (2013) para os representantes das classes Reptilia, Amphibia e Mammalia. Um resultado esperado, uma vez que o genoma das aves apresenta tamanho e complexidade menores, em comparação aos outros tetrápodes (ELLEGREN, 2005; KAPUSTA; SUH; FESCHOTTE, 2017).

Com a adição do novo conjunto de 22 *loci* de microssatélites descritos para *Bubulcus ibis*, temos um total de 34 *loci* de microssatélites polimórficos e espécie-específicos, disponíveis para a espécie (CAMPANINI et al., 2012; MIÑO et al., 2017). A taxa de sucesso de amplificação e os níveis detectados de polimorfismos demonstraram que a utilização de dados NGS e o estabelecimento de critérios restritivos para o desenho de primers conduziram a uma maior eficiência no isolamento de *loci* de microssatélites. Originalmente, 11 *loci* de microssatélites foram prospectados por Campanini et al. (2012), mas cinco destes, apresentaram polimorfismos muito baixos e, posteriormente mostraram-se monomórficos em muitas populações. Além da eficiência baixa, a primeira prospecção foi realizada por um método tradicional que demandou um tempo e um esforço maiores de desenvolvimento (CAMPANINI et al., 2012). Para isolar os 11 *loci* com 28 alelos, Campanini et al. (2012) desenharam 32 pares de oligonucleotídeos a partir das sequências obtidas com 432 clones recombinantes. Além dos 34 *loci* espécie-específicos atualmente disponíveis, um *locus* heterólogo desenvolvido para a garça-azul-grande, *Ardea herodias* (Linnaeus, 1758), mostrou-se polimórfico em *Bubulcus ibis* e foi empregado em estudos anteriores (MIÑO et al., 2017; MORALEZ-SILVA; DEL LAMA, 2017).

Os valores médios de PIC (Tabela 5) demonstraram que o conjunto de 22 *loci* aqui descritos foi mais informativo em comparação com o conjunto de 11 *loci* reportados por Campanini et al. (2012) (valor médio de PIC =  $0,27 \pm 0,14$ ) e com o conjunto de seis *loci* relatados por Miño et al. (2017) (valor médio de PIC =  $0,56 \pm 0,06$ ). Os índices de diversidade genética aqui determinados foram significativamente maiores do que os calculados usando os valores reportados por Campanini et al. (2012) (média de  $H_O = 0,30 \pm 0,18$  e média de  $H_E = 0,33 \pm 0,18$ , valores de  $p < 0,05$  para todas as comparações usando o teste de Wilcoxon ranqueado). Em contraste, os índices de diversidade genética aqui determinados não diferiram significativamente dos reportados por Miño et al. (2017) (média de  $H_O = 0,53 \pm 0,07$  e média de  $H_E = 0,51 \pm 0,06$ ,  $p$ -valores  $> 0,05$  para todas as comparações usando o teste de Wilcoxon ranqueado). Os valores de heterozigosidade esperada determinados neste estudo convergiram com os níveis de diversidade genética

compilados para populações de aves aquáticas, usando informações de *loci* de microssatélites (média de  $H_E$  por *locus* por população =  $0,56 \pm 0,21$ ) (EO; DOYLE; DEWOODY, 2011). Com base no número de alelos por *locus*, nas frequências alélicas e nas estimativas de riqueza alélica (Figura 4; Tabelas 4 e 5), pode-se afirmar que os índices de diversidade genética foram estimados com precisão a partir do novo conjunto de *loci* de microssatélites. Essa suposição se faz porque não encontramos uma baixa riqueza alélica no conjunto de *loci* analisados, o que poderia afetar a precisão na estimativa de parâmetros genéticos populacionais, levando a erros significativos na avaliação da diversidade genética de populações alvo (BASHALKHANOV; PANDEY; RAJORA, 2009). Os valores calculados de poder *multilocus* para a inferência de parentesco (PWR, Figura 5) também foram maiores do que os apresentados com um conjunto de 11 *loci* de microssatélites em uma investigação anterior em uma população de *Bubulcus ibis* amostrada em São Paulo (MIÑO et al., 2017). No trabalho anterior, Miño et al. (2017) obtiveram valores de PWR entre 0,45 e 0,50 para as comparações entre irmãos completos e meio irmãos e entre 0,40 e 0,45 nas comparações entre irmãos e indivíduos não relacionados.

O novo conjunto de *loci* de microssatélites tem o potencial de ser usado em estudos genéticos com outras espécies de aves. Os *loci* de microssatélites previamente descritos para *Bubulcus ibis* demonstraram uma alta taxa de sucesso de amplificação cruzada (entre 81,8% e 100%) quando testados entre outras oito espécies de ardeídeos (CAMPANINI et al., 2012). Para a classe das Aves, a transferibilidade de microssatélites entre espécies não-modelos e filogeneticamente relacionadas é alta devido às contrições filogenéticas do grupo e a taxa alta de sintenia (BURT et al., 1999; DAWSON et al., 2010).

### 5.3 *Aplicação dos loci de microssatélites prospectados no estudo genético em populações de Bubulcus ibis*

Os *loci* de microssatélites descritos também podem contribuir para a caracterização genética de populações de *Bubulcus ibis* em áreas colonizadas e auxiliar no desenvolvimento e na avaliação de estratégias de manejo da espécie, particularmente em ilhas oceânicas, onde seu impacto negativo sobre populações de espécies nativas tem sido documentado (BARBOSA-FILHO et al., 2009; SILVA-JR; PERES-JR; SAZIMA, 2005). Neste estudo, avaliou-se a aplicação dos 22 *loci* prospectados, juntamente com três

*loci* descritos por Campanini et al. (2012) e constatou-se que o conjunto de dados obtido permitiu determinar a estrutura genética e verificar se há conectividade entre as populações de *Bubulcus ibis* amostradas em São Paulo, Fernando de Noronha e África do Sul (Figura 6 e Tabela 6). Os resultados de  $F_{ST}$  par-a-par, AMOVA e as análises de clusterização genética, obtidos a partir desse conjunto de genótipos, convergiram para a existência de estruturação genética entre as populações amostradas (Figura 6). Nas análises Bayesianas de clusterização genética, os resultados divergiram no quanto as populações estavam estruturadas e pode ser interpretado como consequência das informações diferentes dadas *a priori* em cada análise: as populações de amostragem no programa STRUCTURE e a exata localização geográfica dessas populações no programa TESS (Figura 6). O modelo do programa STRUCTURE com informações prévias da população é recomendado para a identificação de migrantes entre as populações amostradas quando há fortes evidências de estruturação populacional (HUBISZ et al., 2009). Como indicado pela estatística  $F$ , a AMOVA e a análise espacial Bayesiana do programa TESS (Figura 6A), o cenário mais provável é o de estruturação genética baixa entre essas populações amostradas.

Os resultados obtidos nas análises de alelos privados, de atribuição populacional e as estimativas das taxas de migração estiveram de acordo com o esperado pelos padrões históricos de distribuição geográfica da espécie (AHMED, 2011; MORALES-SILVA; NASSIF DEL LAMA, 2014). O número maior de alelos privados detectados na população sul-africana (oito alelos) do que nas populações brasileiras (três em SP e dois em FN) está dentro do esperado, já que as Américas foram ocupadas por *Bubulcus ibis* durante um processo de colonização mais recente, possivelmente com poucos eventos de migração (ARENDDT, 2006). Esses resultados também estão em concordância com o fato de nenhum indivíduo amostrado na população sul-africana ter sido atribuído a uma população brasileira. As taxas de migração estimadas indicaram que um número menor de indivíduos migra a partir da população de SP (4% a cada geração) em comparação com as populações da ZA (32% a cada geração) e FN (30% a cada geração) (Tabela 6). Esses resultados foram confirmados pelas evidências da análise de atribuição populacional, em que SP recebe indivíduos oriundos tanto de FN quanto da ZA e de que FN recebe indivíduos apenas da ZA.

#### 5.4 O estudo de caso com a população de *Bubulcus ibis* no arquipélago de Fernando de Noronha

Os dados simulados para a população de *Bubulcus ibis* do arquipélago de Fernando de Noronha indicaram oscilações no tamanho populacional e na probabilidade de sobrevivência da população durante o período entre 2008 e 2020 (Tabela 7). Esses dados indicaram uma tendência negativa no tamanho populacional médio entre 2008 e 2015, quando foi realizado o manejo ecológico da espécie no arquipélago (Tabela 7). Após 2015, previu-se o crescimento populacional da espécie na região, embora a probabilidade de sobrevivência tenha permanecido baixa até 2017 (Tabela 7). Em ambos os períodos, as mudanças no tamanho populacional não foram lineares, como reportado para o período entre 1986 e 2008 (NUNES et al., 2010). Esses resultados indicam que o manejo ecológico da espécie, pelo menos durante o período em que foi executado, cumpriu o seu propósito de controlar o tamanho populacional de *Bubulcus ibis* no arquipélago. Um cenário similar ao inferido pelas simulações foi observado durante o manejo ecológico de *Bubulcus ibis* realizado no Havaí (EUA) entre 1982 e 1983 (FELLOWS; PANTON, 1988). No Havaí, uma população estimada em 547 indivíduos foi manejada com a destruição de ovos e eventos de captura e eutanásia de centenas de adultos e filhotes (FELLOWS; PANTON, 1988). Entretanto, assim que as atividades de manejo foram encerradas, a população restabeleceu o seu tamanho original em apenas uma única temporada reprodutiva (FELLOWS; PANTON, 1988). Na ocasião, o crescimento exponencial da população pode ter ocorrido pelo aumento da taxa de sucesso reprodutivo ou por um comportamento de reprodução compensatório para restaurar a perda de ovos e filhotes rapidamente (FELLOWS; PANTON, 1988; KENNEDY, 1991).

A extinção em 2015 de 33,35% das populações simuladas, provavelmente, motivou a manutenção de probabilidades baixas de sobrevivência nos anos seguintes após o término das atividades de manejo (Tabela 7). Essa probabilidade baixa deve estar associada ao insucesso na recolonização que é totalmente dependente da dispersão de indivíduos a partir de outras populações. A chegada de indivíduos em Fernando de Noronha seria mais provável a partir da dispersão de indivíduos provenientes da costa brasileira ou do continente africano (ARENDDT, 1988; ELLISON, 1978). Para garantir o controle do tamanho populacional no arquipélago populacional, o manejo ecológico deveria ocorrer de forma ininterrupta. A manutenção dos níveis diversidade genética na população, apesar das oscilações no tamanho populacional (Tabela 7) indica uma chegada regular de migrantes no arquipélago.

Sem manejo o cenário futuro mais provável é o da população crescendo exponencialmente em tamanho até atingir a sua capacidade máxima (NUNES et al., 2010). Entretanto, é difícil prever qual seria essa capacidade uma vez que é notável a flexibilidade ecológica da espécie que frequentemente explora novos nichos (ARENDRT, 1988; KREBS; RIVEN-RAMSEY; HUNTE, 1994). Como exemplo, *Bubulcus ibis* colonizou a ilha de Barbados nas Antilhas em 1956 e em apenas 34 anos o tamanho estimado da população ultrapassou 11.000 indivíduos com uma densidade populacional de 25 indivíduos/km<sup>2</sup> (KREBS; RIVEN-RAMSEY; HUNTE, 1994). Apesar da densidade populacional alta e da tendência de crescimento populacional, não foram encontradas evidências de limitações em termos de disponibilidade alimentar e espacial para o estabelecimento de novas colônias (KREBS; RIVEN-RAMSEY; HUNTE, 1994). A colonização eficiente e o crescimento exponencial das populações de *Bubulcus ibis* em áreas recém-colonizadas têm sido atribuídos a algumas características como a dieta generalista, estações reprodutivas longas e a ausência de predadores, especialmente nas regiões insulares (ARENDRT, 1988; BLACKBURN et al., 2001).

#### 5.5 SNPs identificados no genoma de *Bubulcus ibis* por *Genotyping-by-sequencing*

Na abordagem de GBS, ao se escolher uma determinada enzima de restrição se determina qual a extensão da redução de complexidade do genoma a ser estudado (ELSHIRE et al., 2011). Como este é o primeiro trabalho a aplicar a técnica de *Genotyping-by-sequencing* (GBS) em uma espécie da família Ardeidae, até onde se tem conhecimento, baseou-se a escolha da enzima de restrição em estudos com espécies de aves pertencentes a outros grupos (LIM; BRAUN, 2016; RANDI et al., 2015). A expectativa de que o uso da enzima Pst1 assegurasse uma profundidade de leitura do genoma suficiente para marcar genótipos heterozigóticos foi confirmada, resultando em 6.043 *loci* contendo 19.385 SNPs. Na análise GBS, cada *locus* pode conter alguns marcadores potenciais, porém o número de *locus* que pode ser convertido com sucesso em marcadores depende do tamanho da população, da diversidade genética e do limite de frequências alélicas menores (MAF) usado nas chamadas de SNPs (BYRNE et al., 2013). Assumindo que as amostras foram suficientemente representativas da diversidade genética de cada população, o limite de MAF estipulado foi o fator principal afetando as chamadas de SNPs. Trabalhos anteriores, comparando diferentes critérios de filtragem de

SNPs, recomendaram que o limite de MAF em 0,05 e de heterozigosidade em 0,75 eram suficientes para garantir a confiança nos dados de genotipagem e eliminar o pareamento incorreto de sequências parálogas (CLARK, 2018; WOLD, 2017). Neste trabalho, os limites de MAF em 0,05 e de no máximo oito bases heterozigóticas por sequência consenso removeram SNPs menos informativos, mantendo SNPs com a heterozigosidade observada, geralmente, acima de 0,1 e as frequências alélicas em um padrão de distribuição bimodal (Figuras 7 e 8). A distribuição bimodal das frequências alélicas é o padrão esperado após a filtragem de SNPs com níveis significativos de distorção de segregação mendeliana (RABBI et al., 2014).

O número de SNPs mantidos no conjunto de dados após as filtrações (2.751 SNPs) apresentou-se dentro do intervalo (de 703 a 4051 SNPs) reportado para outras espécies de aves, utilizando bibliotecas de representação reduzida do genoma e critérios semelhantes de filtrações (LIM; BRAUN, 2016; RANDI et al., 2015).

## 6 CONCLUSÕES

Os 22 novos *loci* de microssatélites e o painel de 2.751 SNPs descritos para a garça-vaqueira (*Bubulcus ibis*) neste estudo são poderosas ferramentas moleculares que podem auxiliar futuras investigações sobre a estrutura genética de populações da espécie e aumentar a resolução de inferências de parentesco. Além disso, o novo conjunto de *loci* de microssatélites pode ser usado para futuros testes de transferibilidade em outras espécies de aves filogeneticamente relacionadas. Os critérios estabelecidos para a seleção de microssatélites e o desenho de oligonucleotídeos garantiram que todos os *loci* de microssatélites selecionados eram polimórficos e podem ser aplicados na prospecção de *loci* espécie-específicos em outros táxons. Aqui, provou-se a utilidade dos dois conjuntos de marcadores, microssatélites e SNPs, para avaliações de diversidade genética. O poder *multilocus* estimado para os *loci* de microssatélites prospectados é um indicativo de que esse novo conjunto de *loci* podem dar uma maior resolução a inferência em estudos genéticos familiares, melhorando a discriminação genética entre diferentes relações de parentesco. Os 22 novos *loci* de microssatélites e o painel de 2.751 SNPs descritos podem contribuir para a caracterização genética de populações de garças em áreas colonizadas e auxiliar na tomada de decisão para o manejo dessa espécie invasora, particularmente em



ilhas oceânicas, onde seu impacto negativo sobre populações de espécies nativas tem sido documentado. No estudo de caso com a população de *Bubulcus ibis* localizada no arquipélago de Fernando de Noronha, estimou-se uma redução no tamanho populacional sem uma perda significativa de diversidade genética durante os anos em atividades de manejo ecológico da espécie foram realizadas na região. Esses dados corroboraram a eficácia do programa implementado para o controle do tamanho populacional, mas indicam que indivíduos provenientes de outras populações podem dispersar regularmente para o arquipélago.

### III - CAPÍTULO II

UM SÉCULO APÓS A COLONIZAÇÃO DO BRASIL PELA GARÇA-VAQUEIRA  
(*BUBULCUS IBIS*): DISTRIBUIÇÃO DA VARIAÇÃO GENÉTICA E  
CONECTIVIDADE ENTRE AS POPULAÇÕES NATURAIS DA ESPÉCIE

## 1 INTRODUÇÃO

### 1.1 *O processo de colonização por espécies com dispersão de longa distância*

A distribuição geográfica, a quantidade e a distribuição da variação genética de uma espécie são resultantes dos processos populacionais como a migração, a colonização de áreas novas e a resposta das populações às condições ambientais (ELITH; LEATHWICK, 2009; TIEDEMANN et al., 2004). Entender como a dinâmica populacional determina tanto a distribuição geográfica de uma espécie quanto a estrutura genética de suas populações são questões centrais na Ecologia, Evolução, Biogeografia e Biologia da Conservação (ELITH; LEATHWICK, 2009; MANEL et al., 2003). Entre os processos populacionais, a colonização de novas áreas pode ser determinante na sobrevivência de populações, para a definição da estrutura genética e ecológica das comunidades e na distribuição de espécies (BURGESS; MARSHALL, 2011; HEDGE; O'CONNOR; JOHNSTON, 2012; JONES et al., 2002; KOKKO; LÓPEZ-SEPULCRE, 2007). E o seu entendimento é um tópico de interesse para prever a resposta de uma espécie frente às mudanças ambientais, assim como para projetar práticas de conservação e manejar espécies invasoras (HIGGINS; RICHARDSON; COWLING, 1996; MCGEOCH et al., 2015; PIERTNEY et al., 2016).

Estudos de colonização por espécies com dispersão de longa distância têm procurado determinar a pressão de propágulo e identificar as rotas de invasão (GILLESPIE et al., 2012; MCGEOCH et al., 2015; YANG et al., 2012). A pressão de propágulo inclui a abundância de indivíduos (tamanho do propágulo) e a frequência com que esses indivíduos chegam a uma área a ser colonizada (número de propágulos) (CARLTON et al., 2002; LOCKWOOD; CASSEY; BLACKBURN, 2005). O sucesso de uma colonização depende fundamentalmente de fatores como a sobrevivência dos indivíduos durante a migração e o estabelecimento de uma população reprodutora sustentável na área explorada (ELLIS-EVANS; WALTON, 1990). Indivíduos estabelecidos em uma área recém-colonizada podem produzir novas pressões de propágulo que se dispersam com o desafio de encontrarem novas áreas para também se estabelecerem (DAVIS, 2009). Esse processo, denominado persistência ou naturalização, garante a expansão do tamanho e da distribuição geográfica dessas populações (HEDGE; O'CONNOR; JOHNSTON, 2012; PYŠEK; RICHARDSON, 2006).

Marcadores genéticos têm sido amplamente utilizados para investigar padrões de movimento histórico em espécies invasoras e para examinar o papel da variação genética no sucesso de um evento de colonização (BLANCHET, 2012; LAWSON HANDLEY et al., 2011). Esses marcadores podem indicar o status das espécies invasoras, a origem das populações invasoras e os níveis de variação genética nas populações nas áreas colonizadas quando comparadas às populações de origem (CHOWN et al., 2015). Embora os marcadores genéticos sejam usados quase que rotineiramente para identificar populações-fonte de espécies, marcadores genômicos prometem um nível maior de resolução e estão sendo cada vez mais aplicados (CHOWN et al., 2015). O uso de marcadores genéticos e genômicos apropriados em uma abordagem conjunta filogeográfica pode dar indícios de um evento mais antigo de colonização, a partir da estrutura genética atual das populações (AVISE, 2000; TIEDEMANN et al., 2004). Por exemplo, a ausência de estruturação genética pode indicar que a população na área colonizada está continuamente conectada a população de origem por fluxo gênico (GILLESPIE et al., 2012). A diversidade genética da população fundadora também pode ser uma medida indireta da pressão de propágulo (LOCKWOOD; CASSEY; BLACKBURN, 2005, 2009). De forma que a comparação das diversidades genéticas determinadas nas populações das áreas nativas e colonizadas pode indicar a intensidade de fluxo gênico durante o processo de colonização (WILSON et al., 2009).

A variação genética neutra encontrada nas populações é vista como o produto da mutação, da migração e da deriva genética (HARTL; CLARK, 2007; WRIGHT, 1943). Como a migração é mais comum entre populações vizinhas, uma correlação entre as distâncias geográfica e genética é esperada em muitas populações naturais, o que caracteriza o padrão de isolamento por distância (WRIGHT, 1943). Presume-se que a maioria das diferenças fenotípicas entre essas populações naturais surgiu por regimes de seleção diferencial, e a deriva genética, nesse caso, é considerada uma hipótese nula contra a qual a seleção pode ser testada (ORR, 1998). Alguns dos modelos de colonização, baseados nesse padrão de isolamento por distância, são o modelo *stepping stone* simples e o hierárquico (KIMURA; WEISS, 1964). Nesses modelos, supõe-se que a existência de barreiras geográficas ou de uma distribuição geográfica muito grande em comparação com o alcance de migração dos indivíduos, impede a formação de uma única unidade populacional panmítica (WRIGHT, 1943). Para espécies distribuídas de forma descontínua, como aves de reprodução colonial, é esperado um padrão de colonização conforme o modelo de migração *stepping-stone* (KIMURA; WEISS, 1964). O modelo

*stepping stone* presume que em populações naturais, os indivíduos frequentemente estão distribuídos mais ou menos descontinuamente para formar numerosas colônias, e prevê ainda que indivíduos migrem entre colônias mais próximas (KIMURA; WEISS, 1964).

Nos casos de populações distribuídas ao longo de gradientes ambientais, o provável é que os indivíduos estejam adaptados às condições locais (SPURGIN et al., 2014). Quando a adaptação local atua contra a migração, é produzido um padrão de estruturação genética neutra, conhecido como isolamento por adaptação, resultante da deriva genética subsequente (ORSINI et al., 2013; SPURGIN et al., 2014). Durante um processo de colonização, o efeito fundador também pode ser o principal responsável pela diferenciação genética e fenotípica entre as populações das áreas nativa e colonizada, levando a um padrão de isolamento por colonização (SPURGIN et al., 2014). Nesse último modelo, as diferenças entre as populações refletem principalmente a história da colonização, ao invés de padrões contemporâneos de migração, deriva genética ou seleção (SPURGIN et al., 2014). Nesse modelo, portanto, não é esperada nenhuma relação entre a distância genética e as distâncias geográfica ou ecológica, em vez disso, espera-se que a estrutura genética reflita a história de colonização e o gargalo das populações estudadas (ORSINI et al., 2013).

Estudos de colonização integram os esforços empreendidos para monitorar e manejar os ecossistemas em todo o mundo (VAN DEN BOSCH; HENGEVELD; METZ, 2006). A dispersão de longa distância (LDD) tem se destacado nesses estudos como um fator fundamental no processo de colonização, influenciando tanto a dinâmica de expansão populacional quanto a distribuição espacial da diversidade genética (BOHRER; NATHAN; VOLIS, 2005; FAYARD; LEFÈVRE; KLEIN, 2009). Do ponto de vista da genética populacional, a LDD pode tanto determinar o tamanho de um efeito fundador quanto promover o fluxo gênico após a colonização (KLEIN; LAVIGNE; GOUYON, 2006; LAMBRINOS, 2004). O efeito da LDD na estrutura genética não é completamente compreendido devido à complexidade introduzida pelo balanço entre o efeito fundador e a manutenção de fluxo gênico que depende da frequência e não apenas da existência de eventos LDD (FAYARD; LEFÈVRE; KLEIN, 2009). Espécies de aves potencialmente invasoras, tais como a garça-vaqueira (*Bubulcus ibis*, Linnaeus, 1758; Pelecaniformes: Ardeidae), tornaram-se relevantes nesses estudos como modelos biológicos para documentar padrões espaciais e temporais de colonização, justamente devido à capacidade de dispersão em longas distâncias (SI BACHIR et al., 2011; VAN DEN BOSCH; HENGEVELD; METZ, 2006). Além das próprias aves que dispersam em

longas distâncias há de se considerar que esses organismos levam consigo outros organismos como invertebrados parasitas e plantas e promovem sua dispersão (GILLESPIE et al., 2012; GREEN; FIGUEROLA, 2005). A determinação de rotas de invasão e da capacidade dos ambientes para o estabelecimento de novas populações invasoras são essenciais no planejamento prévio de ações de manejo, pois permitem a identificação de áreas vulneráveis à colonização por espécies exóticas.

### *1.2 Evidências históricas de migração e da expansão da distribuição geográfica de *Bubulcus ibis**

Observações reais de eventos de dispersão de longa distância são limitadas, apesar dos avanços tecnológicos que tornaram a marcação de indivíduos mais factível (SHAFFER et al., 2006). Registros oficiais de migração de *Bubulcus ibis* são escassos e dependem da recuperação de anilhas de aves mortas e da captura ou observação de aves com algum tipo de identificação. Na América do Norte, por exemplo, 61.162 indivíduos de *Bubulcus ibis* foram anilhados entre as décadas de 1950 e 1990, e desse total somente 577 (0,94%) foram recuperados ou avistados (ARENDDT, 1988). Apesar da taxa baixa de recuperação, Arendt (1988) identificou três rotas de migração partindo da América do Norte durante o período de invernada em direção à América Central e o extremo norte da América do Sul. Uma delas partindo de Nova Escócia, no Canadá, e chegando até a Colômbia na América do Sul, após passar pela costa atlântica dos EUA, pelo México e pela América Central (ARENDDT, 1988). Quando a distribuição inclui muitos países e dois continentes a recuperação de indivíduos anilhados é rara (VALVERDE, 2003).

No continente africano e no Sul da Europa, provavelmente, o número de indivíduos produzidos a cada temporada reprodutiva está além da capacidade suporte do ambiente, incentivando a dispersão anual de indivíduos (SIEGFRIED, 1971). Depois de 1900, as populações de *Bubulcus ibis* expandiram em tamanho e distribuição geográfica dentro da sua área nativa, sendo que em algumas regiões essa expansão coincidiu com a intensificação de atividades agropecuárias (KUSHLAN; HANCOCK, 2005). Em resposta às mudanças de temperatura e padrões de chuva, algumas populações africanas de *Bubulcus ibis* migram em busca de condições ambientais mais favoráveis (KUSHLAN; HANCOCK, 2005). Evitando temperaturas mais baixas durante o inverno,

populações ao sul do continente migram em direção à África Central (KUSHLAN; HANCOCK, 2005). Populações altamente nômades da África Ocidental (região centro-oeste africana), migram sentido ao sul ao longo da costa atlântica evitando os períodos de seca para se reproduzir (HANCOCK; KUSHLAN, 1984). O comportamento nômade das populações do centro-oeste africano e os registros de aves em ilhas do Atlântico Sul dão suporte à hipótese de que seja dessa região a origem das aves que colonizaram o continente americano (TELFAIR, 1994). No Sul da Europa, as aves que ocupavam principalmente a região de Andaluzia na Espanha, se estabeleceram na década de 1950 na França e no entorno do mar Cáspio (TELFAIR, 1994; VALVERDE; WEICKERT, 1956). Em 1956, 2.000 indivíduos de *Bubulcus ibis* foram anilhados no Parque Nacional de Doñanas na Espanha e 28 indivíduos tiveram as anilhas recuperadas na metade inferior da Península Ibérica, desde a foz do rio Tajo até o município de Elche na província espanhola de Alicante (VALVERDE, 2003). Os dados de recuperação de anilhas, indicaram que durante os períodos secos no norte da África e no sul da Espanha, pela menos uma parte da população de *Bubulcus ibis* migrou sentido ao norte e não ao sul, como previsto anteriormente (VALVERDE, 2003). Em Portugal, o primeiro registro de uma colônia reprodutiva de *Bubulcus ibis* foi em 1962, embora haja relatos de que a reprodução da espécie tenha se iniciado na região em 1930 (MARION; MARION, 1982; VOISIN, 1991).

Nas Américas, supõe-se que essas aves tenham chegado à costa norte da América do Sul por muitos anos antes de finalmente se estabelecerem (TELFAIR, 1994). Os primeiros registros da espécie são de 1877-1882 no Suriname e em 1911-1912 na Guiana, onde populações reprodutivas se estabeleceram em 1950 (TELFAIR, 1983). Apesar dos avistamentos anteriores, a primeira captura de um indivíduo da espécie nas Américas ocorreu em 1937 na Guiana (BLAKE, 1939). Os registros seguintes ocorreram na Venezuela em 1943, na Colômbia em 1951, na Bolívia em 1953, no Peru em 1956 e no arquipélago de Galápagos em 1964 (LEVEQUE et al., 1966; POST, 1970; SPRUNT, 1955). Na América Central, os primeiros registros de *Bubulcus ibis* se deram em Aruba em 1944 e em Cuba em 1954 (DRURY, 1953; SPRUNT, 1955). Os primeiros registros de *Bubulcus ibis* no Brasil foram em 1964 na Ilha de Marajó (Pará) e em 1971 em Brasília, porém a reprodução da espécie no país só foi confirmada em 1973 com o registro da primeira colônia reprodutiva no Rio Grande do Sul (SICK, 1965; BELTON, 1974).

A hipótese mais discutida para a colonização das Américas prevê que as aves que deram origem as populações do continente americano são originárias da costa centro-

oeste da África (MASSA; DOYLE; FORTUNATO, 2014). Evidências para apoiar a travessia transatlântica são fornecidas por registros da espécie no arquipélago de São Pedro e São Paulo, localizado entre a África e a América do Sul (BOWEN; NICHOLLS, 1968). Os avistamentos de *Bubulcus ibis* cruzando o oceano Atlântico também dão suporte à hipótese de que a rota de migração mais provável é a de menor distância entre os dois continentes (BLAKER, 1971; CROSBY, 1972). A hipótese de travessia transatlântica é suportada pela capacidade de dispersão da espécie a partir dos sítios natais, mantendo o vôo por mais de 40h em uma velocidade variando entre 50 e 90 km/h (BROWDER, 1973; MADDOCK; BRIDGMAN, 1992). Alguns autores defendem que essas aves africanas provavelmente ainda estão entrando nas Américas, devido ao comportamento desorientado de dispersão dos juvenis após a temporada reprodutiva (BROWDER, 1973; TELFAIR, 1994). A dispersão dos juvenis parece ser um mecanismo de controle populacional, em que o número de indivíduos que dispersam a partir dos sítios natais é função do número de jovens produzidos na temporada reprodutiva ou da densidade populacional (WYNNE-EDWARDS, 1962). Embora seja surpreendente a colonização recente das Américas, considerando a possibilidade de travessias transatlânticas no passado, apenas no último século se intensificou a formação de pastagens em áreas anteriormente inadequadas para *Bubulcus ibis* (TELFAIR, 1994).

A atmosfera altamente dinâmica tem um impacto importante em todos os aspectos do voo migratório de uma ave, afetando desde a decisão de partida, até a sua velocidade, altitude e direção (ALERSTAM, 1979; LIECHTI, 2006). Essas aves selecionam as condições favoráveis de vento para economizar energia e para isso, geralmente, os ventos de cauda são indispensáveis (LIECHTI, 2006). Além dos registros de *Bubulcus ibis* em ilhas oceânicas, Siegfried (1978) indicou, com base em cálculos fisiológicos, que a espécie possui o potencial de voo necessário para cruzar o oceano Atlântico sem escalas, talvez auxiliada por ventos alísios de cauda. A dependência das aves nesses ventos entre a África e a América do Sul pode explicar a rota de invasão da população fundadora de *Bubulcus ibis* no continente americano (TELFAIR, 1994). Travessias transatlânticas favorecidas por sistemas de ventos e realizadas em intervalos relativamente curtos já foram documentadas em espécies de aves marinhas com o uso de geolocalizadores (EGEVANG et al., 2010; FELICÍSIMO; MUÑOZ; GONZÁLEZ-SOLIS, 2008; HEDD et al., 2012). Outras espécies pertencentes à família Ardeidae, *Ardea cinerea* (Linnaeus, 1758) e *Egretta garzetta* (Linnaeus, 1758), tiveram travessias transatlânticas registradas (DOWNS, 1959; BADOUIN-BODIN, 1960). Indivíduos de



ambas as espécies foram anilhados na Europa e tiveram suas anilhas recuperadas nas Américas (DOWNS, 1959; BADOUIN-BODIN, 1960).

Para *Bubulcus ibis*, simulações baseadas em padrões climáticos e de circulação atmosférica entre 1871 e 1920 foram usadas para calcular rotas ao longo da costa da África Ocidental e das Américas Central e do Sul, avaliando teorias existentes sobre a possibilidade da espécie cruzar o oceano Atlântico (MASSA; DOYLE; CALLICÓ FORTUNATO, 2014). Foram supostas saídas a partir do sul, centro e norte do continente africano e chegadas na Guiana Francesa, Suriname, Guiana, Venezuela, nas ilhas do Caribe e no Brasil, considerando somente a costa acima de 5° S de latitude (MASSA; DOYLE; CALLICÓ FORTUNATO, 2014). As rotas foram simuladas a aproximadamente 1km de altitude sob a pressão atmosférica de 850 a 900 hPa, onde é mais provável que a espécie migre (MADDOCK; BRIDGMAN, 1992). De 192.864 trajetórias analisadas, 1.695 rotas foram encontradas saindo da África Central e alcançando a costa nordeste da América do Sul e as ilhas do Caribe em menos de uma semana, com o maior número de chegadas ao leste da área brasileira analisada (MASSA; DOYLE; CALLICÓ FORTUNATO, 2014).

### 1.3 A proposta de trabalho: Capítulo II

Neste capítulo, a meta foi inferir a rota mais provável de invasão durante o processo de colonização do Brasil por *Bubulcus ibis*. Para alcançá-la, avaliou-se como a variação genética está distribuída entre as populações da espécie em áreas nativas (continente africano e sul da Europa) e colonizadas (Brasil), aproximadamente um século após o processo de colonização. Dois conjuntos de dados foram utilizados, um com os SNPs identificados ao longo do genoma de *Bubulcus ibis* e apresentados no capítulo I e outro composto por um subconjunto de oito *loci* microssatélites entre os 25 também apresentados anteriormente. Esses oito microssatélites foram selecionados porque foram utilizados anteriormente no estudo intitulado “Diversidade e estruturação genética em populações africanas e brasileiras de garça-vaqueira (*Bubulcus ibis*, Linnaeus, 1758) determinadas por microssatélites” desenvolvido paralelamente em colaboração (GERONIMO, 2018).

Dentro dessa proposta, estimou-se o fluxo gênico e determinou-se os níveis de diversidade e estruturação genética entre as populações das duas regiões. A expectativa

foi de que os dados de estruturação genética e de fluxo gênico obtidos a partir de marcadores de maior resolução (SNPs) permitissem inferir uma rota de invasão durante o processo de colonização do Brasil por *Bubulcus ibis*. Ainda nessa proposta, foi avaliado se a estruturação genética encontrada entre as populações de *Bubulcus ibis* poderia ser explicada pela distância geográfica e/ou por seleção natural. Assumimos que uma correlação entre a distribuição da variação genética e a distribuição geográfica da espécie seria um indício de colonização seguindo um modelo de progressão *stepping stone*. Da mesma forma, previmos que a identificação de SNPs potencialmente sob seleção e com maior contribuição para a estruturação genética populacional seria um indicativo de um fator adaptativo contribuindo para o padrão de estruturação genética.

## 2 HIPÓTESES E OBJETIVOS

### 2.1 Hipóteses

- a. Níveis mais altos de diversidade genética são esperados nas populações de *Bubulcus ibis* da área nativa do que nas populações da área colonizada brasileira, uma vez que não há indícios que suportem o fluxo gênico contínuo entre estas áreas e supõe-se um efeito fundador no Brasil;
- b. A rota mais provável de invasão de *Bubulcus ibis* no continente americano é suposta com partidas da região centro-oeste africana e chegadas na região norte da América do Sul devido aos registros históricos de avistamentos;
- c. Mesmo na ausência de fluxo gênico contínuo é esperada uma estruturação genética baixa entre as populações de *Bubulcus ibis* nas áreas colonizadas e nas nativas, considerando o intervalo de tempo decorrido desde a colonização;
- d. A colonização do Brasil por *Bubulcus ibis* pode ter ocorrido em um modelo de progressão *stepping stone*, como esperado para espécies de aves aquáticas coloniais;
- e. Um fator adaptativo pode estar contribuindo na determinação da estruturação genética entre as populações de *Bubulcus ibis*, considerando a heterogeneidade ambiental ao longo da distribuição da espécie e a possibilidade de adaptação local.

### 2.2 Objetivo geral

Inferir o padrão de estruturação genética atual e uma possível rota de invasão durante o processo de colonização do Brasil pela garça-vaqueira, *Bubulcus ibis*, a partir da distribuição da variação genética e a conectividade entre populações naturais da espécie em áreas da sua distribuição nativa (continente africano e sul da Europa) e áreas colonizadas do território brasileiro.

### 2.3 *Objetivos específicos*

- a. Determinar a diversidade genética em populações de *Bubulcus ibis* em áreas nativas e colonizadas dentro da distribuição da espécie;
- b. Inferir o fluxo gênico entre as populações de *Bubulcus ibis* estudadas;
- c. Determinar a diferenciação genética entre as populações de *Bubulcus ibis* estudadas;
- d. Avaliar se há uma correlação entre a distâncias genética e geográfica determinadas entre as populações de *Bubulcus ibis* estudadas;
- e. Identificar e mapear no genoma de referência de *Bubulcus ibis* SNPs potencialmente sob seleção, com maior contribuição com o padrão de estruturação genética entre as populações e/ou associados a variáveis ambientais;

## 3 MÉTODOS

### 3.1 *Dados de microssatélites no estudo da colonização do Brasil por Bubulcus ibis*

A partir do conjunto de genótipos de microssatélites apresentados para as populações de São Paulo, Fernando de Noronha e África do Sul no Capítulo I, foi definido subconjunto de oito *loci* de microssatélites (Bi39, Bi40, Bi42, Bi45, Bi46, Bi49, Bi50 e Bi55). Esses dados foram agrupados com os dados produzidos nesses mesmos *loci* em colaboração com Gerônimo (2018) para outras oito populações: Rio Grande do Sul ( $n = 29$ ), Pernambuco ( $n = 20$ ), Rio Grande do Norte ( $n = 32$ ), Pará ( $n = 30$ ), Guiné-Bissau ( $n = 23$ ), Senegal ( $n = 30$ ), Gana ( $n = 20$ ) e Nigéria ( $n = 30$ ). Com esse conjunto de dados ( $n = 306$ ), avaliou-se a diferenciação genética e calculou-se as taxas de migração entre as populações de *Bubulcus ibis* amostradas nas áreas nativas e recém colonizadas. Para avaliar a diferenciação genética foram realizadas uma análise de  $F_{ST}$  par-a-par (WEIR;

COCKERHAM, 1984) no pacote Genepop v4.7 do programa R v3.4.3 (ROUSSET, 2008) e uma análise de variância molecular (AMOVA,  $10^5$  passos de dememorização e MCMC) no programa ARLEQUIN v3.5.2.2 (EXCOFFIER; LISCHER, 2010). Na AMOVA, as populações foram divididas em dois grupos, um com as populações amostradas na área nativa da e outro com as populações brasileiras. Além disso, os algoritmos Bayesianos de agrupamento genético implementados nos programas STRUCTURE v2.3.4 (PRITCHARD; STEPHENS; DONNELLY, 2000) e TESS v3 (CAYE et al., 2015) foram utilizados para investigar a existência de estruturação genética entre essas populações. No programa STRUCTURE, assumiu-se um modelo de mistura com frequências alélicas correlacionadas, sem a informação *a priori* sobre os locais de amostragem e testando valores de  $K$  entre 1 e 5 (20 corridas por valor de  $K$ , com  $10^6$  iterações de MCMC e *burn-in* de  $5 \times 10^4$  por corrida). O número mais provável de grupos genéticos foi determinado avaliando os valores de Delta  $K$  no programa POPHELPER v1.0 (EVANNO; REGNAUT; GOUDET, 2005; FRANCIS, 2017). No programa TESS, as análises foram feitas testando os três modelos disponíveis na aplicação (sem mistura, BYM e CAR) e valores de  $K$  entre 1 e 10 (5 corridas por valor de  $K$  usando  $7 \times 10^5$  *sweeps* e *burn-in* de  $2 \times 10^4$  por corrida) e incorporando a localização geográfica das populações como *prior*.

As taxas de migração entre as populações foram estimadas ( $6 \times 10^8$  iterações de MCMC e *burn-in* de  $4 \times 10^7$ ) pelo método Bayesiano de Wilson e Rannala (2003) no programa BayesAss v3.04 (RANNALA, 2007). O programa Tracer v1.7 (RAMBAUT et al., 2018) foi utilizado para visualizar e analisar as amostras geradas por MCMC a partir da distribuição posterior.

### 3.2 Dados de SNPs no estudo da colonização do Brasil por *Bubulcus ibis*

Para se testar as hipóteses levantadas nesse capítulo (ver item 2) sobre a colonização do Brasil pela espécie-foco foi utilizado um conjunto de genótipos, determinados a partir do painel de SNPs identificados nas 11 populações de *Bubulcus ibis* amostradas (Tabela 2 e Figura 3 do Capítulo I) e descritos no Capítulo I. Com esse conjunto de dados determinou-se os níveis de diversidade genética para cada população, bem como a diferenciação genética e a conectividade por fluxo gênico entre elas. A maior parte das análises com esse conjunto de dados foi realizada em um servidor (IBM x3859 x5) localizado no Departamento de Genética e Biologia Evolutiva do Instituto de Biociências da Universidade de São Paulo.

As análises de diversidade genética foram realizadas no programa R v3.4.3, utilizando os pacotes *diveRsity* (KEENAN et al., 2013) e *hierfstat* (GOUDET, 2005), e no programa *ARLEQUIN* v3.5.2.2 (EXCOFFIER; LISCHER, 2010). Os parâmetros de diversidade genética estimados foram: o número de alelos por população ( $N_a$ ), a riqueza alélica ( $A_R$ ), as heterozigosidades observada ( $H_o$ ) e esperada ( $H_e$ ), o coeficiente de endogamia ( $F_{IS}$ ) (WEIR; COCKERHAM, 1984) e os números de substituições (transições e transversões). Estimativas do  $\theta$  de Watterson (WATTERSON, 1975) foram feitas a partir do número observado de sítios segregantes ( $\theta_S$ ) e dos valores médios de diferenças entre pares de SNPs por população ( $\theta\pi$ ). O teste de Wilcoxon ranqueado foi utilizado para testar se ocorria diferenças significativas entre os níveis de diversidade genética, determinados para as populações de *Bubulcus ibis* amostradas ( $\alpha = 0,05$ ).

A diferenciação genética entre as populações estudadas foi investigada, inicialmente, a partir de uma análise de variância molecular (AMOVA) e de estimativas de  $F_{ST}$  par-a-par (WEIR; COCKERHAM, 1984) realizadas no programa *ARLEQUIN* v3.5.2.2, bem como a partir de uma análise espacial de variância molecular (SAMOVA) realizada no programa *SAMOVA* v2.0 (DUPANLOUP; SCHNEIDER; EXCOFFIER, 2002; EXCOFFIER; LISCHER, 2010). A AMOVA foi realizada considerando todas as populações como pertencentes a um grupo único e dividindo-as em dois grupos: área nativa e colonizada. Na SAMOVA avaliou-se a divisão das populações em dois a cinco grupos, sem informar *a priori* os agrupamentos mais prováveis.

Para investigar a existência de estruturação genética entre as populações amostradas de *Bubulcus ibis*, empregou-se uma análise de componentes principais (PCA; JOLLIFFE, 2011) e uma análise discriminante de componentes principais no pacote *adegenet* (DAPC; JOMBART; DEVILLARD; BALLOUX, 2010) do programa R. Além disso, estimativas das proporções de ancestralidade dos genótipos observados foram modeladas pelos métodos Bayesianos implementados nos programas *Admixture* v1.23 (ALEXANDER; NOVEMBRE; LANGE, 2009), *STRUCTURE* v2.3.4 (PRITCHARD; STEPHENS; DONNELLY, 2000) e *BAPS* v6.0 (CORANDER et al., 2004; CORANDER; WALDMANN; SILLANPA, 2003), por uma análise baseada na fatoração de uma matriz geograficamente restrita no pacote *tess3r* (CAYE et al., 2015) e por um método de fatoração de matriz esparsa não negativa (função *snmf*) do pacote *LEA* do programa R (FRICHOT; FRANÇOIS, 2015).

O método de PCA fornece as coordenadas de cada indivíduo ao longo de eixos de variação genética em um plano cartesiano e serve como uma aproximação focada na

descrição da diversidade genética global da amostra (JOLLIFFE, 2011). A DAPC encontra variáveis sintéticas (funções discriminantes) que maximizem o componente de variação entre as populações definidas *a priori*, minimizando a variação dentro de cada população (JOMBART; DEVILLARD; BALLOUX, 2010). NA PCA, os dados faltantes foram imputados com base nas frequências alélicas determinadas em cada população. Na DAPC, valores de  $K$  entre 1 e 20 foram testados e comparados usando o Critério de Informação Bayesiana (BIC) para se definir o número de agrupamentos genéticos ( $K$ ) que maximizassem a variação entre os grupos. O melhor valor de  $K$  foi definido como o primeiro ponto de inflexão na curva de crescimento dos valores de BIC plotados.

No programa Admixture, cinco corridas foram realizadas para os valores de  $K$  de 1 a 12, usando um procedimento de validação cruzada com 200 iterações. Definiu-se o número mais provável de grupos genéticos nas amostras avaliando para qual valor de  $K$  foi obtido o menor erro de validação cruzada.

No programa STRUCTURE, assumiu-se um modelo de mistura com frequências alélicas correlacionadas, sem a informação *a priori* sobre os locais de amostragem, utilizando um valor de lambda igual a 0,24, e testando valores de  $K$  entre 1 e 12 (10 corridas por valor de  $K$ , com  $2 \times 10^5$  iterações de MCMC e *burn-in* de  $5 \times 10^4$  por corrida). Ainda no programa STRUCTURE, foi analisado um segundo conjunto de dados, construído a partir do das 6.043 sequências consenso nas quais os SNPs foram identificados. Para gerar esse conjunto de dados, o arquivo de saída do programa ipyrad no formato \*.alleles, que inclui as sequências completas dos *loci*, foi convertido especificamente para o formato \*.str por um *script* customizado ([https://github.com/jordansatler/pyRAD\\_to\\_Structure](https://github.com/jordansatler/pyRAD_to_Structure)), mantendo a informação completa dos haplótipos. Esse conjunto foi analisado, testando valores de  $K$  de 1 a 8 e seguindo os mesmos parâmetros descritos acima. Por último, uma análise foi feita com os subconjuntos de populações amostradas no Brasil e na área nativa de distribuição da espécie, testando valores de  $K$  entre 1 e 6 para cada grupo. O número mais provável de grupos genéticos foi determinado, para todas as análises realizadas no programa STRUCTURE, avaliando os valores de Delta  $K$  no programa POPHELPER v1.0 (EVANNO; REGNAUT; GOUDET, 2005; FRANCIS, 2017).

No programa BAPS, realizou-se uma análise particionada com  $5 \times 10^3$  iterações, assumindo um modelo de mistura com dois grupos pré-definidos de populações, o primeiro composto pelas populações brasileiras e o segundo composto pelas populações da área nativa de *Bubulcus ibis*.

No pacote tess3r, as análises foram feitas incorporando a localização geográfica das populações e testando valores de  $K$  entre 1 e 20. O número mais provável de agrupamento genéticos no pacote tess3r foi baseado nas pontuações obtidas por um procedimento de validação cruzada, semelhante ao realizado no programa Admixture.

As análises usando a função *snmf* foram realizadas com 20 repetições por valor de  $K$ , tolerância de 0,00001, testando valores de  $K$  entre 1 e 10 e valores de  $\alpha$  entre 0 e 1.000 (0, 10, 100 e 1000). A escolha do número mais provável de grupos genéticos inferidos pela função *snmf* foi baseada no critério de entropia cruzada.

Para se investigar a conectividade entre as populações de *Bubulcus ibis* amostradas, três abordagens principais foram empregadas: a inferência de padrões de divisão e mistura populacional histórica no programa TreeMix v1.13 (PICKRELL; PRITCHARD, 2012), a modelagem da relação entre a variação genética e a distribuição geográfica usando o conceito de “migração efetiva” na aplicação EEMS (PETKOVA; NOVEMBRE; STEPHENS, 2015) e uma análise de atribuição populacional foi realizada no programa GENODIVE (MEIRMANS; VAN TIENDEREN, 2004). O modelo implementado no programa Treemix utiliza as frequências alélicas das populações e uma aproximação Gaussiana da deriva genética para inferir uma árvore de máxima verossimilhança que mostra uma relação de ancestralidade e, opcionalmente, de fluxo gênico entre as populações amostradas (PICKRELL; PRITCHARD, 2012). Enquanto, o programa EEMS permite visualizar padrões de fluxo gênico em modelos não estacionários de isolamento por distância, assumindo a premissa que a migração efetiva é baixa em regiões onde a similaridade genética decai rapidamente (PETKOVA; NOVEMBRE; STEPHENS, 2015). Na análise realizada no programa Treemix, foram testados 16 modelos, considerando desde a ausência de eventos de migração até a ocorrência de 15 eventos de migração entre as populações. O modelo mais provável foi definido com base nos erros padrões médios estimados pelo programa para uma matriz de covariância da amostra populacional. No programa EEMS, a convergência dos logs posteriores da análise foi obtida com  $3 \times 10^7$  iterações de MCMC e  $10^7$  de *burn-in*. Para a análise de atribuição populacional realizada no programa GENODIVE, estimou-se a probabilidade de que o genótipo de cada indivíduo fosse encontrado em uma determinada população, utilizando um procedimento de reamostragem com  $10^4$  iterações de MCMC ( $\alpha = 0,05$ ) (PAETKAU et al., 1995). As atribuições foram realizadas com base nas frequências alélicas calculadas após a remoção do indivíduo-alvo da população de origem.

### 3.3 Padrão geográfico e/ou adaptativo na distribuição da variação genética populacional de *Bubulcus ibis*

Será que algum modelo geográfico e/ou adaptativo explicaria a distribuição da variação genética encontrada com o conjunto de genótipos de SNPs determinados para as populações amostradas de *Bubulcus ibis*? Para responder essa questão procurou-se inicialmente identificar se existia um fator geográfico contribuindo para o padrão de estruturação genética determinado para as populações de *Bubulcus ibis* estudadas. O teste de Mantel no pacote *adegenet* (JOMBART; AHMED, 2011) do programa R foi utilizado para avaliar se a distribuição da variação genética populacional se enquadrava em um modelo de isolamento por distância (IBD). Na análise, testou-se a correlação entre uma matriz de distâncias genéticas euclidianas estimadas a partir da estatística  $F$  de Weir e Cockerham (1984) e uma matriz de distâncias geográficas calculadas na aplicação online *GeoDataSource™* (<https://www.geodatasource.com/distance-calculator>), testando a hipótese nula de ausência de IBD em um modelo com  $10^5$  réplicas de simulações.

Em seguida, para identificar se existia um fator adaptativo influenciando a estruturação genética entre as populações de *Bubulcus ibis* estudadas, identificou-se os SNPs com maior contribuição para a diferenciação entre os grupos genéticos inferidos ( $\geq 0,02\%$ ) e buscou-se por SNPs potencialmente sob seleção e/ou correlacionados a variáveis ambientais. Os SNPs com maior contribuição para o padrão de estruturação genética foram identificados, a partir dos dados da DAPC, pela função *loading.plot* do pacote *adegenet* do programa R (JOMBART; DEVILLARD; BALLOUX, 2010). Para identificar SNPs potencialmente sob seleção foram utilizados o método de *FST-outlier* ( $5 \times 10^4$  corridas pilotos e *burn-in*) implementado no programa *BayeScan* v2.01 (BEAUMONT; NICHOLS, 1996; FOLL; GAGGIOTTI, 2008), o método de coalescência (100 demes e  $10^5$  simulações) implementado no programa *ARLEQUIN* v3.5.2.2 (EXCOFFIER; LISCHER, 2010) e o teste baseado na diferenciação das frequências alélicas e nos coeficientes de ancestralidade espacialmente variáveis do pacote *tess3r* (CAYE et al., 2015). O procedimento de correção sequencial de Benjamini-Hochberg foi utilizado para ajustar os valores de significância nos testes múltiplos (FERREIRA; ZWINDERMAN, 2006).

Para detectar SNPs potencialmente associados a variáveis ecológicas, utilizou-se a função *lmm* do pacote *LEA* (FRICHOT; FRANÇOIS, 2015) do programa R. Foram



selecionadas sete variáveis ambientais (Tabela 8). A escolha das variáveis ambientais foi baseada em dois critérios: a) a sua disponibilidade nas bases de dados WorldClim v2.0 (<http://worldclim.org/version2>) e EarthData v2.2 (<https://earthdata.nasa.gov/>) (FICK; HIJMANS, 2017) e cobrindo um período de pelo menos 20 anos, e b) a representatividade, à primeira vista, da heterogeneidade ambiental observada entre as localidades amostradas. Uma análise de componentes principais usando a função *princomp* no programa R foi realizada para avaliar quais das variáveis ambientais selecionadas eram significativamente informativas.

Foi utilizada ainda a ferramenta *makeblastdb* do programa *ncbi-blast* v2.6.0 (ALTSCHUL et al., 1990) foi utilizada para criar um banco de dados local com o genoma de referência de *Bubulcus ibis*. Mapeou-se no genoma de referência as sequências que incluíram os SNPs indicados pela DAPC como maiores contribuidores para o padrão de estruturação genética encontrado e os SNPs com correlação positiva com alguma das variáveis ambientais testadas. Além do genoma de referência, a ferramenta *blastn* (<https://blast.ncbi.nlm.nih.gov/>) foi utilizada para buscar sequências de nucleotídeos correspondentes na base de dados NCBI.

**Tabela 8** - Variáveis ambientais extraídas para as localidades das populações de garça-vaqueira (*Bubulcus ibis*) amostradas. As médias anuais de temperatura (Tmed), temperaturas máxima (Tmax) e mínima (Tmin), vento (V), incidência de radiação solar (Rad), pressão do vapor de água (PV) e precipitação (P) são exibidas para as localidades das populações amostradas em Portugal (PT), Guiné Bissau (GW), Senegal (SN), Gana (GH), África do Sul (ZA), Nigéria (NG), Pará (PA), Fernando de Noronha (FN), Pernambuco (PE), São Paulo (SP) e Rio Grande do Sul (RS).

Pop	Tmed (°C)	Tmax (°C)	Tmin (°C)	V (m s <sup>-1</sup> )	Rad (kJ m <sup>-2</sup> day <sup>-1</sup> )	PV (kPa)	P (mm)
PT	16,62	28,74	2,58	3,13	17191,33	1,25	557,00
GW	26,51	33,27	18,58	2,01	19714,33	2,29	1500,00
SN	24,52	32,98	17,45	4,03	21011,67	2,21	360,00
GH	27,28	31,95	21,21	2,32	16453,33	2,97	845,00
NG	21,91	29,82	12,54	2,43	17455,58	1,48	1200,00
ZA	17,40	25,84	8,56	4,74	18305,17	1,46	569,00
PA	27,02	33,32	21,90	1,68	15937,17	2,88	3108,00
FN	26,58	29,61	23,17	4,00	19830,42	2,74	1429,00
PE	20,74	27,78	13,07	1,94	18039,00	2,01	828,00
SP	20,57	29,42	10,64	1,98	14871,17	1,83	1301,00
RS	19,66	28,54	9,53	2,57	14602,58	1,78	1429,00

Fonte: Base de dados WorldClim v2.0 (FICK; HIJMANS, 2017).

## 4 RESULTADOS

### 4.1 A colonização do Brasil por *Bubulcus ibis*: dados de microssatélites

As análises realizadas com o conjunto de genótipos determinados em oito *loci* de microssatélites mostraram um padrão de estruturação genética entre as populações de *Bubulcus ibis* amostradas no continente africano e no Brasil (Tabela 9 e Figura 9). Baseando-se nos valores significativos ( $p < 0,05$ ) de  $F_{ST}$  par-a-par obtidos, três principais observações puderam ser feitas: a) a população de Pernambuco foi a única a não diferir geneticamente de nenhuma outra população amostrada e, b) as populações de Guiné Bissau, Senegal e Gana, no centro-oeste do continente africano, apresentaram um número menor de diferenças significativas na comparação entre populações brasileiras e africanas e c) a população da África do Sul diferiu geneticamente de todas as outras populações africanas (Tabela 9). A AMOVA indicou um padrão de estruturação genética baixa entre as populações amostradas, sendo que as diferenças encontradas entre os dois continentes e entre as populações de cada continente explicaram 0,56% e 1,34%, respectivamente, da variação genética global ( $F_{ST} = 0,019$ ,  $F_{SC} = 0,013$ ,  $F_{CT} = 0,006$ ,  $p < 0,01$ ).

**Tabela 9** - Diferenciação genética determinada entre as populações de garça-vaqueira (*Bubulcus ibis*) amostradas a partir da informação de oito *loci* de microssatélites. Valores de  $F_{ST}$  (diagonal inferior) e de significância ( $p$ ) correspondentes (diagonal superior) foram calculados para um conjunto de 11 populações (Pop,  $n = 306$ ): Guiné Bissau (GW,  $n = 23$ ), Senegal (SN,  $n = 30$ ), Gana (GH,  $n = 20$ ), África do Sul (ZA,  $n = 29$ ) e Nigéria (NG,  $n = 30$ ), no continente africano, e Pará (PA,  $n = 30$ ), Fernando de Noronha (FN,  $n = 31$ ), Pernambuco (PE,  $n = 20$ ), Rio Grande do Norte (RN,  $n = 32$ ), São Paulo (SP,  $n = 32$ ) e Rio Grande do Sul (RS,  $n = 29$ ).

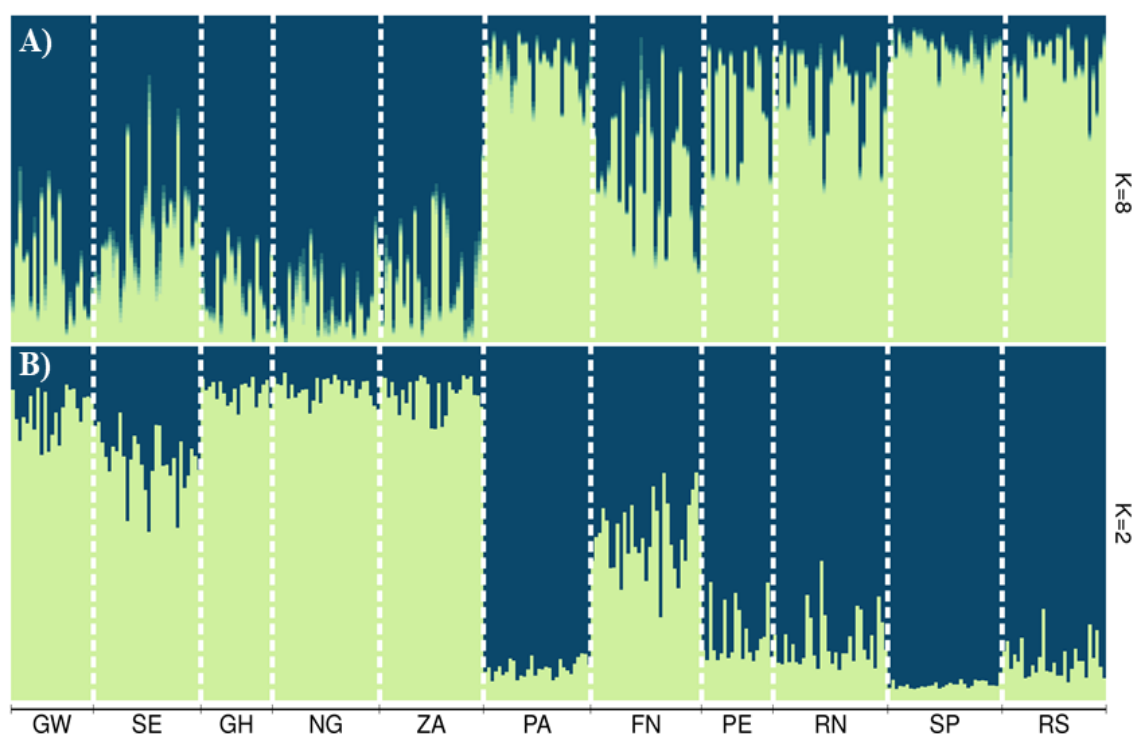
Pop	GW	SN	GH	ZA	NG	PA	FN	PE	RN	SP	RS
GW		0,66	0,75	0,00	0,29	0,01	0,01	0,71	0,01	0,01	0,20
SN	0,00		0,58	0,01	0,70	0,01	0,00	0,85	0,00	0,01	0,36
GH	0,00	0,00		0,00	0,72	0,00	0,05	0,75	0,04	0,00	0,03
ZA	0,03	0,01	0,02		0,00	0,00	0,00	0,11	0,00	0,00	0,00
NG	0,00	0,00	0,00	0,02		0,00	0,00	0,34	0,00	0,00	0,01
PA	0,01	0,01	0,02	0,03	0,02		0,04	0,13	0,01	0,02	0,58
FN	0,02	0,02	0,01	0,03	0,03	0,00		0,10	0,00	0,00	0,00
PE	0,00	0,00	0,00	0,01	0,00	0,01	0,01		0,13	0,37	0,59
RN	0,02	0,02	0,01	0,04	0,02	0,01	0,05	0,01		0,00	0,06
SP	0,02	0,02	0,03	0,03	0,04	0,01	0,03	0,00	0,04		0,01
RS	0,00	0,00	0,01	0,03	0,01	0,00	0,03	0,00	0,01	0,10	

Legenda: Foram destacados em vermelho os valores de  $F_{ST} > 0,010$  e de  $p$  significativos ( $\alpha < 0,05$ ).

Fonte: Produção do próprio autor.

Os resultados das análises Bayesianas de clusterização realizadas nos programas TESS v3 ( $K = 8$ ) e STRUCTURE v2.3.4 ( $K = 2$ ) foram concordantes sobre a existência de dois grupos genéticos mais representativos, respectivamente, entre as populações africanas e do Brasil (Figuras 9A e 9B). Essas análises colocaram a população de *Bubulcus ibis* amostrada em Fernando de Noronha em uma posição intermediária entre os dois principais grupos genéticos inferidos, apresentando proporções semelhantes de ancestralidade de ambos (Figuras 9A e 9B). Entre as demais populações brasileiras, as populações do Pará e de São Paulo foram as que mais diferiram em relação a área nativa, apresentando uma menor proporção de ancestralidade com o grupo genético mais representativo na África (Figuras 9A e 9B). No continente africano, a população de Senegal exibiu uma proporção maior do agrupamento majoritariamente brasileiro (Figuras 9A e 9B).

**Figura 9** - Evidência de diferenciação genética entre as populações de garça-vaqueira (*Bubulcus ibis*) amostradas obtida a partir da informação de oito *loci* de microssatélites. Os gráficos mostram as proporções de ancestralidade genética estimadas nos programas A) TESS ( $K = 8$ ) e B) STRUCTURE ( $K = 2$ ) a partir dos dados de 11 populações amostradas ( $n = 306$ ): Guiné Bissau (GW,  $n = 23$ ), Senegal (SN,  $n = 30$ ), Gana (GH,  $n = 20$ ), África do Sul (ZA,  $n = 29$ ) e Nigéria (NG,  $n = 30$ ), no continente africano, e Pará (PA,  $n = 30$ ), Fernando de Noronha (FN,  $n = 31$ ), Pernambuco (PE,  $n = 20$ ), Rio Grande do Norte (RN,  $n = 32$ ), São Paulo (SP,  $n = 32$ ) e Rio Grande do Sul (RS,  $n = 29$ ).



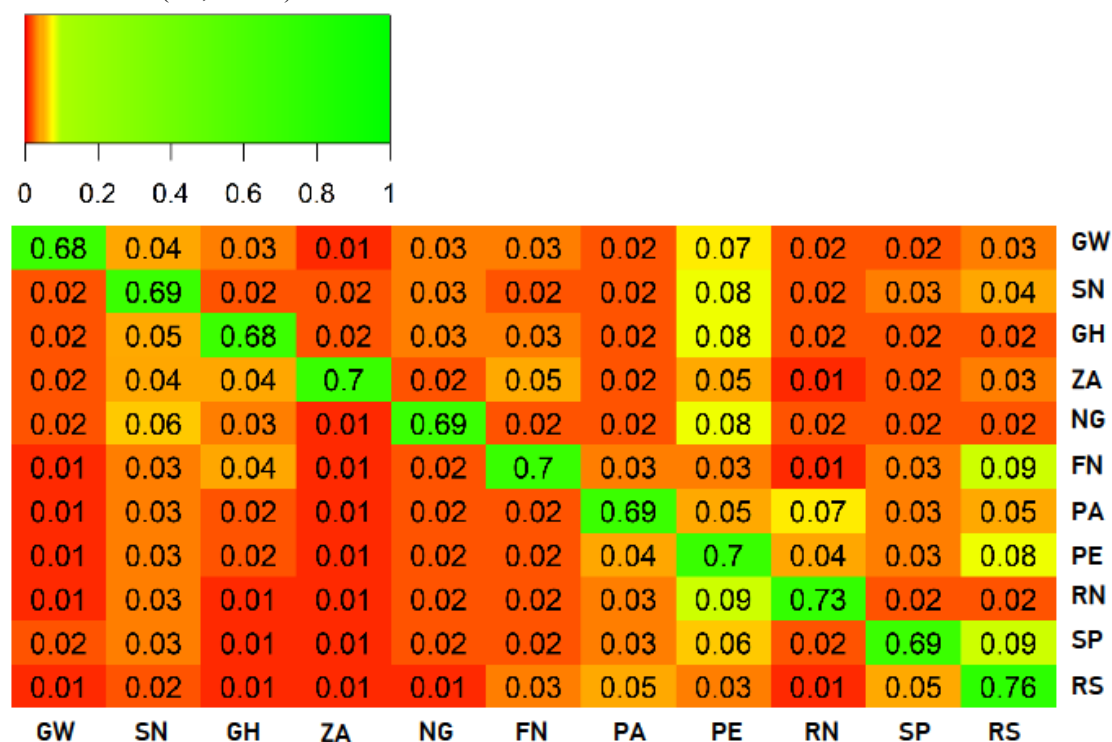
Legenda: A divisão entre as populações é indicada por uma linha branca tracejada e cada barra vertical corresponde às proporções de ancestralidade dos grupos genéticos, caracterizados por cores diferentes, estimados para cada indivíduo amostrado.

Fonte: Produção do próprio autor.

As taxas de migração calculadas entre populações foram muito baixas, oscilando

entre 0,01 e 0,09 (Figura 10). E as taxas calculadas para as próprias populações, que indicam a taxa de não-migrantes por geração, oscilaram entre 0,68 (GW e GH) e 0,76 (RS) (Figura 10). Entre a distribuição dos valores calculados, uma conectividade maior entre a população do Pernambuco, no Brasil, e as populações do centro-oeste africano (GW, SN, GH e NG) foi encontrada (Figura 10). Além disso, a população do Rio Grande do Sul exibiu taxas de migração maiores dentro do país em relação às outras populações brasileiras amostradas (Figura 10). No continente africano, a África do Sul foi a população que apresentou menor conectividade tanto com as populações brasileiras quanto com o restante das populações africanas (Figura 10).

**Figura 10** - Taxas de migração (m) estimadas entre as populações da garça-vaqueira (*Bubulcus ibis*) amostradas a partir da informação de oito *loci* de microssatélites. Os valores, em frequência de indivíduos por geração, determinados usando o programa BayesAss v3.04 são mostrados para 11 populações amostradas ( $n = 306$ ): Guiné Bissau (GW,  $n = 23$ ), Senegal (SN,  $n = 30$ ), Gana (GH,  $n = 20$ ), África do Sul (ZA,  $n = 29$ ) e Nigéria (NG,  $n = 30$ ), no continente africano, e Pará (PA,  $n = 30$ ), Fernando de Noronha (FN,  $n = 31$ ), Pernambuco (PE,  $n = 20$ ), Rio Grande do Norte (RN,  $n = 32$ ), São Paulo (SP,  $n = 32$ ) e Rio Grande do Sul (RS,  $n = 29$ ).



Legenda: Valores plotados em um mapa de calor com um gradiente entre as cores vermelho (valores menores) e verde (valores maiores).

Fonte: Produzido pelo próprio autor no pacote gplot do programa R.

#### 4.2 A colonização do Brasil por *Bubulcus ibis*: dados de SNPs

A partir do conjunto de dados composto por 2.751 SNPs genotipados em 174 indivíduos amostrados a partir de 11 populações de *Bubulcus ibis* no Brasil, em Portugal

e no continente africano, determinou-se como a variação genética está distribuída geograficamente entre essas populações e buscou-se por evidências de conectividade por fluxo gênico entre elas. Nesse conjunto de dados, o número total de alelos identificados variou entre 7.744 (população de PT) e 8.084 (população da ZA), porém a riqueza alélica, corrigida pelo tamanho amostral, foi idêntica entre todas as populações ( $A_R = 1,23$ ) (Tabela 10). Apesar da maioria dos valores médios de diversidade genética estimados terem sido menores para o grupo das populações brasileiras ( $H_O = 0,056$ ,  $H_E = 0,088$ ,  $T_S = 1.748$ ,  $T_V = 481$ ,  $\theta S = 547$  e  $\theta\pi = 568$ ) do que para o grupo das populações da área nativa ( $H_O = 0,058$ ,  $H_E = 0,093$ ,  $T_S = 1.879$ ,  $T_V = 510$ ,  $\theta S = 586$  e  $\theta\pi = 590$ ), os índices de diversidade não diferiram significativamente entre os dois grupos de populações ( $p < 0,05$  usando o teste de Wilcoxon ranqueado) (Tabela 10).

**Tabela 10** - Diversidade genética determinada para as populações de garça-vaqueira (*Bubulcus ibis*) a partir da informação de 2.751 SNPs. O número total de alelos ( $N_a$ ), a riqueza alélica ( $A_R$ ), as heterozigosidades observada ( $H_O$ ) e esperada ( $H_E$ ), o coeficiente de endogamia ( $F_{IS}$ ), os números de substituições por transições ( $T_S$ ) e transversões ( $T_V$ ), e os estimadores Thetas de Waterson calculados a partir do número de sítios segregantes ( $\theta S$ ) e dos valores médios de diferenças entre pares de SNPs ( $\theta\pi$ ) foram estimados para cada uma das 11 populações analisadas ( $n = 174$ ): Portugal (PT,  $n = 9$ ), Guiné Bissau (GW,  $n = 17$ ), Senegal (SN,  $n = 16$ ), Gana (GH,  $n = 16$ ), Nigéria (ZA,  $n = 15$ ) e África do Sul (ZA,  $n = 17$ ) na área nativa, e Pará (PA,  $n = 17$ ), Fernando de Noronha (FN,  $n = 17$ ), Pernambuco (PE,  $n = 17$ ), São Paulo (SP,  $n = 17$ ) e Rio Grande do Sul (RS,  $n = 16$ ) no Brasil.

Pop	$N_a$	$A_R$	$H_O$	$H_E$	$F_{IS}$	$T_S$	$T_V$	$\theta S$	$\theta\pi$
PT	7744	1,23	0,059	0,089	0,499	1943	520	605	602
GW	8105	1,23	0,063	0,097	0,501	1831	508	563	581
SN	8074	1,23	0,057	0,094	0,048	1777	488	552	562
GH	8054	1,23	0,057	0,093	0,014	1842	492	573	571
NG	8072	1,23	0,057	0,092	0,496	1916	517	601	614
ZA	8084	1,23	0,056	0,092	0,099	1969	534	623	612
PA	8076	1,23	0,055	0,085	0,006	1753	489	535	548
FN	8053	1,23	0,056	0,088	0,502	1894	538	577	595
PE	8113	1,23	0,056	0,089	0,504	1767	469	618	650
SP	8062	1,23	0,060	0,095	0,165	1590	439	478	494
RS	8044	1,23	0,053	0,085	0,497	1735	468	529	552

Fonte: Produção do próprio autor.

Os valores globais encontrados para a estatística  $F$  pela AMOVA ( $F_{ST} = 0,010$ ,  $F_{IS} = 0,024$  e  $F_{IT} = 0,034$ ;  $p < 0,05$ ) evidenciaram uma estruturação genética entre as populações amostradas, sendo que somente 1,02% da variação genética pode ser explicada pelas diferenças encontradas entre as populações. Na AMOVA realizada com as populações divididas em dois grupos (Brasil e área nativa), 0,64% da variação genética estava contida dentro dos grupos e 0,67% entre os grupos de populações das áreas nativa

e colonizada ( $F_{IS} = 0,031$ ,  $F_{SC} = 0,006$ ,  $F_{CT} = 0,007$ ,  $F_{IT} = 0,004$ ;  $p < 0,05$ ). Na SAMOVA, conforme mais grupos foram assumidos *a priori*, maiores foram as proporções de variação genética explicada pelas diferenças entre os grupos em detrimento das proporções explicadas pelas diferenças entre as populações dentro dos grupos (Tabela 11). A população de FN e PT foram as populações mais distantes geneticamente entre os grupos das populações amostradas no Brasil e na área nativa, respectivamente (Tabela 11). A partir dos valores de  $F_{ST}$  par-a-par calculados (Tabela 12), identificou-se que: a) no Brasil, as populações amostradas no nordeste brasileiro (FN e PE) exibiram valores menores de  $F_{ST}$  significativos nas comparações com as populações africanas de GW, SN, ZA e NG, b) valores menores de  $F_{ST}$  foram encontrados entre as populações da área nativa do que entre as populações brasileiras, c) SN não diferiu significativamente com nenhuma outra população amostrada na área nativa e d) PT foi a população que apresentou valores maiores de  $F_{ST}$  significativos tanto nas comparações com as populações brasileiras quanto nas comparações com as populações da área nativa (Tabela 12).

**Tabela 11** - Distribuição da variação genética em grupos de populações inferidos por uma análise espacial de variância molecular (SAMOVA) a partir da informação de 2.751 SNPs. Os percentuais e índices de fixação da variação genética explicada pelas diferenças entre os grupos inferidos ( $G$  e  $F_{CT}$ ), entre as populações dentro dos grupos ( $P$  e  $F_{SC}$ ) e entre os indivíduos dentro das populações ( $I$  e  $F_{ST}$ ) foram calculados para os grupos de populações inferidos.

Nº de grupos	$G$ (%)	$P$ (%)	$I$ (%)	$F_{CT}$	$F_{SC}$	$F_{ST}$	Grupos inferidos
2	1,14	1,21	97,64	0,011	0,012	0,024	1 (PA, PE, SP e RS) 2 (PT, GW, SN, GH, NG, ZA e FN)
3	1,40	0,92	97,67	0,014	0,009	0,023	1 (PA, PE, SP e RS) 2 (PT, GW, SN, GH, NG e ZA) 3 (FN)
4	1,52	0,77	97,71	0,015	0,008	0,023	1 (PA, PE, SP e RS) 2 (GW, SN, GH, NG e ZA) 3 (FN) 4 (PT)
5	1,61	0,57	97,81	0,016	0,006	0,022	1 (PA e PE) 2 (SP e RS) 3 (GW, SN, GH, NG e ZA) 4 (FN) 5 (PT)

Fonte: Produção do próprio autor.

A PCA e a DAPC indicaram a existência de dois grupos genéticos majoritariamente representativos entre as populações de *Bubulcus ibis* amostradas (Figuras 11, 12, 13 e Tabela 13). Os grupos genéticos identificados pela PCA foram concordantes nos três pares de componentes principais avaliados, dividindo em dois

grupos principais as amostras das populações brasileiras e das populações da área nativa (Figuras 11A e B, 12A e B e 13A e B). Os indivíduos da população de FN (população 8) ocuparam uma posição intermediária no plano cartesiano, entre os dois grupos de populações identificados (Figuras 11B e 12B). Dez indivíduos mostraram-se geneticamente mais distantes dos demais indivíduos no plano cartesiano: um de GW, um da NG, um do PA, um do PE, dois de GH e quatro da ZA (Figuras 11B, 12B e 13B).

**Tabela 12** - Diferenciação genética determinada entre as populações de garça-vaqueira (*Bubulcus ibis*) amostradas a partir da informação de 2.751 SNPs. Valores de  $F_{ST}$  (diagonal superior) e de significância ( $p$ ) correspondentes (diagonal inferior) foram calculados para um conjunto de 11 populações ( $n = 174$ ): Portugal (PT,  $n = 9$ ), Guiné Bissau (GW,  $n = 17$ ), Senegal (SN,  $n = 16$ ), Gana (GH,  $n = 16$ ), Nigéria (ZA,  $n = 15$ ) e África do Sul (ZA,  $n = 17$ ) na área nativa, e Pará (PA,  $n = 17$ ), Fernando de Noronha (FN,  $n = 17$ ), Pernambuco (PE,  $n = 17$ ), São Paulo (SP,  $n = 17$ ) e Rio Grande do Sul (RS,  $n = 16$ ) no Brasil.

	PT	GW	SN	GH	NG	ZA	PA	FN	PE	SP	RS
PT		0.013	0.007	0.019	0.012	0.017	0.024	0.025	0.016	0.020	0.026
GW	0.013		0.004	0.008	0.007	0.010	0.024	0.016	0.010	0.009	0.018
SN	0.060	0.104		0.000	0.004	0.000	0.014	0.009	0.008	0.012	0.011
GH	0.007	0.025	0.645		0.002	0.002	0.025	0.016	0.012	0.020	0.014
NG	0.003	0.013	0.077	0.384		0.001	0.017	0.013	0.009	0.011	0.011
ZA	0.002	0.003	0.679	0.361	0.415		0.014	0.010	0.013	0.014	0.014
PA	0.000	0.000	0.000	0.000	0.000	0.000		0.020	0.006	0.007	0.006
FN	0.000	0.000	0.002	0.000	0.000	0.001	0.000		0.014	0.015	0.018
PE	0.000	0.001	0.003	0.003	0.004	0.000	0.050	0.000		0.007	0.008
SP	0.000	0.001	0.000	0.000	0.000	0.000	0.008	0.000	0.004		0.004
RS	0.000	0.000	0.000	0.006	0.000	0.000	0.007	0.000	0.003	0.056	

Legenda: Foram destacados em vermelho os valores de  $F_{ST} > 0,010$  e de  $p$  significativos ( $\alpha < 0,05$ ).

Fonte: Produção do próprio autor.

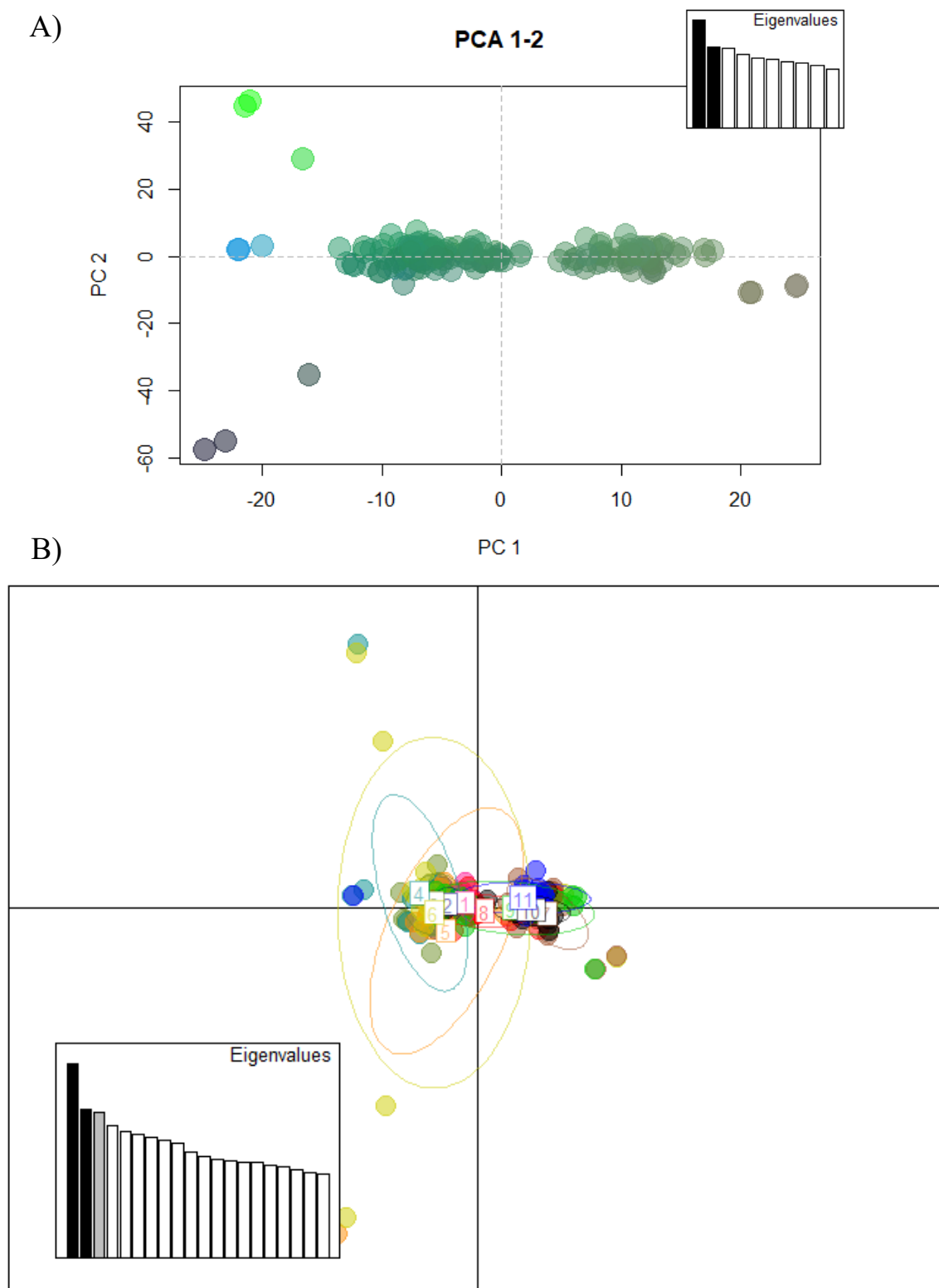
Os dois grupos genéticos inferidos pela DAPC ( $K = 2$ ,  $BIC = 819,2253$ ), grupo 1 ( $n = 68$ ) e grupo 2 ( $n = 106$ ), foram majoritariamente compostos por indivíduos amostrados no Brasil e na área nativa de *Bubulcus ibis*, respectivamente (Tabela 13). No total, 26,50% dos indivíduos amostrados no Brasil foram atribuídos ao grupo 2, majoritariamente composto por indivíduos amostrados na área nativa. As populações de FN e PE, no nordeste brasileiro, contribuíram com 90,91% (20 indivíduos) dos 26,19% (22 indivíduos) atribuídos ao grupo das populações amostradas na área nativa. Em contrapartida, somente 6,67% dos indivíduos amostrados na área nativa foram atribuídos ao grupo 1, majoritariamente composto por indivíduos brasileiros (Tabela 13).

As análises de clusterização Bayesiana realizadas no programa STRUCTURE detectaram estruturação genética entre os grupos de populações brasileiras e da área nativa (Figuras 14A e 14B), e a ausência de estruturação genética entre as populações amostradas dentro de cada um desses grupos (Figura 14C e 14D). A partir da informação

dos 2.751 SNPs foram inferidos dois grupos genéticos ( $K = 2$ , Figura 14A) e a partir da informação de 6.043 haplótipos foram inferidos seis grupos genéticos ( $K = 6$ , Figura 14B). Em ambos os resultados (Figuras 14A e 14B), um grupo genético foi mais representativo entre as populações brasileiras e outro entre as populações da área nativa. As populações amostradas no nordeste brasileiro (FN e PE) apresentaram proporções maiores de ancestralidade genética com o grupo mais representativo entre as populações da área nativa (Figura 14A e 14B). As proporções de ancestralidade de alguns indivíduos amostrados nas populações de NG e ZA também sugerem uma conectividade entre essas duas populações e as populações da área colonizada (Figura 14A, 14B e 14C).

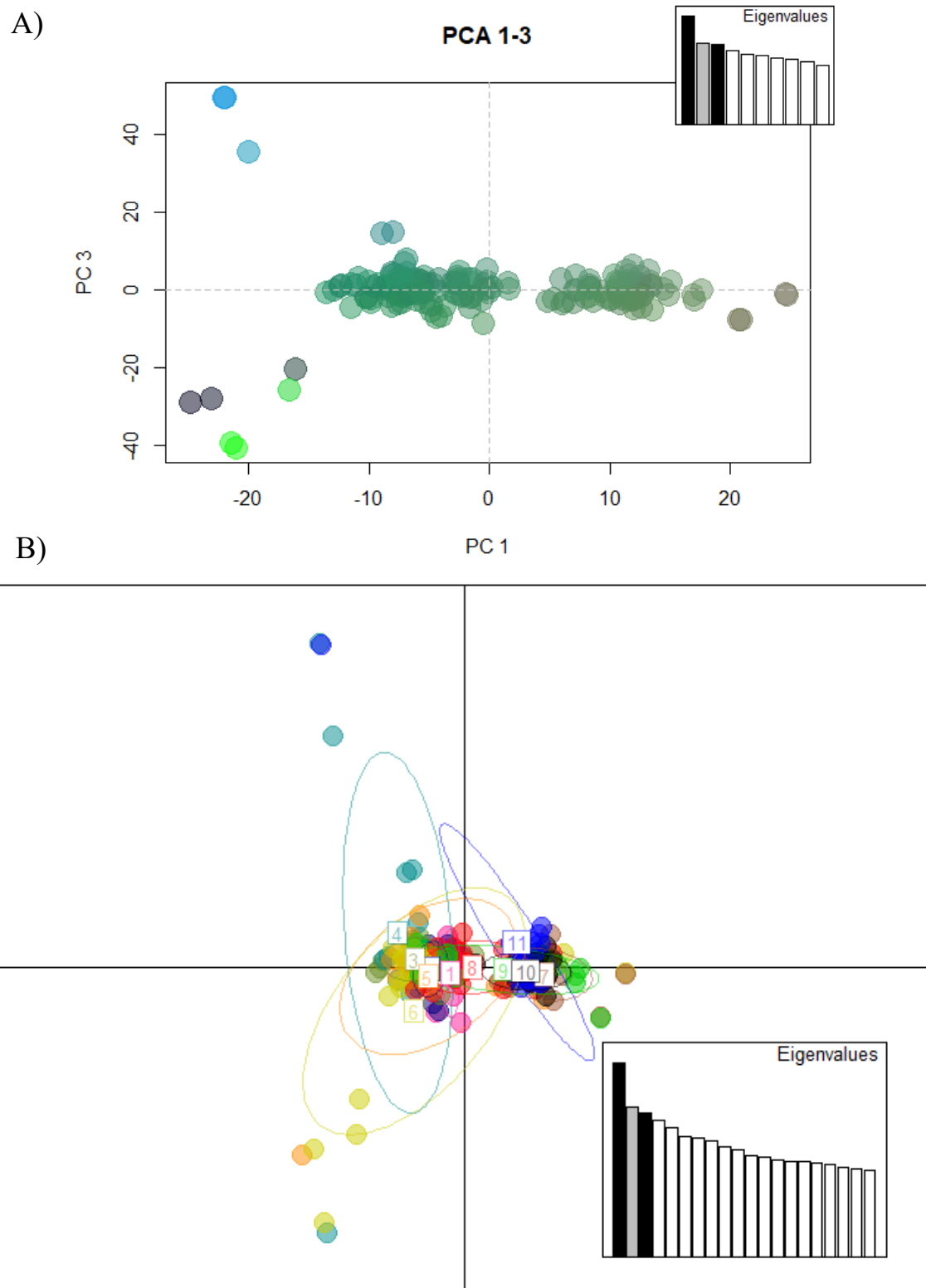


**Figura 11** - Coordenadas dos indivíduos amostrados de garça-vaqueira (*Bubulcus ibis*) ao longo de dois eixos (1 e 2) de variação genética determinados pela análise de componentes principais usando os dados de 2.751 SNPs. Em A) além da disposição dos indivíduos no plano cartesiano, a distância genética entre eles também é mostrada pelas mudanças de cores e em B) cada o esquema de cores e a numeração indica cada uma das 11 populações amostradas ( $n = 174$ ): Portugal (PT,  $n = 9$ ), Guiné Bissau (GW,  $n = 17$ ), Senegal (SN,  $n = 16$ ), Gana (GH,  $n = 16$ ), Nigéria (ZA,  $n = 15$ ) e África do Sul (ZA,  $n = 17$ ) na área nativa, e Pará (PA,  $n = 17$ ), Fernando de Noronha (FN,  $n = 17$ ), Pernambuco (PE,  $n = 17$ ), São Paulo (SP,  $n = 17$ ) e Rio Grande do Sul (RS,  $n = 16$ ) no Brasil.

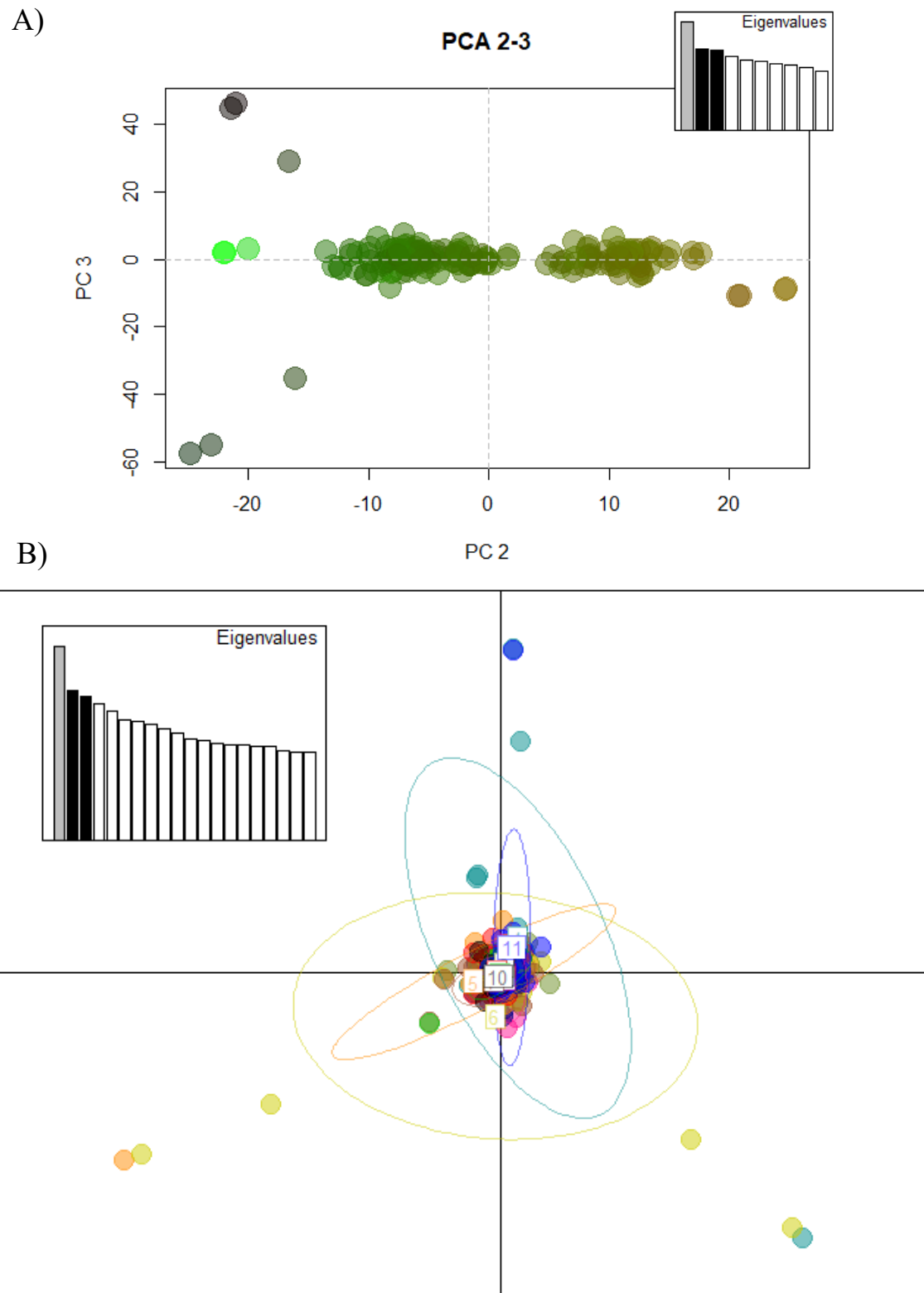


Fonte: Produção do próprio autor.

**Figura 12** - Coordenadas dos indivíduos amostrados de garça-vaqueira (*Bubulcus ibis*) ao longo de dois eixos (1 e 3) de variação genética determinados pela análise de componentes principais usando os dados de 2.751 SNPs. Em A) além da disposição dos indivíduos no plano cartesiano, a distância genética entre eles também é mostrada pelas mudanças de cores e em B) cada o esquema de cores e a numeração indica cada uma das 11 populações amostradas ( $n = 174$ ): Portugal (PT,  $n = 9$ ), Guiné Bissau (GW,  $n = 17$ ), Senegal (SN,  $n = 16$ ), Gana (GH,  $n = 16$ ), Nigéria (ZA,  $n = 15$ ) e África do Sul (ZA,  $n = 17$ ) na área nativa, e Pará (PA,  $n = 17$ ), Fernando de Noronha (FN,  $n = 17$ ), Pernambuco (PE,  $n = 17$ ), São Paulo (SP,  $n = 17$ ) e Rio Grande do Sul (RS,  $n = 16$ ) no Brasil.



**Figura 13** - Coordenadas dos indivíduos amostrados de garça-vaqueira (*Bubulcus ibis*) ao longo de dois eixos (2 e 3) de variação genética determinados pela análise de componentes principais usando os dados de 2.751 SNPs. Em A) além da disposição dos indivíduos no plano cartesiano, a distância genética entre eles também é mostrada pelas mudanças de cores e em B) cada o esquema de cores e a numeração indica cada uma das 11 populações amostradas ( $n = 174$ ): Portugal (PT,  $n = 9$ ), Guiné Bissau (GW,  $n = 17$ ), Senegal (SN,  $n = 16$ ), Gana (GH,  $n = 16$ ), Nigéria (ZA,  $n = 15$ ) e África do Sul (ZA,  $n = 17$ ) na área nativa, e Pará (PA,  $n = 17$ ), Fernando de Noronha (FN,  $n = 17$ ), Pernambuco (PE,  $n = 17$ ), São Paulo (SP,  $n = 17$ ) e Rio Grande do Sul (RS,  $n = 16$ ) no Brasil.



**Tabela 13** - Atribuição dos indivíduos amostrados de garça-vaqueira (*Bubulcus ibis*) aos grupos genéticos inferidos pela análise discriminante de componentes principais. Os números de indivíduos atribuídos a cada um dos grupos genéticos inferidos são mostrados para cada uma das 11 populações de *Bubulcus ibis* amostradas: Portugal (PT), Guiné Bissau (GW), Senegal (SN), Gana (GH), Nigéria (ZA) e África do Sul (ZA) na área nativa, e Pará (PA), Fernando de Noronha (FN), Pernambuco (PE), São Paulo (SP) e Rio Grande do Sul (RS) no Brasil.

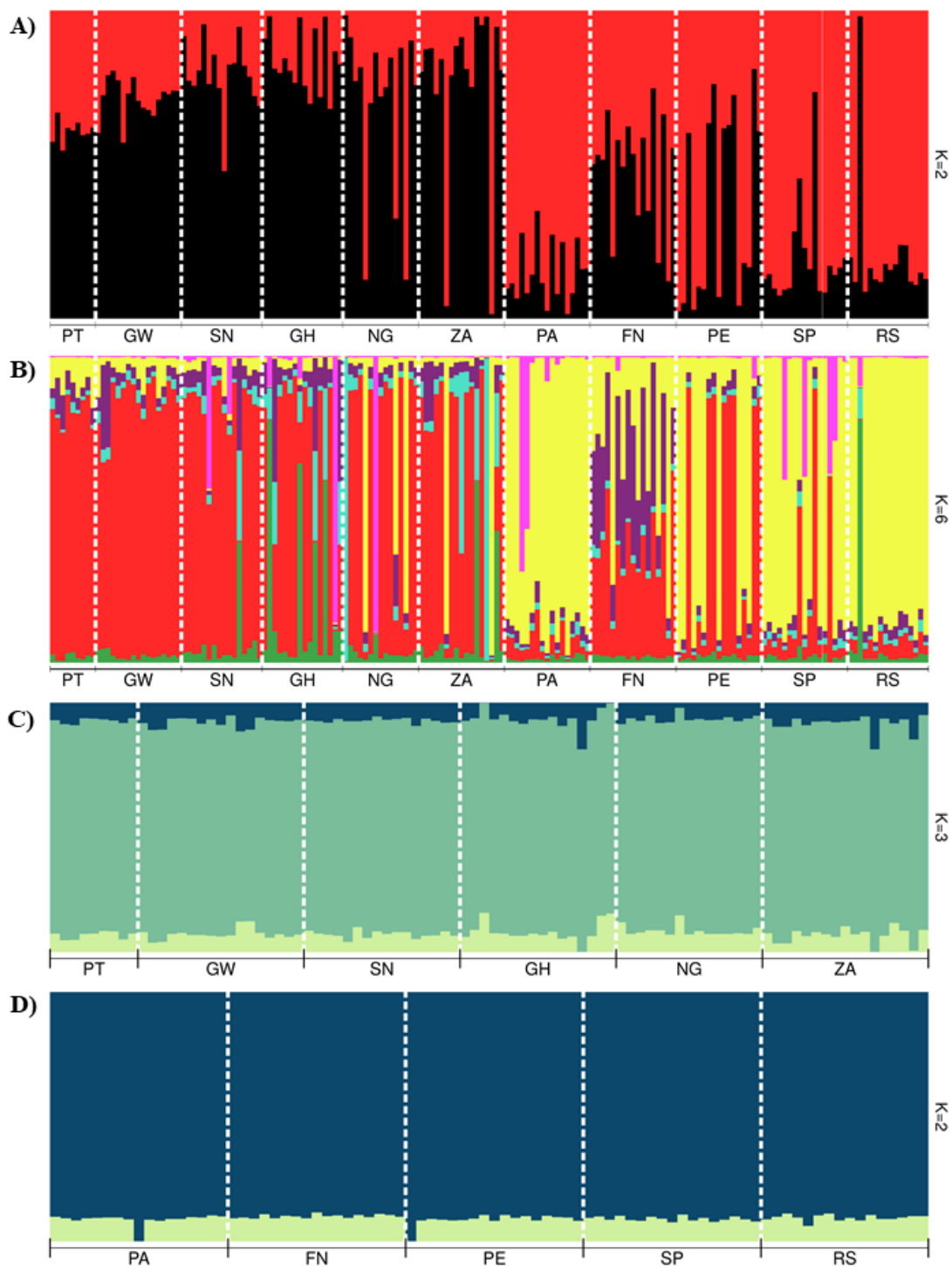
População	Grupo 1	Grupo 2	<i>n</i>
PT	0	9	9
GW	0	17	17
SN	1	15	16
GH	0	16	16
NG	3	12	15
ZA	2	15	17
PA	17	0	17
FN	5	12	17
PE	9	8	17
SP	16	1	17
RS	15	1	16
Total	68	106	174

Fonte: Produção do próprio autor.

As análises de estruturação genética realizadas nos programas Admixture e BAPs também identificaram dois grupos genéticos distribuídos de modo similar ao descrito anteriormente para os resultados obtido usando o programa STRUCTURE (Figuras 15A e 15B). No programa BAPS, além das populações de FN e PE que compartilharam proporções de ancestralidade com o grupo das populações da área nativa, a população da NG apresentou proporções de ancestralidade com o grupo genético inferido para o Brasil (Figura 15B). A função *snmf* indicou que dez era o número mais provável de agrupamentos genéticos ( $K = 10$ ), mas os resultados mantiveram o padrão de um grupo genético mais representativo para o Brasil e outro para a área nativa (Figura 15C).

A análise espacial Bayesiana realizada no programa no pacote tess3r mostrou um número maior de agrupamentos genéticos quando a posição geográfica das populações foi utilizada como *prior* ( $K = 7$ , Figura 16). Entre as áreas nativa e colonizada da distribuição de *Bubulcus ibis*, as proporções de ancestralidade genética inferidas pela análise espacial mostraram um fluxo gênico mais provável entre a população de PE e as populações de GW e SN, e uma conectividade maior entre quatro duplas de populações (SP e RS, PA e PE, GW e SN, e NG e ZA) (Figura 16). Nessa análise, a população de PT foi a que apresentou o maior número de diferenças genéticas entre o grupo das populações da área nativa, enquanto a população de FN foi a que mais diferiu entre as populações brasileiras (Figura 16).

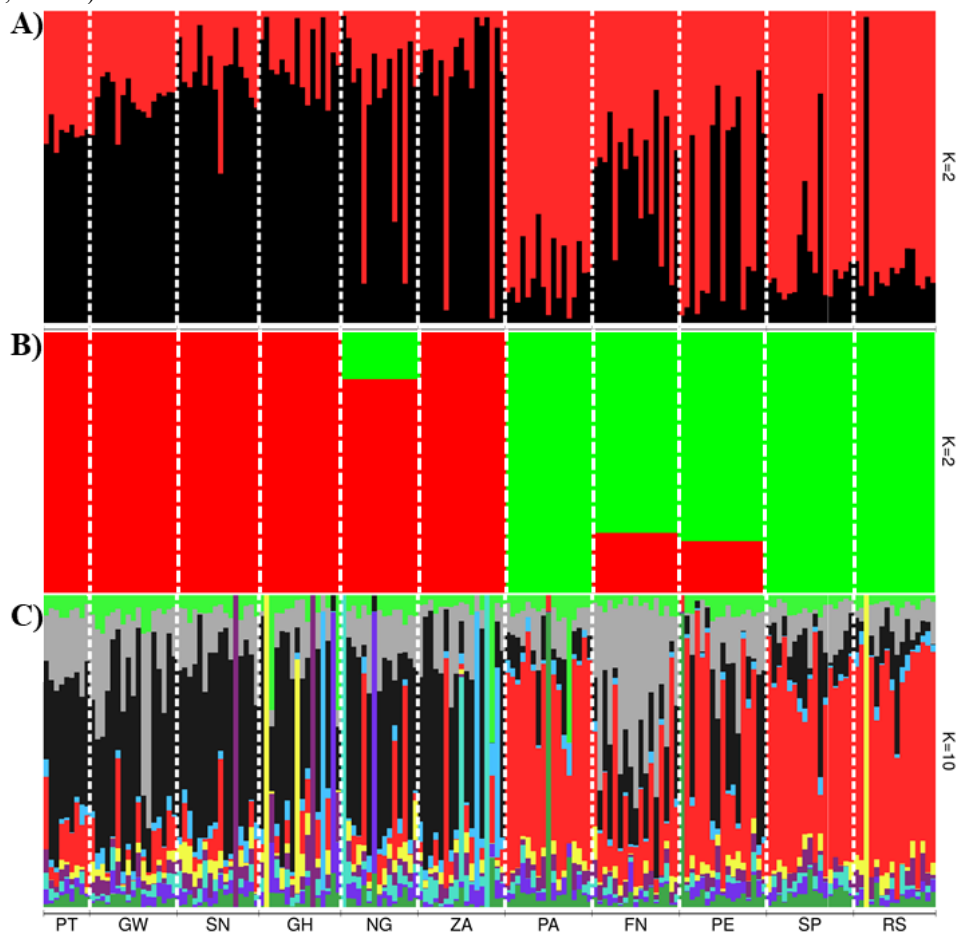
**Figura 14** - Estrutura genética inferida para as populações de garça-vaqueira (*Bubulcus ibis*) a partir de dados genômicos. Os gráficos mostram as proporções de ancestralidade genética estimadas no programa STRUCTURE a partir da informação de A) 2.751 SNPs ( $K = 2$ ) e B) 6.043 haplótipos ( $K = 2$ ), para as 11 populações amostradas ( $n = 174$ ): Portugal (PT,  $n = 9$ ), Guiné Bissau (GW,  $n = 17$ ), Senegal (SN,  $n = 16$ ), Gana (GH,  $n = 16$ ), Nigéria (ZA,  $n = 15$ ) e África do Sul (ZA,  $n = 17$ ) na área nativa, e Pará (PA,  $n = 17$ ), Fernando de Noronha (FN,  $n = 17$ ), Pernambuco (PE,  $n = 17$ ), São Paulo (SP,  $n = 17$ ) e Rio Grande do Sul (RS,  $n = 16$ ) no Brasil.



Legenda: A divisão entre as populações é indicada por uma linha branca tracejada e cada barra vertical corresponde às proporções de ancestralidade dos grupos genéticos, caracterizados por cores diferentes, estimados para cada indivíduo amostrado.

Fonte: Produção do próprio autor.

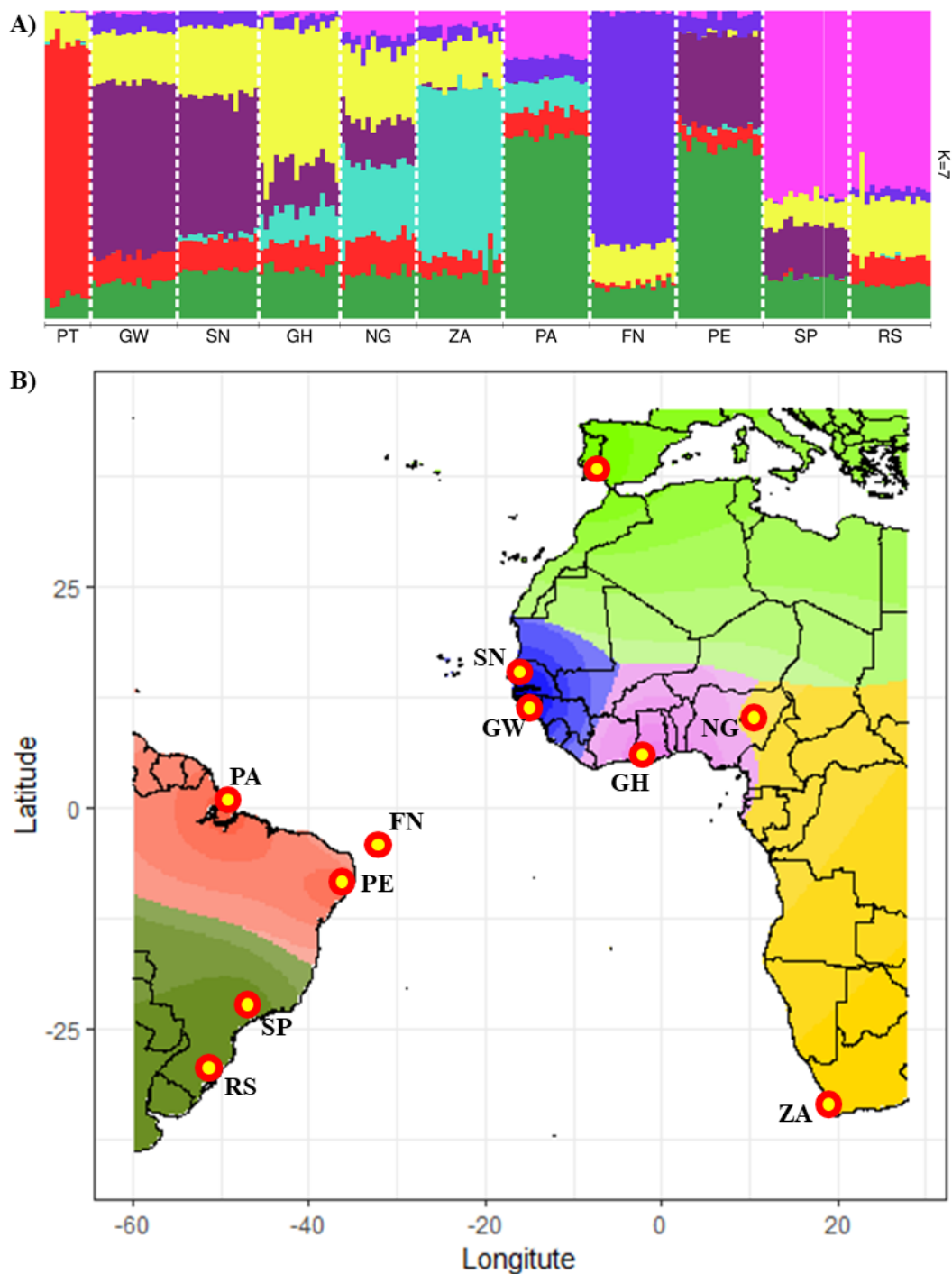
**Figura 15** - Estrutura genética inferida para as populações de garça-vaqueira (*Bubulcus ibis*) a partir de 2.751 SNPs. Os gráficos mostram as proporções de ancestralidade genética estimadas nos programas A) Admixture ( $K = 2$ ) e B) BAPS ( $K = 2$ ), e C) usando a função *snmf* do pacote LEA do R ( $K = 10$ ), para as 11 populações amostradas ( $n = 174$ ): Portugal (PT,  $n = 9$ ), Guiné Bissau (GW,  $n = 17$ ), Senegal (SN,  $n = 16$ ), Gana (GH,  $n = 16$ ), Nigéria (ZA,  $n = 15$ ) e África do Sul (ZA,  $n = 17$ ) na área nativa, e Pará (PA,  $n = 17$ ), Fernando de Noronha (FN,  $n = 17$ ), Pernambuco (PE,  $n = 17$ ), São Paulo (SP,  $n = 17$ ) e Rio Grande do Sul (RS,  $n = 16$ ) no Brasil.



Legenda: A divisão entre as populações é indicada por uma linha branca tracejada e cada barra vertical corresponde às proporções de ancestralidade dos grupos genéticos, caracterizados por cores diferentes, estimados para cada indivíduo amostrado.

Fonte: Produção do próprio autor.

**Figura 16** - Estrutura genética inferida para as populações de garça-vaqueira (*Bubulcus ibis*) a partir de 2.751 SNPs, incorporando a localização geográfica das populações. A partir da análise realizada na aplicação tess3r, são mostradas A) as proporções de ancestralidade genética ( $K = 7$ ) e B) a distribuição espacial dos grupos genéticos inferidos para as 11 populações amostradas ( $n = 174$ ): Portugal (PT,  $n = 9$ ), Guiné Bissau (GW,  $n = 17$ ), Senegal (SN,  $n = 16$ ), Gana (GH,  $n = 16$ ), Nigéria (ZA,  $n = 15$ ) e África do Sul (ZA,  $n = 17$ ) na área nativa, e Pará (PA,  $n = 17$ ), Fernando de Noronha (FN,  $n = 17$ ), Pernambuco (PE,  $n = 17$ ), São Paulo (SP,  $n = 17$ ) e Rio Grande do Sul (RS,  $n = 16$ ) no Brasil.

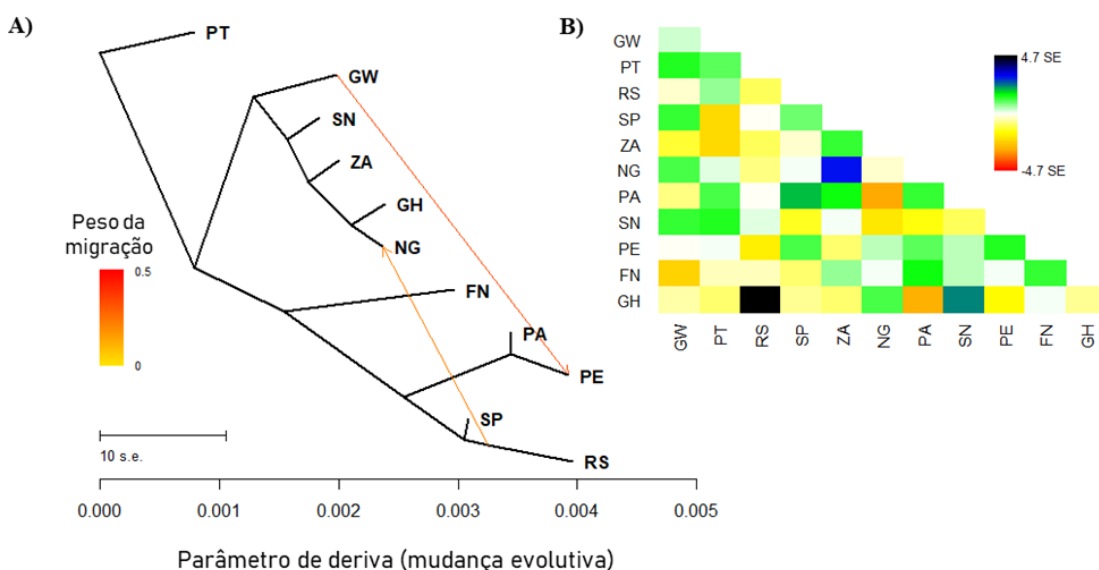


Legenda: Em A) a divisão entre as populações é indicada por uma linha branca tracejada e cada barra vertical corresponde às proporções de ancestralidade dos sete grupos genéticos, caracterizados por cores diferentes, estimados para cada indivíduo amostrado.

Fonte: Produção do próprio autor.

As análises de divisão e mistura populacional realizada no programa Treemix determinaram que o modelo com dois eventos de migração era o mais provável dentre os modelos testados (Figura 17). O filograma de ancestralidade inferido separou as populações brasileiras e da área nativa, colocando a população de FN entre o grupo das outras populações brasileiras e o grupo das populações africanas e separando a população de PT do restante das populações da área nativa (Figura 17). Segundo o melhor modelo recuperado na análise, o evento de migração mais provável entre as populações amostradas supôs indivíduos partindo da população de GW com destino à população de PE (Figura 17). Uma segunda rota com menor peso de migração foi indicada, pela qual indivíduos partiriam da população do RS com destino à população africana da NG (Figura 17).

**Figura 17** - Padrão de divisão e mistura histórica estimado entre as populações de garça-vaqueira (*Bubulcus ibis*) a partir da informação de 2.751 SNPs. Em A) é exibido um filograma de máxima verossimilhança com as relações de ancestralidade inferidas no programa Treemix para o modelo com dois eventos de migração e em B) é exibido o erro padrão médio da covariância residual do filograma distribuído para cada par das 11 populações amostradas ( $n = 174$ ): Portugal (PT,  $n = 9$ ), Guiné Bissau (GW,  $n = 17$ ), Senegal (SN,  $n = 16$ ), Gana (GH,  $n = 16$ ), Nigéria (ZA,  $n = 15$ ) e África do Sul (ZA,  $n = 17$ ) na área nativa, e Pará (PA,  $n = 17$ ), Fernando de Noronha (FN,  $n = 17$ ), Pernambuco (PE,  $n = 17$ ), São Paulo (SP,  $n = 17$ ) e Rio Grande do Sul (RS,  $n = 16$ ) no Brasil.



Legenda: Os comprimentos dos ramos são proporcionais à mudança evolutiva (uma medida da deriva genética). As setas que indicam o sentido de cada evento de migração foram coloridas de acordo com o peso da migração (a porcentagem de ancestralidade recebida da população de origem).

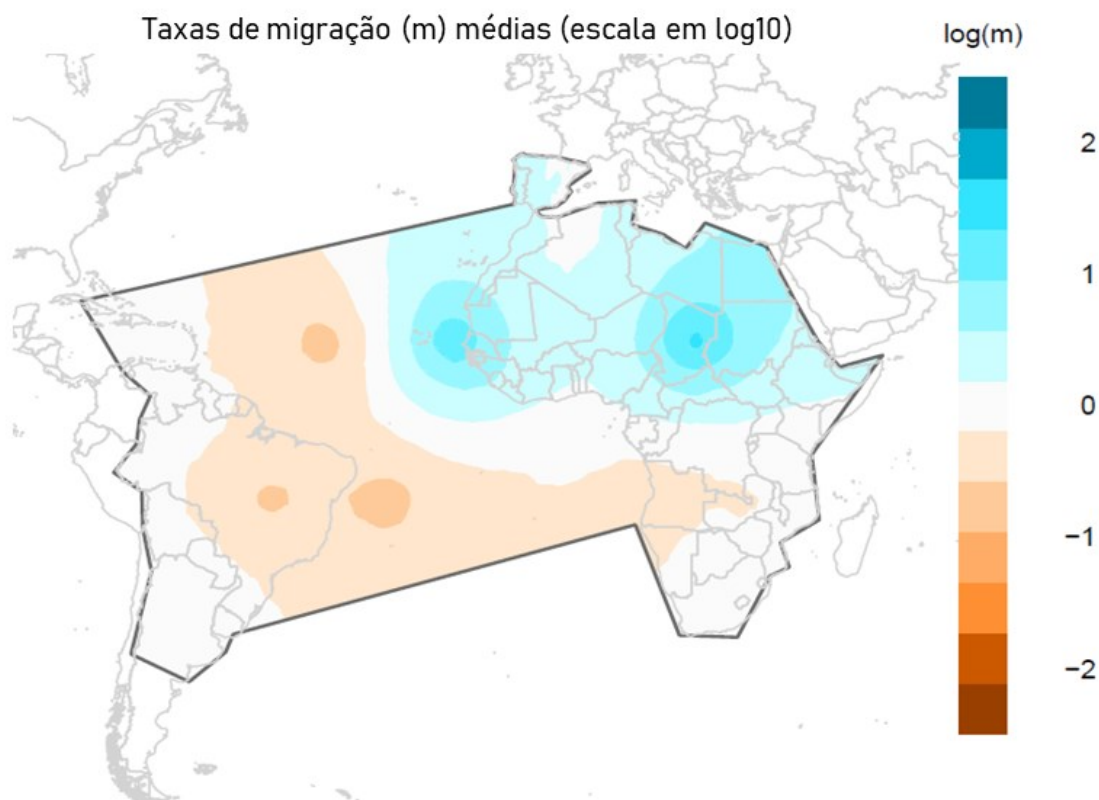
Fonte: Produção do próprio autor.

Na modelagem realizada no programa EEMS não foram detectadas superfícies efetivas de migração entre a área nativa e o Brasil, mas foi detectada uma resistência ao fluxo gênico entre as populações brasileiras e a população da ZA (Figura 18). Uma superfície efetiva de migração foi determinada entre as populações do centro-oeste



africano e Portugal, com maior intensidade de migração entre as populações de GW e SN, e a região da população da NG (Figura 18).

**Figura 18** - Superfícies de migração efetiva estimadas entre as populações de garça-vaqueira (*Bubulcus ibis*) a partir da informação de 2.751 SNPs e a localização geográfica das populações. As taxas de migração médias foram modeladas no programa EEMS entre as 11 populações amostradas ( $n = 174$ ): Portugal (PT,  $n = 9$ ), Guiné Bissau (GW,  $n = 17$ ), Senegal (SN,  $n = 16$ ), Gana (GH,  $n = 16$ ), Nigéria (ZA,  $n = 15$ ) e África do Sul (ZA,  $n = 17$ ) na área nativa, e Pará (PA,  $n = 17$ ), Fernando de Noronha (FN,  $n = 17$ ), Pernambuco (PE,  $n = 17$ ), São Paulo (SP,  $n = 17$ ) e Rio Grande do Sul (RS,  $n = 16$ ) no Brasil.



Legenda: No modelo assumido, a migração efetiva é baixa em regiões onde a similaridade genética decai rapidamente. A escala de cores indica a ocorrência e a intensidade de migração: marrom para potenciais barreiras ao fluxo gênico e azul para probabilidade de fluxo gênico.

Fonte: Produção do próprio autor.

Pela análise de atribuição populacional, 76,67% dos indivíduos amostrados na área nativa foram atribuídos a uma população dentro dessa área, enquanto 83,33% dos indivíduos amostrados a partir de uma população brasileira foram atribuídos a alguma população no Brasil (Tabela 14). Entre as populações da área nativa, a população de SN foi a que apresentou o menor percentual de indivíduos atribuídos dentro da área nativa: 88,89% (PT), 82,35% (GW), 68,75% (SN), 81,25% (GH), 73,33% (NG) e 70,59% (ZA) (Tabela 14). No Brasil, a população do PE apresentou o menor percentual de indivíduos atribuídos dentro do país: 94,12% (PA), 82,35% (FN), 64,71% (PE), 88,23% (SP) e 87,5% (RS) (Tabela 14). Com exceção das populações de FN e do PA, os percentuais de indivíduos atribuídos

corretamente à sua população de origem foram baixos: 22,22% (PT), 47,05% (GW), 12,50% (SN), 31,25% (GH), 15% (NG), 11,76% (ZA), 58,82% (PA), 76,47% (FN), 29,41% (PE), 35,30% (SP) e 50% (RS) (Tabela 14).

**Tabela 14** - Atribuição populacional dos indivíduos de garça-vaqueira (*Bubulcus ibis*) realizada a partir da informação de 2.751 SNPS. Com base nas frequências alélicas calculadas para cada uma das 11 populações amostradas (n = 174): Portugal (PT), Guiné Bissau (GW), Senegal (SN), Gana (GH), Nigéria (NG) e África do Sul (ZA) na área nativa, e Pará (PA), Fernando de Noronha (FN), Pernambuco (PE), São Paulo (SP) e Rio Grande do Sul (RS) no Brasil.

Pop	n	PT	GW	SN	GH	NG	ZA	PA	FN	PE	SP	RS
PT	9	2	1	3	0	1	1	0	1	0	0	0
GW	17	0	8	4	0	0	2	0	1	2	0	0
SN	16	0	2	2	5	0	2	0	1	4	0	0
GH	16	0	1	5	5	1	1	0	0	2	0	1
NG	15	0	0	2	2	3	4	1	0	2	0	1
ZA	17	0	0	2	3	5	2	1	3	0	1	0
PA	17	0	0	0	0	0	1	10	1	2	2	1
FN	17	0	0	2	0	1	0	0	13	1	0	0
PE	17	0	2	0	2	1	1	3	2	5	0	1
SP	17	0	0	0	1	0	1	3	1	1	6	4
RS	16	0	1	0	1	0	0	3	0	1	2	8

Fonte: Produção do próprio autor.

#### 4.3 Padrão geográfico e/ou adaptativo na distribuição da variação genética populacional de *Bubulcus ibis*

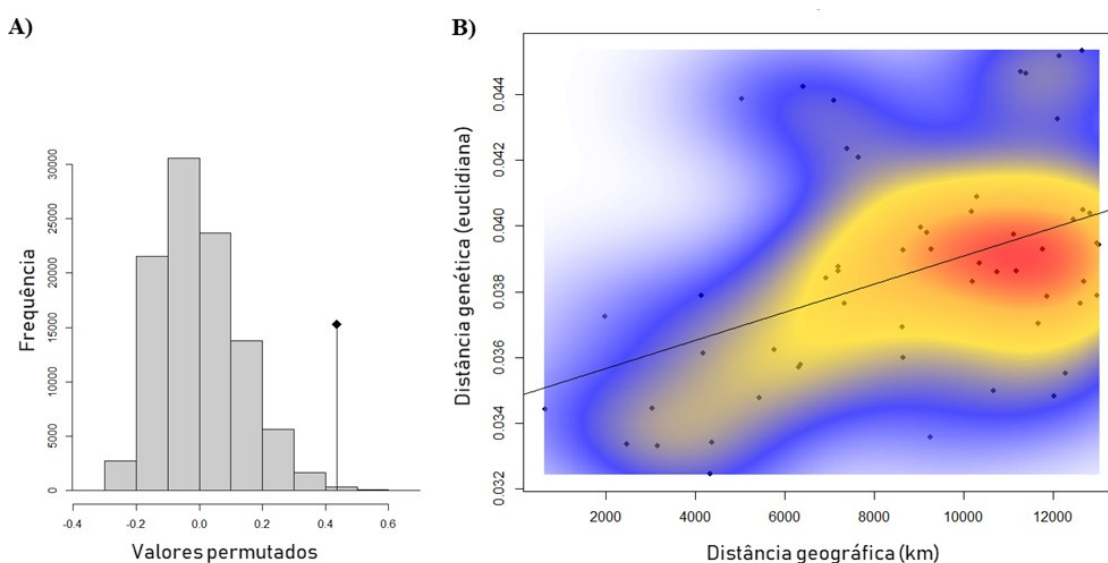
Pelo teste de Mantel, a distribuição da variação genética correspondeu a um modelo de isolamento por distância (Figura 19 A e B). Na análise, encontrou-se uma correlação positiva significativa entre as distâncias genéticas e geográficas das populações de *Bubulcus ibis* amostradas ( $r = 0,44$ ,  $p = 0,0021$ ).

Não foram encontrados SNPs potencialmente sob seleção, após a correção sequencial de Benjamini-Hochberg ( $\alpha = 0,05$ ). A DAPC indicou apenas 11 SNPs com contribuição maior do que 0,02% ao padrão de estruturação genética encontrado, sendo que a maior contribuição encontrada foi de ~0,037% (Figura 20).

Entre as variáveis ambientais selecionadas para identificar SNPs potencialmente correlacionados à heterogeneidade ambiental, a PCA indicou que a temperatura média anual e a incidência média anual de radiação solar eram as variáveis mais informativas (dois primeiros componentes principais). Usando a função *lfgmm* do pacote LEA, 17 SNPs foram apontados como potencialmente associados à temperatura

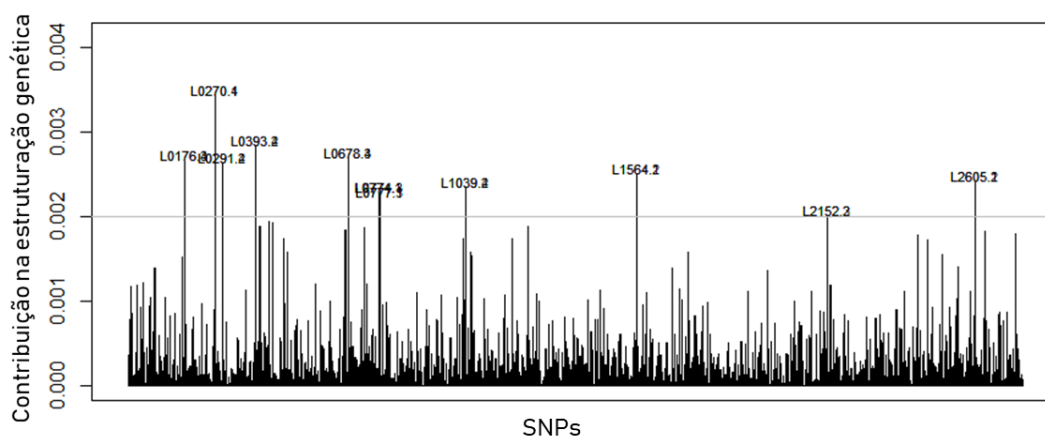
média anual e 22 SNPs foram apontados como potencialmente associados à incidência média de radiação solar ( $p < 0,05$ ) (Figura 21). Entretanto, não houve correspondência entre os SNPs indicados pela DAPC e pela função *lfgmm*. O mapeamento desses SNPs no genoma de referência de *Bubulcus ibis* e na base dados NCBI encontrou 14 compatibilidades (Tabela 15).

**Figura 19** - Padrão de isolamento por distância encontrado para as populações de garça-vaqueira (*Bubulcus ibis*) a partir da informação de 2.751 SNPs. Em A) é mostrada a distribuição dos valores permutados para a hipótese nula de que não haveria um modelo de isolamento por distância e o valor calculado a partir dos dados de SNPs e em B) é mostrada uma estimativa de densidade de núcleo bidimensional para a correlação entre a distância genética euclidiana e a distância geográfica em km ( $r = 0,44$  indicada na linha diagonal,  $p = 0,0021$ ) determinada pelo teste de Mantel.



Fonte: Produção do próprio autor.

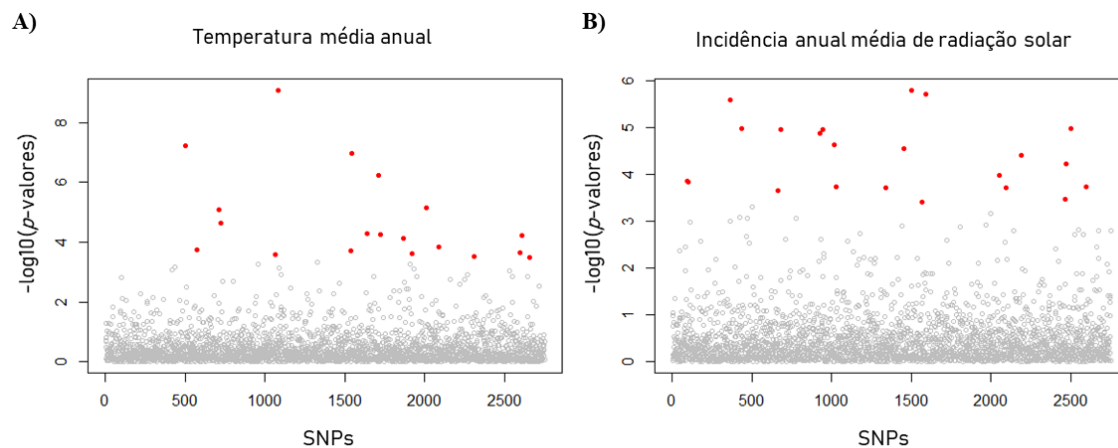
**Figura 20** - SNPs com maior contribuição para o padrão de estruturação genética encontrado entre as populações de garça-vaqueira (*Bubulcus ibis*) amostradas. O resultado refere-se à contribuição de cada um dos 2.751 SNPs ao padrão de estruturação genética ( $K = 2$ ) inferido na análise discriminante de componentes principais (DAPC).



Legenda: Cada barra vertical representa um SNP e a linha na horizontal representa o limite de contribuição estabelecido para identificar os SNPs.

Fonte: Produção do próprio autor.

**Figura 21** - SNPs potencialmente correlacionados a variáveis ambientais das localidades onde as populações de garça-vaqueira (*Bubulcus ibis*) foram amostradas. Os valores de significância (p) para a correlação da variação genética e os valores médios anuais de A) temperatura e B) incidência de radiação solar são mostrados em  $-\log_{10}$  para cada SNP.



Legenda: SNPs, representados por círculos, foram coloridos em vermelho quando apresentaram correlação significativa com as variáveis ambientais.

Fonte: Produção do próprio autor.

**Tabela 15** - SNPs mapeados no genoma de referência da garça-vaqueira (*Bubulcus ibis*) e na base de dados NCBI. Para cada SNP mapeado são mostrados: o locus cuja sequência foi utilizada para o mapeamento, o tamanho da sequência compatível, o percentual de compatibilidade entre as sequências (Hit), o organismo no qual a sequência foi originalmente descrita e a descrição da sequência compatível.

SNP	Locus	Tamanho	Hit (%)	Organismo	Descrição da sequência
SNP_252 <sup>1</sup>	locus_900	305	100	<i>Meleagris gallopavo</i>	MARCH11 Ligase de ubiquitina associada à membrana
SNP_522 <sup>1</sup>	locus_1866	1276	98	<i>Homo sapiens</i>	Proteína lipofuscínose ceróide neuronal 6
SNP_604 <sup>1</sup>	locus_2132	2094	99	<i>Bos taurus</i>	Proteína transmembrana 214
SNP_822 <sup>1</sup>	locus_3112	484	100	<i>Calidris pugnax</i>	Não caracterizado, transcrito X1
SNP_1222 <sup>1</sup>	locus_4658	1274	100	<i>Danio rerio</i>	Regulador negativo da sinalização do receptor do fator de crescimento epidérmico
SNP_4818 <sup>1</sup>	locus_47460	1228	100	<i>Mus musculus</i>	Subunidade reguladora de fosfatase piruvato desidrogenase
SNP_9037 <sup>1</sup>	locus_102432	2677	97	<i>Homo sapiens</i>	Proteína de ligação à GTP-RAD1
SNP_9045 <sup>1</sup>	locus_102432	2677	97	<i>Homo sapiens</i>	Proteína de ligação à GTP-RAD1
SNP_12312 <sup>1</sup>	locus_164287	238	99	<i>Aquila chrysaetos canadensis</i>	Proteína 4 de ligação ao fator de crescimento semelhante à insulina (IGFBP4)
SNP_12316 <sup>2</sup>	locus_164300	637	100	<i>Pandion haliaetus haliaetus</i>	Microssatélite
SNP_13348 <sup>2</sup>	locus_176498	807	100	<i>Anser cygnoides domesticus</i>	Proteína transmembrana 108
SNP_17965 <sup>1</sup>	locus_202375	443	100	<i>Gallus gallus</i>	Homólogo a enzima lisil oxidase
SNP_17977 <sup>1</sup>	locus_202468	1278	100	<i>Zonotrichia albicollis</i>	Não caracterizado, ncRNA
SNP_14589 <sup>3</sup>	locus_182840	877	96	<i>Meleagris gallopavo</i>	Não caracterizado, transcrito X5, ncRNA

Legenda: <sup>1</sup>SNPs identificados pela análise discriminante de componentes principais como maiores contribuidores para o padrão de estruturação genética encontrado, <sup>2</sup>SNPs correlacionados à temperatura média anual e <sup>3</sup>SNPs correlacionados à incidência média anual de radiação solar.

Fonte: Produção do próprio autor.

## 5 DISCUSSÃO

### 5.1 Níveis de diversidade genética das populações amostradas de *Bubulcus ibis*

A hipótese de que níveis maiores de diversidade genética seriam encontrados nas populações de *Bubulcus ibis* da área nativa do que nas populações da área colonizada brasileira foi refutada. Não foram encontradas diferenças significativas entre os índices de diversidade genética dos dois grupos de populações ( $\alpha = 0,05$ , Tabela 10). A diversidade genética também não diferiu significativamente entre essas populações, quando avaliada a partir da informação de 14 *loci* de microssatélites e sequências nucleares e mitocondriais do DNA (CASTILLO, 2013; GERONIMO, 2018; MORALES-SILVA; DEL LAMA, 2014). Essa hipótese foi formulada com base na possibilidade de um efeito fundador na população brasileira durante o processo de colonização e na ausência de evidências que suportassem um fluxo gênico contínuo entre essas áreas. Na ausência de fluxo gênico, uma população fundadora, geralmente, apresenta níveis reduzidos de diversidade genética em comparação às populações fontes e um consequente aumento na divergência genética das populações mais antigas para as mais recentes (ABDELKRIM; PASCAL; SAMADI, 2005; CLEGG et al., 2002). Nessas populações, a menor diversidade genética é consequência do tamanho populacional pequeno e de um aumento correspondente na taxa de deriva genética (NEI; TAJIMA, 1981). Além disso, após o estabelecimento inicial, alguns padrões de expansão populacional podem resultar em novos gargalos populacionais (CORRE; KREMER, 1998; PASCUAL et al., 2001).

Os níveis similares de diversidade genética encontrados entre as populações brasileiras e nativas de *Bubulcus ibis* podem ser explicados pela capacidade de dispersão em longas distâncias da espécie ou pela pressão de propágulo no processo de colonização. É provável que a pressão de propágulo durante o processo de colonização nunca será conhecida, mas múltiplos propágulos, potencialmente, permitiriam que níveis equivalentes de diversidade genética fossem compartilhados entre as populações nativas e colonizadas. Em áreas continentais, grupos com até 75 aves foram observados migrando ao longo da costa do Texas nos EUA durante a migração de outono (TELFAIR, 1983). Considerando que a quantidade de diversidade genética de populações fundadoras pode influenciar o sucesso dos eventos de colonização (CRAWFORD; WHITNEY, 2010; DLUGOSCH; PARKER, 2008), esses dados podem explicar porque a colonização do

Brasil por *Bubulcus ibis* não é mais antiga. Propágulos anteriores com níveis baixos de diversidade genética podem ter falhado no estabelecimento de uma população fundadora (DLUGOSCH; PARKER, 2008).

Nos modelos de colonização sem fluxo gênico subsequente são previstas reduções significativas na diversidade genética e rápida diferenciação entre as populações fonte e fundadora (MAYR, 1954; NEI; MARUYAMA; CHAKRABORTY, 1975), mas nem sempre esse padrão é observado em sistemas naturais (ARDERN et al., 1997). Em alguns casos, são encontrados níveis altos de diversidade genética em populações fundadoras e às vezes, superiores aos das populações fontes têm sido reportados (ARDERN et al., 1997; GENTON; SHYKOFF; GIRAUD, 2005; KOLBE et al., 2004). Nas populações de robins-da-ilha-do-sul, *Petroica australis* (Sparrman, 1788), na Nova Zelândia, foram encontrados níveis moderados de diversidade genética em populações fundadas por apenas algumas aves (ARDERN et al., 1997). Os mecanismos que permitem populações fundadoras apresentarem níveis altos de diversidade genética inclui a mistura de múltiplas populações geográficas e geneticamente distintas como populações fontes e/ou múltiplas introduções independentes (BRODERSEN; LAVERGNE; MOLOFSKY, 2008; DLUGOSCH; PARKER, 2008). A colonização da ilha Daphne Major em Galápagos por pintassilgos, *Geospiza magnirostris* Gould, 1837, ocorreu a partir de pequenas populações de imigrantes (39 indivíduos genotipados), oriundos de até cinco populações-fontes geneticamente distintas (GRANT, P.; GRANT, B.; PETREN, 2001). Em populações de olhos-de-prata, *Zosterops lateralis* (Latham, 1802), na Nova Zelândia, observou-se que após três ou quatro eventos fundadores os níveis de diversidade alélica e divergência genética da população na área colonizada se aproximaram dos determinados para a população fonte (CLEGG et al., 2002).

Outra hipótese que explicaria os níveis similares de diversidade genética entre as populações de *Bubulcus ibis* é a existência de fluxo gênico entre áreas nativa e colonizada. Um fluxo gênico contínuo não parece ser provável, considerando-se o padrão de estruturação genética encontrado (Figuras 9 e 14), porém um fluxo gênico intermitente e/ou de menor intensidade, pode ser compatível com um padrão de estruturação entre o Brasil e a área nativa, com níveis similares de diversidade genética (HANSKI, 1999). Esse cenário também condiz com o comportamento não regular de dispersão em longas distâncias da espécie (TELFAIR, 1994). Quando o fluxo gênico entre populações estabelecidas é raro, as populações podem evoluir de forma independente (WRIGHT, 1982), o que explicaria os níveis de diversidade genética encontrados e o padrão de

estruturação genética. Nichols e Hewitt (1994) simularam as consequências do processo de colonização nas frequências alélicas, mostrando a importância da migração de longa distância e indicando que uma dispersão rara de longa distância poderia produzir introgressão genética entre as populações nativa e colonizada.

## 5.2 *A rota de invasão mais provável durante o processo de colonização do Brasil por *Bubulcus ibis**

A hipótese de que a rota mais provável de invasão de *Bubulcus ibis* no continente americano ocorreu com partidas da região centro-oeste africana e chegadas na região norte da América do Sul foi parcialmente corroborada. Apesar dos resultados indicarem as populações de *Bubulcus ibis* do centro-oeste africano como as fontes mais prováveis, a constituição genética atual das populações brasileiras amostradas, indicaram o Nordeste brasileiro como o ponto de entrada mais provável ao invés do norte. O noroeste do centro-oeste africano (SN e GW) foi suportado como ponto de partida pelas proporções de ancestralidade estimadas com oito *loci* de microssatélites (Figura 9), pelo modelo mais provável de divisão histórica inferido (Figura 17) e pela análise de atribuição populacional (Tabela 14). Castillo (2013) encontrou que um grupo de indivíduos de *Bubulcus ibis* amostrados na República de Cabo Verde, também no noroeste do centro-oeste africano, não diferiu geneticamente das populações brasileiras. Como não há registros de colônias reprodutivas da garça vaqueira nas ilhas de Cabo Verde, o autor considerou que essa região funcione como um local de pouso para indivíduos que estão se deslocando a partir do continente para o oceano (CASTILLO, 2013). Os resultados obtidos pelas análises de clusterização Bayesiana, pela PCA e pela DAPC a partir do conjunto de dados de SNPs indicaram que as duas populações amostradas no nordeste brasileiro (FN e PE) foram geneticamente mais relacionadas às populações da área nativa (Figuras 11, 12, 13, 15A, 15B e 16, Tabela 13). A população de FN foi apontada como ponto de entrada no Brasil pela SAMOVA (2 grupos) e pelas proporções de ancestralidade estimadas por indivíduo a partir de oito *loci* de microssatélites (Figura 9 e Tabela 11). Enquanto que as taxas de migração e os valores de  $F_{ST}$  calculados a partir do conjunto de dados com oito *loci* de microssatélites (Figura 10 e Tabela 10) e as análises de atribuição populacional e de divisão e mistura histórica indicaram a população de PE como a que menos divergiu em relação às populações da área nativa (Figura 17 e Tabela



14).

Esses resultados não se alinham com os primeiros registros históricos de *Bubulcus ibis* no Pará no Brasil (SICK, 1965). Até o momento, todos os trabalhos que investigaram a colonização do Brasil por *Bubulcus ibis* formularam a hipótese de que o norte do Brasil ou o norte da América do Sul seria o ponto de chegada na rota de invasão (CASTILLO, 2013; GERONIMO, 2018; MASSA; DOYLE; FORTUNATO, 2014; MORALES-SILVA; DEL LAMA, 2014). Moralez-Silva e Del Lama (2014) propuseram em análises preliminares de diferenciação genética feitas com sequências do mtDNA que a colonização poderia ter se iniciado na região sul do Brasil. Os dados genéticos produzidos por Castillo (2013) e Geronimo (2018) não apresentaram resolução suficiente para discriminar uma única região na costa brasileira como ponto de chegada durante o processo de colonização. Os resultados obtidos com a informação de 14 *loci* de microssatélites indicaram que tanto a região norte quanto a região nordeste do Brasil poderiam ser o final da rota de invasão de *Bubulcus ibis* (GERONIMO, 2018).

Os dados obtidos neste trabalho dão suporte à teoria descartada por Blaker (1971) e Crosby (1972) de que a rota de invasão de *Bubulcus ibis* no Brasil ocorreu por uma rota de menor distância entre a África e a América do Sul (~2.450 km entre SN e FN, e ~2.790 km entre SN e FN). Esses autores discutiram que, apesar da menor distância entre o centro-oeste africano e o nordeste brasileiro, era improvável que os indivíduos cruzassem o oceano Atlântico neste ponto porque a espécie foi registrada primeiramente nas Guianas e por não existirem ventos favoráveis à travessia entre estes dois pontos (BLAKER, 1971; CROSBY, 1972). Para Blaker (1971) e Crosby (1972) era mais provável que imigrantes saíssem do norte da África ou do sul da Europa (Portugal e Espanha) e utilizassem os ventos alísios para chegarem diretamente nas Guianas, no nordeste da América do Sul. Em um estudo realizado por meio da análise de trajetórias eólicas históricas e dos padrões climáticos associados a elas, o sul da Europa foi descartado já na concepção do estudo por não apresentar ventos de cauda que permitissem a travessia do oceano Atlântico por *Bubulcus ibis* (MASSA; DOYLE; FORTUNATO, 2014). Embora nesse estudo com base nos dados meteorológicos do período entre 1871 e 1920 tenham sido consideradas somente áreas da costa brasileira acima de 5° S de latitude, foi reportado um grande número de rotas de invasão possíveis partindo do centro-oeste da África e atingindo o nordeste brasileiro (MASSA; DOYLE; FORTUNATO, 2014). No total, 304 rotas foram encontradas saindo da África central e alcançando a costa nordeste do Brasil, enquanto três rotas com sucesso foram encontradas partindo do norte

da África e uma partindo do Sul da África (MASSA; DOYLE; FORTUNATO, 2014). Apesar do maior número de rotas com sucesso simuladas a partir dos dados meteorológicos chegarem no nordeste brasileiro, a menor trajetória encontrada partiu do Senegal e chegou no Suriname em 132 h (MASSA; DOYLE; FORTUNATO, 2014). Na possibilidade de múltiplos propágulos durante a colonização das Américas por *Bubulcus ibis*, os dados apresentados aqui mostram que o nordeste brasileiro é a área mais provável de chegada no país, porém é possível a existência de outras rotas de invasão com pontos de chegada em diferentes localidades do continente.

Estudos em espécies de aves marinhas rastreadas por geolocalizadores suportam a travessia transatlântica entre o nordeste brasileiro e o centro-oeste africano (EGEVANG et al., 2010; FELICÍSIMO; MUÑOZ; GONZÁLEZ-SOLIS, 2008; HEDD et al., 2012). Indivíduos da andorinha-do-ártico, *Sterna paradisaea* Pontoppidan, 1763, durante a migração pós-reprodutiva, saíram da Groenlândia e cruzaram o oceano Atlântico entre uma área próxima à costa do centro-oeste africano e a costa do nordeste brasileiro (Paraíba e Pernambuco) (EGEVANG et al., 2010). Um padrão similar de migração foi encontrado para o bobo-escuro, *Puffinus griseus* (Gmelin, 1789), com um indivíduo cruzando o oceano Atlântico entre a costa da Maurîtânia e o Rio Grande do Norte (HEDD et al., 2012). Para o bobo-grande, *Calonectris diomedea* (Scopoli, 1769), foram inferidas as rotas de migração entre o continente africano e as Américas baseadas em modelos de resistência ao movimento aéreo próximo à superfície do oceano (FELICÍSIMO; MUÑOZ; GONZÁLEZ-SOLIS, 2008). Felicísimo, Muñoz e González-Solis (2008) determinaram que o menor custo energético para a travessia entre o oceano seria entre o noroeste do centro-oeste africano (Maurîtânia, Senegal e Gâmbia) e o nordeste brasileiro (Rio Grande do Norte, passando por Fernando de Noronha).

### 5.3 Padrão de estruturação genética entre as populações amostradas de *Bubulcus ibis*

Não foram encontradas evidências que suportassem a hipótese de que, mesmo na ausência de fluxo gênico contínuo, as populações de *Bubulcus ibis* apresentariam uma estruturação genética baixa entre as áreas nativa e colonizada. As primeiras análises de diferenciação genética (AMOVA e  $F_{ST}$  par-a-par) indicaram uma estruturação genética significativa entre as populações de *Bubulcus ibis* amostradas tanto com os dados de microsatélites quanto com os dados de SNPs (Tabelas 9 e 12). Os resultados das análises

de agrupamento populacional (análises Bayesianas, DAPC e PCA) indicaram que a estruturação genética detectada correspondia principalmente à divisão entre as populações das áreas nativa e colonizada (Figuras 9, 11, 12, 13, 14A, 14B e 15; Tabela 13). Esses dados descartaram o norte do Brasil como ponto de entrada na rota de invasão de *Bubulcus ibis* durante a colonização do Brasil e indicaram não só a conectividade entre o nordeste brasileiro (FN e PE) com as populações africanas, mas uma conectividade maior entre o nordeste e o restante da área colonizada. Recentemente, o mesmo padrão de estruturação genética foi encontrado para as populações de *Bubulcus ibis* a partir da análise de 14 *loci* de microssatélites (GERONIMO, 2018). Entretanto, no estudo anterior, o conjunto de genótipos analisado indicou que a população do Pará, assim como as populações amostradas no nordeste brasileiro, possuíam uma maior similaridade genética com as populações africanas (GERONIMO, 2018).

A estrutura genética de uma população é determinada pela interação histórica e contemporânea entre as forças evolutivas de seleção natural, mutação, fluxo gênico e deriva genética (WRIGHT, 1931, 1950). A hipótese sobre a estrutura genética das populações de *Bubulcus ibis* foi formulada em um cenário de colonização seguido de expansão populacional, considerando que menos de 120 gerações se passaram desde a colonização. Nesse cenário, caso o efeito fundador não fosse muito intenso, esperava-se uma estruturação genética baixa até que tivesse passado tempo suficiente para a deriva genética e a seleção natural dissipassem a variação genética ancestral compartilhada (BIRKY; FUERST; MARUYAMA, 1989; SLARKIN, 1985). Uma estruturação genética baixa entre áreas colonizadas e populações fontes também tem sido atribuída a padrões de recolonização inferidos para populações que tiveram uma perda significativa de diversidade genética durante o último máximo glacial (BICKNELL et al., 2012; YANNIC et al., 2016). Para as populações africanas de *Bubulcus ibis*, foi encontrado níveis de diversidade genética no mtDNA inferiores aos reportados para outras espécies de aves adaptadas a habitats áridos e semi-áridos (CONGRAINS et al., 2016; MORALEZ-SILVA; DEL LAMA, 2014). Pelo menos para as populações africanas de *Bubulcus ibis*, a análise de sequências do nDNA e mtDNA revelaram baixos níveis de diferenciação genética, sinais de isolamento por distância e aumentos no tamanho efetivo populacional desde o último máximo glacial (CONGRAINS et al., 2016).

O padrão de estruturação genética encontrado para as populações de *Bubulcus ibis* poderia ser explicado pela atuação da deriva genética e/ou seleção natural na população fundadora. O efeito fundador foi apontado como o fator determinante na

estruturção genética durante os processos de colonização recente da Nova Zelândia pelo olhos-de-prata, *Zosterops lateralis*, e do Alasca (EUA) e da Columbia Britânica (Canadá) pelo pardal-do-noroeste, *Melospiza melodia* (Wilson, 1810) (CLEGG et al., 2002; PRUETT; WINKER, 2005). Entretanto, nesse cenário, a ocorrência de fluxo gênico subsequente seria necessária para nivelar a diversidade genética entre as duas regiões, mantendo a o padrão de estruturação genética encontrado. O grau de conectividade migratória entre populações de aves tem sido apontado como o principal fator determinando os padrões de estruturação genética encontrados (FRIESEN, 2007). É esperado que o fluxo gênico atue contra a deriva genética e a adaptação local, reduzindo a diferenciação populacional, e esse efeito é proporcional a intensidade do fluxo gênico (KAWECKI; EBERT, 2004). Dessa forma, em um modelo de divergência alopátrica, barreiras extrínsecas ao fluxo gênico facilitariam a diferenciação das populações (KAWECKI; EBERT, 2004).

As únicas evidências de subestruturação populacional dentro dos grupos de populações das áreas nativa e colonizadas foram obtidas quando informações *a priori* foram incorporadas na SAMOVA e na análise complementar de agrupamento espacial do programa tess3r com os dados de SNPs (Figura 16, Tabela 11). Por esses resultados, as populações geograficamente mais próximas apresentaram uma menor divergência genética, mostrando proporções de ancestralidade compartilhadas entre pares de populações: SP e RS, PA e PE, GW e SN, NG e ZA (Figura 16, Tabela 11). Esses resultados convergiram com um modelo de isolamento por distância (Figura 19) (GANDON; ROUSSET, 1999; WRIGHT, 1943). As populações de Portugal e Fernando de Noronha foram as que mais divergiram, respectivamente, dentro do grupo de populações da área nativa e do grupo de populações brasileiras (Figura 16, Tabela 11). Embora, esse padrão de estruturação não tenha sido suportado pela análise de clusterização Bayesiana realizada no programa STRUCTURE (Figuras 14C e 14D), ele pode indicar um menor fluxo gênico para essas populações dentro dos grupos de populações amostradas (FRIESEN; BURG; MCCOY, 2007).

O modelo selecionado na análise de divisão e mistura histórica realizada no programa Treemix previu um evento de migração entre as populações do RS e NG (Figura 17). Apesar dessa rota no sentido contrário da colonização ter um peso menor de migração, esse resultado foi suportado pela DAPC e por algumas análises de clusterização que indicaram proporções de ancestralidade do grupo majoritariamente brasileiro não só na população da NG, mas também na população da ZA (Figuras 14A, 14B, 15A e 15B,

Tabela 13). Uma possível rota com maior custo energético foi inferida, por Felicísimo, Muñoz e González-Solis (2008) com base nos padrões de ventos, conectando parte da costa brasileira e a África do Sul. Entretanto, a partir dos dados de SNPs apresentados aqui, detectou-se uma resistência ao fluxo gênico entre as populações brasileiras e a população da ZA (Figura 18). Outra possibilidade é que essas análises detectaram sinais de polimorfismos ancestrais que foram mantidos durante o processo de colonização, uma vez que antes mesmo da expansão de *Bubulcus ibis* no continente africano, a Nigéria fazia parte da distribuição geográfica original da espécie (KUSHLAN; HANCOCK, 2005). Dados do mtDNA também indicaram uma divergência genética baixa entre a população da NG em relação às outras populações africanas, e Castillo (2013) atribuiu essa diferença a posição geográfica da população, uma vez que ela foi amostrada em uma região mais interior do continente. Segundo Castillo (2013), a posição da população da NG no interior do continente africano implicaria na possibilidade de ela estar fora de rotas preferenciais de dispersão pela costa do continente africano.

#### 5.4 Padrão geográfico e/ou adaptativo na distribuição da variação genética populacional de *Bubulcus ibis*

Considerando que a estrutura genética das populações naturais pode ser influenciada pela paisagem via isolamento geográfico e variação ambiental (SEXTON; HANGARTNER; HOFFMANN, 2014; WANG, 2013), duas hipóteses foram formuladas sobre a contribuição de um padrão geográfico e/ou adaptativo na distribuição da variação genética populacional de *Bubulbus ibis*. Foram encontradas evidências de que a colonização do Brasil por *Bubulcus ibis* pode ter ocorrido em um modelo de progressão *stepping stone*, como esperado para espécies de aves coloniais (TIEDEMANN et al., 2004; TIGANO et al., 2017). Entretanto, os resultados das análises de seleção não suportaram a hipótese de que um fator adaptativo estaria contribuindo na determinação da estruturação genética entre as populações de *Bubulcus ibis*.

Os resultados obtidos determinaram um padrão global de isolamento por distância entre as populações amostradas (Figura 19). Esse resultado convergiu com os valores de  $F_{ST}$  par-a-par calculados, com as análises de agrupamento espacial e com o filograma de ancestralidade inferido pelo programa Treemix (Figuras 16 e 17, Tabelas 9, 11 e 12). O padrão de isolamento por distância foi reportado em populações de várias espécies invasoras (CAO et al., 2016; FOLINO-ROREM; DARLING; D'AUSILIO,

2009; ROLLINS et al., 2009). No entanto, na possibilidade de fluxo gênico recorrente dentro dos grupos de populações, o padrão de isolamento por distância encontrado refletiria majoritariamente os grupos genéticos distintos das populações nativas e colonizadas de *Bubulcus ibis*. Além disso, o padrão de isolamento por distância não permite afirmar com certeza que a colonização do Brasil por *Bubulcus ibis* tenha seguido um modelo de progressão *stepping stone*, porém descarta modelos que não preveem que indivíduos migrem com maior frequência entre colônias mais próximas. Dentre os modelos possíveis, o modelo *stepping stone* hierárquico prevê que as taxas de migração entre as populações dentro de grupos podem ser diferentes das taxas de migração entre grupos de populações devido a existência de barreiras ao fluxo gênico (MURPHY et al., 2008). Outros modelos de colonização que se encaixam no padrão de isolamento por distância são os modelos *leading edge* (vanguarda) e *phalanx* (falange). O modelo de colonização *leading edge* foi descartado porque prevê uma diminuição da diversidade genética em populações fundadoras recém-estabelecidas e uma diminuição gradual na variação genética entre as populações encontradas na direção da expansão (HEWITT, 1996, 2000). Enquanto que o modelo *phalanx* prevê uma mistura mais uniforme de populações ancestrais, resultando em uma diminuição marginal na diversidade genética local (HEWITT, 1996, 2000). Embora o modelo de colonização *phalanx* não tenha sido descartado, ele tem sido documentado apenas em populações com expansão lenta, com altas densidades populacionais, alto fluxo gênico e ampla frente de colonização (HEWITT, 1996, 2000).

Nesse estudo, não foram identificados SNPs que individualmente contribuíram de forma desproporcional ao padrão de estruturação genética encontrado. Apesar de terem sido encontrados SNPs potencialmente associados a variáveis ambientais, não foram detectados SNPs significativamente sob seleção (Figuras 20 e 21). A expectativa de encontrar *outliers* (SNPs exibindo frequências alélicas com diferenças extremas entre as populações) era a de identificar indícios de adaptação às condições locais, considerando a heterogeneidade ambiental ao longo da distribuição da espécie (SPURGIN et al., 2014). Mesmo que não necessariamente sob seleção, os *outliers* poderiam ser úteis na identificação de padrões de estruturação genética sutil em populações homogêneas (ALLENDORF; HOHENLOHE; LUIKART, 2010).

## 6 CONCLUSÕES

O conjunto de resultados obtidos foi suficiente para avaliarmos todas as hipóteses levantadas no início deste trabalho sobre a colonização do Brasil pela garça-vaqueira, *Bubulcus ibis*. A hipótese de que a diversidade genética nas populações de *Bubulcus ibis* da área nativa seria maior do que nas populações brasileiras não foi sustentada pelos dados de diversidade genética determinados com os SNPs. Níveis de diversidade genética similares entre as áreas nativa e colonizada indicam que a diversidade genética foi mantida durante o processo de colonização e é um indício de que um número considerável de indivíduos se estabeleceu no Brasil durante a colonização, provavelmente com múltiplas introduções. A estruturação genética com dois grupos genéticos mais representativos, um na área nativa e outro na área colonizada, corrobora a hipótese de que não há um fluxo gênico intenso entre essas duas áreas, mas não descarta a possibilidade de que ainda haja ou houve um fluxo gênico contínuo e de baixa intensidade. Nesse estudo, pela primeira vez a região Norte do Brasil foi descartada como uma região preferencial de entrada da espécie no país. Segundo os dados obtidos, a rota de invasão mais provável teve como ponto de chegada o Nordeste brasileiro e não o Norte do país (onde o primeiro registro foi feito em 1964), com indícios de que as partidas na área nativa ainda ocorram ou ocorreram na região noroeste do centro-oeste africano. Demonstrou-se ainda que o fluxo gênico entre as duas áreas não pode ser muito antigo, devido às evidências de que após a estruturação ocorreram eventos de migração. Não foram encontradas evidências de que haja um fator adaptativo contribuindo para o padrão de estruturação populacional encontrado, descartando que a heterogeneidade ambiental tenha um papel importante na distribuição da variação genética. Por outro lado, o padrão de isolamento por distância encontrado, não só ressalta a importância de fator geográfico na estruturação genética das populações, mas também suporta que a colonização ocorreu por um modelo de progressão *stepping-stone* hierárquico.

#### IV REFERÊNCIAS BIBLIOGRÁFICAS

- ABDELKRIM, J. et al. Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. **BioTechniques**, [s. l.], v. 46, n. 3, p. 185–192, 2009.
- ABDELKRIM, J.; PASCAL, M.; SAMADI, S. Island colonization and founder effects: The invasion of the Guadeloupe islands by ship rats (*Rattus rattus*). **Molecular Ecology**, [s. l.], v. 14, n. 10, p. 2923–2931, 2005.
- ADEFN. **Plano de Manejo do Perigo Aviário com Uso da Falcoaria – 1º Etapa**. Administração do Distrito Estadual de Fernando de Noronha, 2012.
- ADEFN. **Plano de Manejo do Perigo Aviário com Uso da Falcoaria – 10º Etapa**. Administração do Distrito Estadual de Fernando de Noronha, 2015.
- AHMED, R. Subspecific identification and status of Cattle Egret. **Dutch Birding**, [s. l.], v. 33, n. August 2009, p. 294–304, 2011.
- ALERSTAM, T. Wind as Selective Agent in Bird Migration. **Ornis Scandinavica**, [s. l.], v. 10, n. 1, p. 76–93, 1979.
- ALEXANDER, D. H.; NOVEMBRE, J.; LANGE, K. Fast model-based estimation of ancestry in unrelated individuals. **Genome Research**, [s. l.], v. 19, p. 1655–1664, 2009.
- ALI, A. et al. Genetic diversity and population structure analysis of *Saccharum* and *Erianthus* genera using microsatellite (SSR) markers. **Scientific Reports**, [s. l.], v. 9, n. 1, p. 1–10, 2019.
- ALLENDORF, F. W.; HOHENLOHE, P. A.; LUIKART, G. Genomics and the future of conservation genetics. **Nature reviews. Genetics**, [s. l.], v. 11, n. 10, p. 697–709, 2010.
- ALTSCHUL, S. F. et al. Basic local alignment search tool. **Journal of Molecular Biology**, [s. l.], v. 215, n. 3, p. 403–410, 1990.
- ANDREWS, K.; LUIKART, G. Recent novel approaches for population genomics data analysis. **Molecular Ecology**, [s. l.], v. 23, n. 7, p. 1661–1667, 2014.
- ARDERN, S. L. et al. Social and Sexual Monogamy in Translocated New Zealand Robin Populations Detected Using Minisatellite DNA. **The Auk**, [s. l.], v. 114, n. 1, p. 120–126, 1997.
- ARENDRT, W. J. Range Expansion of the Cattle Egret (*Bubulcus ibis*) in the Greater Caribbean Basin. **Colonial Waterbirds**, [s. l.], v. 11, n. 2, p. 252, 1988.
- ARENDRT, W. J. Range Expansion of the Cattle Egret (*Bubulcus ibis*) in the Greater Caribbean Basin. **Colonial Waterbirds**, [s. l.], v. 11, n. 2, p. 252, 2006.
- AVERY, M.; TILLMAN, E. Alien birds in North America – challenges for wildlife managers. In: WILDLIFE DAMAGE MANAGEMENT CONFERENCES - PROCEEDINGS 2005, **Anais...** [s.l.: s.n.]
- BADOUIN-BODIN, J. Des hérons cendrés bagues au lac de Grand-Lieu repris aux Antill. **L'Oiseau et RFO**, n. 30, p. 270, 1960.
- BAIRD, N. A. et al. Rapid SNP Discovery and Genetic Mapping Using Sequenced



- RAD Markers. **PLoS ONE**, [s. l.], v. 3, n. 10, p. e3376, 2008.
- BARBOSA-FILHO, R. C. et al. A garça-vaqueira ( *Bubulcus ibis* Linnaeus , 1758 ) e o atobá-de-pé-vermelho ( *Sula sula* Linnaeus , 1766 ) no Arquipélago de Fernando de Noronha : uma abordagem ecológica comparativa. **Ornithologia**, [s. l.], v. 3, n. 2, p. 101–114, 2009.
- BASHALKHANOV, S.; PANDEY, M.; RAJORA, O. P. A simple method for estimating genetic diversity in large populations from finite sample sizes. [s. l.], v. 10, p. 1–10, 2009.
- BASTÍAS, A. et al. Identification and characterization of microsatellite loci in maqui ( *aristotelia chilensis* [molina] stunz) using next-generation sequencing (ngs). **PLoS ONE**, [s. l.], v. 11, n. 7, p. 1–17, 2016.
- BEAUMONT, M. A.; NICHOLS, R. A. Evaluating loci for use in the genetic analysis of population structure. **Proceedings of the Royal Society B: Biological Sciences**, [s. l.], v. 263, n. 1377, p. 1619–1626, 1996.
- BELTON, W. Cattle Egret in Rio Grande do Sul, Brazil. **Bird Banding**, v. 45, p. 59. 1974.
- BICKNELL, A. W. J. et al. Population genetic structure and long-distance dispersal among seabird populations: Implications for colony persistence. **Molecular Ecology**, [s. l.], v. 21, n. 12, p. 2863–2876, 2012.
- BIRDLIFE INTERNATIONAL, 2009. In: DEVENISH, C. et al. **Important Bird Areas Americas - Priority sites for biodiversity conservation**. Quito: BirdLife International, p. 369-382 (BirdLife Conservation Series, n.16), 2009.
- BIRKY, C. W.; FUERST, P.; MARUYAMA, T. Organelle gene diversity under migration, mutation, & drift. **Genetics**, [s. l.], v. 121, p. 613–627, 1989.
- BLAKE, E. R. 1939. African Cattle Heron taken in British Guiana. **Auk**, n. 56, p. 470-471.
- BLACKBURN, T. M. et al. Determinants of establishment success in introduced birds. **Nature**, [s. l.], n. 414, p. 195–197, 2001.
- BLACKBURN, T. M. et al. Avian extinction and mammalian introductions on oceanic islands. **Science**, [s. l.], v. 305, n. 5692, p. 1955–1958, 2004.
- BLAKER, D. Range expansion of the cattle egret. **Ostrich**, [s. l.], v. 42, n. December 2014, p. 27–30, 1971.
- BLANCHET, S. The use of molecular tools in invasion biology: An emphasis on freshwater ecosystems. **Fisheries Management and Ecology**, [s. l.], v. 19, n. 2, p. 120–132, 2012.
- BOECKMANN, B. et al. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. **Nucleic Acids Research**, [s. l.], v. 31, n. 1, p. 365–370, 2003.
- BOEHM, J. T. et al. Population genomics reveals seahorses (*Hippocampus erectus*) of the western Mid-Atlantic coast to be residents rather than vagrants. **PLoS ONE**, [s. l.], v. 10, n. 1, 2015.
- BOHRER, G.; NATHAN, R.; VOLIS, S. Effects of long-distance dispersal for metapopulation survival and genetic structure at ecological time and spatial scales.

- Journal of Ecology**, [s. l.], v. 93, n. 5, p. 1029–1040, 2005.
- BOWEN, V. T.; NICHOLLS, G. D. An egret observed on St Paul's rocks, equatorial Atlantic Ocean. **Auk**, n. 85, p. 130- 131, 1968.
- BRODERSEN, C.; LAVERGNE, S.; MOLOFSKY, J. Genetic variation in photosynthetic characteristics among invasive and native populations of reed canarygrass (*Phalaris arundinacea*). **Biological Invasions**, [s. l.], v. 10, n. 8, p. 1317–1325, 2008.
- BROWDER, J. A. Long-distance movements of Cattle Egrets. *Bird-Banding*, v. 44, n. 3, p. 158-170, 1973.
- BURGER, J. Competition between Cattle Egrets and Native North American Herons, Egrets, and Ibises. **The Condor**, [s. l.], v. 80, n. 1, p. 15, 1978.
- BURGESS, S. C.; MARSHALL, D. J. Are numbers enough? Colonizer phenotype and abundance interact to affect population dynamics. **Journal of Animal Ecology**, [s. l.], v. 80, n. 3, p. 681–687, 2011.
- BURT, D. W. et al. The dynamics of chromosome evolution in birds and mammals. **Nature**, [s. l.], v. 402, n. 6760, p. 411–413, 1999.
- BYRNE, S. et al. Genome Wide Allele Frequency Fingerprints (GWAFs) of Populations via Genotyping by Sequencing. **PLoS ONE**, [s. l.], v. 8, n. 3, 2013.
- CAMPANINI, E. **Populações brasileiras da espécie exótica invasora *Bubulcus ibis*: distribuição da diversidade genética avaliada pelos microsatélites**. 2011. Universidade Federal de São Carlos, [s. l.], 2011.
- CAMPANINI, E. B. et al. Isolation and characterization of 11 microsatellite loci from cattle egret (*Bubulcus ibis*) and cross-amplification in other Ardeidae species. **Conservation Genetics Resources**, [s. l.], v. 4, n. 3, p. 707–709, 2012.
- CAO, L. J. et al. Rapid genetic structuring of populations of the invasive fall webworm in relation to spatial expansion and control campaigns. **Diversity and Distributions**, [s. l.], v. 22, n. 12, p. 1276–1287, 2016.
- CAREY, J. R.; JUDGE, D. S. **Longevity Records: Life Spans of Mammals, Birds, Amphibians, Reptiles, and Fish**. Odense University Press, Odense, 2000.
- CARLTON, J. T. et al. Invasion of Coastal Marine Communities in North America: Apparent Patterns, Processes, and Biases. **Annual Review of Ecology and Systematics**, [s. l.], v. 31, n. 1, p. 481–531, 2002.
- CASTILLO, C. C. **Filogeografia de *Bubulcus ibis* (Linnaeus, 1758) na África e o processo de colonização do continente americano por essa espécie**. 2013. Universidade Federal de São Carlos, [s. l.], 2013.
- CASTOE, T. A. et al. Rapid microsatellite identification from illumina paired-end genomic sequencing in two birds and a snake. **PLoS ONE**, [s. l.], v. 7, n. 2, 2012.
- CAYE, K. et al. TESS3: fast inference of spatial population structure and genome scans for selection. **Molecular Ecology Resources**, [s. l.], v. 16, n. 2, p. 540–548, 2015.
- CHOWN, S. L. et al. Biological invasions, climate change and genomics. **Evolutionary Applications**, [s. l.], v. 8, n. 1, p. 23–46, 2015.
- CLARK, N. A **“Jaunt” Across the Pond: Investigating Patterns of Genetic**

**Differentiation and Trans-Atlantic Migration in a Temperate Seabird (*Morus Bassanus*)**. 2018. Queen's University, [s. l.], 2018.

CLEGG, S. M. et al. Genetic consequences of sequential founder events by an island-colonizing bird. **Proceedings of the National Academy of Sciences**, [s. l.], v. 99, n. 12, p. 8127–8132, 2002.

CONGRAINS, C. et al. Genetic and paleomodelling evidence of the population expansion of the cattle egret *Bubulcus ibis* in Africa during the climatic oscillations of the Late Pleistocene. **Journal of Avian Biology**, [s. l.], v. 47, n. 6, p. 846–857, 2016. a.

CONGRAINS, C. et al. Genetic and paleomodelling evidence of the population expansion of the cattle egret *Bubulcus ibis* in Africa during the climatic oscillations of the Late Pleistocene. **Journal of Avian Biology**, [s. l.], v. 47, n. 6, p. 846–857, 2016. b.

CONSORTIUM INTERNATIONAL CHICKEN GENOME SEQUENCING. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. **Nature**, [s. l.], v. 432, n. December, p. 695–716, 2004.

CORANDER, J. et al. BAPS 2: Enhanced possibilities for the analysis of genetic population structure. **Bioinformatics**, [s. l.], v. 20, n. 15, p. 2363–2369, 2004.

CORANDER, J.; WALDMANN, P.; SILLANPA, M. J. Bayesian Analysis of Genetic Differentiation Between Populations. **Genetics**, [s. l.], v. 374, n. January, p. 367–374, 2003.

CORRE, V. Le; KREMER, A. Cumulative effects of founding events during colonisation on genetic diversity and differentiation in an island and stepping-stone model. **Journal of Evolutionary Biology**, [s. l.], v. 11, p. 495–512, 1998.

COURCHAMP, F. et al. Climate change, sea-level rise, and conservation: Keeping island biodiversity afloat. **Trends in Ecology and Evolution**, [s. l.], v. 29, n. 3, p. 127–130, 2014.

COWEN, S. J. et al. Identification and characterisation of ten microsatellite loci in the Noisy Scrub-bird *Atrichornis clamosus* using next-generation sequencing technology. **Conservation Genetics Resources**, [s. l.], v. 5, n. 3, p. 623–625, 2013.

CRAWFORD, K. M.; WHITNEY, K. D. Population genetic diversity influences colonization success. **Molecular Ecology**, [s. l.], v. 19, n. 6, p. 1253–1263, 2010.

CROSBY, B. G. T. Spread of the Cattle Egret in the Western Hemisphere. **Bird-Banding**, [s. l.], v. 43, n. 3, p. 205–212, 1972.

DAVEY, J. W. et al. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. **Nature Reviews Genetics**, [s. l.], v. 12, n. 7, p. 499–510, 2011.

DAWSON, D. A. et al. New methods to identify conserved microsatellite loci and develop primer sets of high cross-species utility - as demonstrated for birds. **Molecular Ecology Resources**, [s. l.], v. 10, n. 3, p. 475–494, 2010.

DE WIT, P.; PESPENI, M. H.; PALUMBI, S. R. SNP genotyping and population genomics from expressed sequences - Current advances and future possibilities. **Molecular Ecology**, [s. l.], v. 24, n. 10, p. 2310–2323, 2015.

DEL HOYO, J.; ELLIOT, A.; SARGATAL, J. (editors). **Handbook of the birds of the world 1**. Barcelona, 1992.

DICKINSON, E. C. (editor). **The Howard and Moore complete checklist of the birds of the world**. Third edition. London, 2003.

DLUGOSCH, K. M.; PARKER, I. M. Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. **Molecular ecology**, [s. l.], v. 17, n. 1, p. 431–49, 2008.

DOWNS, W. G. Little Egret banded in Spain taken in Trinidad. **Auk**, n. 76, p. 241–242, 1959.

DRURY, W. H., MORGAN A. H. JR., STACKPOLE, R. Occurrence of an African Cattle Egret (*Ardeola ibis ibis*) in Massachusetts. **Auk**, n. 70, p. 364–365, 1953

DUPANLOUP, I.; SCHNEIDER, S.; EXCOFFIER, L. A simulated annealing approach to define the genetic structure of populations. **Molecular Ecology**, [s. l.], v. 11, n. 12, p. 2571–2581, 2002.

EATON, D. A. R. PyRAD: Assembly of de novo RADseq loci for phylogenetic analyses. **Bioinformatics**, [s. l.], v. 30, n. 13, p. 1844–1849, 2014.

EDEVANG, C. et al. Tracking of Arctic terns *Sterna paradisaea* reveals longest animal migration. **Proceedings of the National Academy of Sciences**, [s. l.], v. 107, n. 5, p. 2078–2081, 2010.

ELITH, J.; LEATHWICK, J. R. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. **Annual Review of Ecology, Evolution, and Systematics**, [s. l.], v. 40, n. 1, p. 677–697, 2009.

ELLEGREN, H. The avian genome uncovered. **Trends in Ecology and Evolution**, [s. l.], v. 20, n. 4, p. 180–186, 2005.

ELLEGREN, H. Genome sequencing and population genomics in non-model organisms. **Trends in Ecology and Evolution**, [s. l.], v. 29, n. 1, p. 51–63, 2014.

ELLIS-EVANS, J.; WALTON, D. The process of colonization in Antarctic terrestrial and freshwater ecosystems. **Polar Biol**, [s. l.], v. 3, n. April, p. 151–163, 1990.

ELLISON, W. G. Cattle Egret. **Species Confirmed as Breeders in Vermont**, [s. l.], n. Crosby 1972, p. 40–41, 1978.

ELSHIRE, R. J. et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. **PLoS ONE**, [s. l.], v. 6, n. 5, p. 1–10, 2011.

EO, S. H.; DOYLE, J. M.; DEWOODY, J. A. Genetic diversity in birds is associated with body mass and habitat type. **Journal of Zoology**, [s. l.], v. 283, n. 3, p. 220–226, 2011.

EVANNO, G.; REGNAUT, S.; GOUDET, J. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. **Molecular Ecology**, [s. l.], v. 14, n. 8, p. 2611–2620, 2005.

EWING, B.; HILLIER, L. D.; WENDL, M. C. Base-Calling of Automated Sequencer Traces Using Phred. **Genome Research**, [s. l.], v. 8, n. 206, p. 186–194, 1998.

EXCOFFIER, L.; LISCHER, H. E. L. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. **Molecular Ecology Resources**, [s. l.], v. 10, n. 3, p. 564–567, 2010.

FAYARD, J.; LEFÈVRE, F.; KLEIN, E. K. Long distance dispersal and the fate of a

- gene from the colonization front. **Journal of Evolutionary Biology**, [s. l.], v. 22, n. 11, p. 2171–2182, 2009.
- FELICÍSIMO, Á. M.; MUÑOZ, J.; GONZÁLEZ-SOLIS, J. Ocean surface winds drive dynamics of transoceanic aerial movements. **PLoS ONE**, [s. l.], v. 3, n. 8, p. 1–7, 2008.
- FELLOWS, D. P.; PATON, P. W. C. (1988). Behavioral response of Cattle Egrets to population control measures in Hawaii, p. 315–318. In: A. C. Crabb and R. E. Marsh (eds). *Proceedings of the 13th Annual Vertebrate Pest Conference*. Davis: University of California. DAVIS, M.A. **Invasion Biology**. Oxford: Oxford University Press, 2009. 243p.
- FERREIRA, J. A.; ZWINDERMAN, A. H. On the Benjamini-Hochberg method. **Annals of Statistics**, [s. l.], v. 34, n. 4, p. 1827–1849, 2006.
- FICK, S. E.; HIJMANS, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. **International Journal of Climatology**, [s. l.], v. 37, n. 12, p. 4302–4315, 2017.
- FOLINO-ROREM, N. C.; DARLING, J. A.; D'AUSILIO, C. A. Genetic analysis reveals multiple cryptic invasive species of the hydrozoan genus *Cordylophora*. **Biological Invasions**, [s. l.], v. 11, n. 8, p. 1869–1882, 2009.
- FOLL, M.; GAGGIOTTI, O. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. **Genetics**, [s. l.], v. 180, n. 2, p. 977–993, 2008.
- FRANCIS, R. M. pophelper: an R package and web app to analyse and visualize population structure. **Molecular Ecology Resources**, [s. l.], v. 17, n. 1, p. 27–32, 2017.
- FRICHOT, E.; FRANÇOIS, O. LEA: An R package for landscape and ecological association studies. **Methods in Ecology and Evolution**, [s. l.], v. 6, n. 8, p. 925–929, 2015.
- FRIESEN, V. L. New roles for molecular genetics in understanding seabird evolution, ecology and conservation. **Marine Ornithology**, [s. l.], v. 35, n. 2, p. 89–96, 2007.
- FRIESEN, V. L.; BURG, T. M.; MCCOY, K. D. Mechanisms of population differentiation in seabirds: Invited review. **Molecular Ecology**, [s. l.], v. 16, n. 9, p. 1765–1785, 2007.
- GANDON, S.; ROUSSET, F. Evolution of stepping-stone dispersal rates. **Proceedings of the Royal Society B: Biological Sciences**, [s. l.], v. 266, n. 1437, p. 2507–2513, 1999.
- GENTLEMAN, R. C. et al. Bioconductor: open software development for computational biology and bioinformatics. **Genome Biology**, [s. l.], v. 5, n. 10, p. R80, 2004.
- GENTON, B. J.; SHYKOFF, J. A.; GIRAUD, T. High genetic diversity in French invasive populations of common ragweed, *Ambrosia artemisiifolia*, as a result of multiple sources of introduction. **Molecular Ecology**, [s. l.], v. 14, n. 14, p. 4275–4285, 2005.
- GERONIMO, C. T. **Diversidade e estruturação genética em populações africanas e brasileiras de garça-vaqueira ( *Bubulcus ibis* , Linnaeus , 1758 ) determinadas por microssatélites**. 2018. Universidade Federal de São Carlos, [s. l.], 2018.

- GILLESPIE, R. G. et al. Long-distance dispersal: A framework for hypothesis testing. **Trends in Ecology and Evolution**, [s. l.], v. 27, n. 1, p. 47–56, 2012.
- GISD. Global Invasive Species Database. 2017. Disponível em <<http://www.iucngisd.org/gisd/search.php> on 09-03-2019>. Acesso em: 05 de junho de 2017.
- GONZÁLEZ-CASTELLANO, I. et al. Isolation and characterization of 21 polymorphic microsatellite loci for the rockpool shrimp *Palaemon elegans* using Illumina MiSeq sequencing. **Scientific Reports**, [s. l.], v. 8, n. 1, p. 8–13, 2018.
- GORJANC, G. et al. Potential of genotyping-by-sequencing for genomic selection in livestock populations. **Genetics Selection Evolution**, [s. l.], v. 47, n. 1, 2015.
- GOUDET, J. HIERFSTAT , a package for R to compute and test hierarchical F-statistics. **Molecular Ecology Notes**, [s. l.], v. 5, p. 184–186, 2005.
- GRANT, P. R.; GRANT, B. R.; PETREN, K. A population founded by a single pair of individuals: Establishment, expansion, and evolution. **Genetica**, [s. l.], v. Special, p. 112–113, 2001.
- GREEN, A. J.; FIGUEROLA, J. Recent advances in the study of long-distance dispersal of aquatic invertebrates via birds. **Diversity and Distributions**, [s. l.], v. 11, n. 2, p. 149–156, 2005.
- GRUBER, B. et al. dartr: An r package to facilitate analysis of SNP data generated from reduced representation genome sequencing. **Molecular Ecology Resources**, [s. l.], v. 18, n. 3, p. 691–699, 2018.
- HANS, J. B. et al. Characterization of MHC class II B polymorphism in multiple populations of wild gorillas using non-invasive samples and next-generation sequencing. **American Journal of Primatology**, [s. l.], v. 77, n. 11, p. 1193–1206, 2015.
- HANSKI, I. Connectivity, Habitat Continuity, and Metapopulations in Dynamic Landscapes. **Oikos**, [s. l.], v. 87, n. 2, p. 209–219, 1999.
- HEDD, A. et al. Trans-equatorial migration and habitat use by sooty shearwaters *Puffinus griseus* from the South Atlantic during the nonbreeding season. **Marine Ecology Progress Series**, [s. l.], v. 449, p. 277–290, 2012.
- HEDGE, L. H.; O’CONNOR, W. A.; JOHNSTON, E. L. Manipulating the intrinsic parameters of propagule pressure: implications for bio-invasion. **Ecosphere**, [s. l.], v. 3, n. 6, p. art48, 2012.
- HEMMER-HANSEN, J.; THERKILDSEN, N. O.; PUJOLAR, J. M. Population genomics of marine fishes: Next-generation prospects and challenges. **Biological Bulletin**, [s. l.], v. 227, n. 2, p. 117–132, 2014.
- HEWITT, G. Some genetic consequences of ice ages, and their role, in divergence and speciation. **Biological Journal of the Linnean Society**, [s. l.], v. 58, n. 3, p. 247–276, 1996.
- HEWITT, G. The genetic legacy of the Quaternary ice ages. **Nature**, [s. l.], v. 405, p. 907–913, 2000.
- HIGGINS, S. I.; RICHARDSON, D. M.; COWLING, R. M. Modeling Invasive Plant Spread : The Role of Plant-Environment Interactions and Model Structure Stable URL : <http://www.jstor.org/stable/2265699> REFERENCES Linked references are available on

- JSTOR for this article : You may need to log in to JSTOR to acce. [s. l.], v. 77, n. 7, p. 2043–2054, 1996.
- HOLM, S. A Simple Sequentially Rejective Multiple Test Procedure. **Scandinavian Journal of Statistics**, [s. l.], v. 6, n. 2, p. 65–70, 1979.
- HORREO, J. L. et al. Identification of polymorphic microsatellite loci for the endangered great bustard (*Otis tarda*) by high-throughput sequencing. **Conservation Genetics Resources**, [s. l.], v. 5, n. 2, p. 549–551, 2013.
- HUBISZ, M. J. et al. Inferring weak population structure with the assistance of sample group information. **Molecular Ecology Resources**, [s. l.], v. 9, n. 5, p. 1322–1332, 2009.
- HUGHES, A. L.; HUGHES, M. K. Small genomes for better flyers. **Nature**, [s. l.], v. 377, n. 6548, p. 391, 1995.
- IBAMA, FUNATURA. Plano de manejo do Parque Nacional Marinho de Fernando de Noronha, 1<sup>a</sup> ed. Brasília: IBAMA, 1990.
- JOLLIFFE, I. **Principal Component Analysis**. 2011. ed. Berlin: Springer Berlin Heidelberg, 2011.
- JOMBART, T.; AHMED, I. adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. **Bioinformatics**, [s. l.], v. 27, n. 21, p. 3070–3071, 2011.
- JOMBART, T.; DEVILLARD, S.; BALLOUX, F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. **BMC Genetics**, [s. l.], v. 11, p. 94, 2010.
- JONES, G. P. et al. Recruitment and the Local Dynamics of Open Marine Populations. **Annual Review of Ecology and Systematics**, [s. l.], v. 27, n. 1, p. 477–500, 2002.
- JONES, H. P. et al. Severity of the effects of invasive rats on seabirds: A global review. **Conservation Biology**, [s. l.], v. 22, n. 1, p. 16–26, 2008.
- KALINOWSKI, S. T. HP-RARE 1.0: A computer program for performing rarefaction on measures of allelic richness. **Molecular Ecology Notes**, [s. l.], v. 5, n. 1, p. 187–189, 2005.
- KALINOWSKI, S. T.; TAPER, M. L.; MARSHALL, T. C. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. **Molecular Ecology**, [s. l.], v. 16, n. 5, p. 1099–1106, 2007.
- KAPUSTA, A.; SUH, A.; FESCHOTTE, C. Dynamics of genome size evolution in birds and mammals. **Proceedings of the National Academy of Sciences**, [s. l.], v. 114, n. 8, p. E1460–E1469, 2017.
- KAWECKI, T. J.; EBERT, D. Conceptual issues in local adaptation. **Ecology Letters**, [s. l.], v. 7, n. 12, p. 1225–1241, 2004.
- KEENAN, K. et al. DiveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. **Methods in Ecology and Evolution**, [s. l.], v. 4, n. 8, p. 782–788, 2013.
- KENNEDY, D. Determinate e indeterminate egg-laying patterns: a review. **The Condor**, [s. l.], n. 93, p. 106–124, 1991.
- KIM, K. S. et al. Utility of EST-derived SSRs as population genetics markers in a beetle. **Journal of Heredity**, [s. l.], v. 99, n. 2, p. 112–124, 2008.

- KIMURA, M.; WEISS, G. H. The stepping stone model of population structure and the decrease of genetic correlation with distance. **Genetics**, [s. l.], v. 49, n. 480, p. 561–576, 1964.
- KLEIN, E. K.; LAVIGNE, C.; GOUYON, P. H. Mixing of propagules from discrete sources at long distance: Comparing a dispersal tail to an exponential. **BMC Ecology**, [s. l.], v. 6, p. 1–12, 2006.
- KOKKO, H.; LÓPEZ-SEPULCRE, A. The ecogenetic link between demography and evolution: Can we bridge the gap between theory and data? **Ecology Letters**, [s. l.], v. 10, n. 9, p. 773–782, 2007.
- KOLAR, K. Breeding of the Cattle Egret (*Bubulcus ibis*). **Avicultural Magazine**, n. 72, p. 45-46, 1966.
- KOLBE, J. J. et al. Genetic variation increases during biological invasion by a Cuban lizard. **Nature**, [s. l.], v. 431, n. 7005, p. 177–181, 2004.
- KORESSAAR, T.; REMM, M. Enhancements and modifications of primer design program Primer3. **Bioinformatics**, [s. l.], v. 23, n. 10, p. 1289–1291, 2007.
- KREBS, E. A.; RIVEN-RAMSEY, D.; HUNTE, W. The Colonization of Barbados by Cattle Egrets (*Bubulcus ibis*) 1956-1990. **Colonial Waterbirds**, [s. l.], v. 17, n. 1, p. 86–90, 1994.
- KUSHLAN, J. A.; HAFNER, H. **Heron Conservation**. New York: Academic Press. 2000.
- LACY, R. C. Vortex: A computer simulation model for population viability analysis. **Wildlife Research**, [s. l.], v. 20, n. 1, p. 1–13, 1993.
- LAMBRINOS, J. G. HOW INTERACTIONS BETWEEN ECOLOGY AND EVOLUTION INFLUENCE CONTEMPORARY INVASION DYNAMICS. **Ecology**, [s. l.], v. 85, n. 8, p. 2061–2070, 2004.
- LANCE, S. L. et al. 32 species validation of a new Illumina paired-end approach for the development of microsatellites. **PLoS ONE**, [s. l.], v. 8, n. 11, p. 1–11, 2013.
- LARSON, W. A. et al. Genotyping by sequencing resolves shallow population structure to inform conservation of Chinook salmon (*Oncorhynchus tshawytscha*). **Evolutionary Applications**, [s. l.], v. 7, n. 3, p. 355–369, 2014.
- LAWSON HANDLEY, L. J. et al. Ecological genetics of invasive alien species. **BioControl**, [s. l.], v. 56, n. 4, p. 409–428, 2011.
- LERNER, H. R. L.; FLEISCHER, R. C. Prospects for the Use of Next-Generation Sequencing Methods in Ornithology Special Reviews in Ornithology Prospects for the Use of Next-Generation Sequencing Methods in Ornithology. [s. l.], v. 127, n. 1, p. 4–15, 2010.
- LEVEQUE, R., BOWMAN, R. I.; BILLEB, S. L. Migrants in the Galapagos area. **Condor**, n. 68, p. 81-101, 1966
- LIECHTI, F. Birds: Blowin' by the wind? **Journal of Ornithology**, [s. l.], v. 147, n. 2, p. 202–211, 2006.
- LIM, H. C.; BRAUN, M. J. High-throughput SNP genotyping of historical and modern samples of five bird species via sequence capture of ultraconserved elements. **Molecular ecology resources**, [s. l.], v. 16, n. 5, p. 1204–1223, 2016.



- LOCKWOOD, J. L.; CASSEY, P.; BLACKBURN, T. The role of propagule pressure in explaining species invasions. **Trends in Ecology and Evolution**, [s. l.], v. 20, n. 5, p. 223–228, 2005.
- LOCKWOOD, J. L.; CASSEY, P.; BLACKBURN, T. M. The more you introduce the more you get: The role of colonization pressure and propagule pressure in invasion ecology. **Diversity and Distributions**, [s. l.], v. 15, n. 5, p. 904–910, 2009.
- LUIKART, G. et al. The power and promise of population genomics: From genotyping to genome typing. **Nature Reviews Genetics**, [s. l.], v. 4, n. 12, p. 981–994, 2003.
- MADDOCK, M.; BRIDGMAN, H. Cattle Egret migration and meteorological conditions. **Notornis**, [s. l.], v. 39, n. 1988, p. 73–86, 1992.
- MANEL, S. et al. Landscape genetics: Combining landscape ecology and population genetics. **Trends in Ecology and Evolution**, [s. l.], v. 18, n. 4, p. 189–197, 2003.
- MARFIM, A. “Bubulcus ibis” (Online), Animal Diversity Web. 2000. Disponível em <[http://animaldiversity.ummz.umich.edu/site/accounts/information/Bubulcus\\_ibis.htm](http://animaldiversity.ummz.umich.edu/site/accounts/information/Bubulcus_ibis.htm)>. Acesso em: 20 de março de 2015.
- MARION, L.; MARION, P. Le Héron garde-bœufs (*Bubulcus ibis*) niche dans l’ouest de la France. Statut de l’espèce en France. **Alauda**, n. 50, p. 161-175, 1982.
- MASSA, C.; DOYLE, M.; CALLICÓ FORTUNATO, R. On how Cattle Egret (*Bubulcus ibis*) spread to the Americas: meteorological tools to assess probable colonization trajectories. **International Journal of Biometeorology**, [s. l.], v. 58, n. 9, p. 1879–1891, 2014.
- MASTRETTA-YANES, A. et al. Gene duplication, population genomics, and species-level differentiation within a tropical mountain shrub. **Genome Biology and Evolution**, [s. l.], v. 6, n. 10, p. 2611–2624, 2014.
- MAYR, E. Geographic speciation in tropical echinoids. **Evolution**, [s. l.], v. 8, n. 1, p. 14–16, 1954.
- MCGEOCH, M. A. et al. Monitoring biological invasion across the broader Antarctic: A baseline and indicator framework. **Global Environmental Change**, [s. l.], v. 32, p. 108–125, 2015.
- MEDINA, F. M. et al. A global review of the impacts of invasive cats on island endangered vertebrates. **Global Change Biology**, [s. l.], v. 17, n. 11, p. 3503–3510, 2011.
- MEESE, R. J. Cattle egret predation causing reproductive failures of nesting tricolored blackbirds. **California Fish and Game**, [s. l.], v. 98, n. 1, p. 47–50, 2012.
- MEGLÉCZ, E. et al. QDD: a user-friendly program to select microsatellite markers and design primers from large sequencing projects. **Bioinformatics**, [s. l.], v. 26, n. 3, p. 403–404, 2009.
- MEIRMANS, P. G.; VAN TIENDEREN, P. H. GENOTYPE and GENODIVE: Two programs for the analysis of genetic diversity of asexual organisms. **Molecular Ecology Notes**, [s. l.], v. 4, n. 4, p. 792–794, 2004.
- MESTRE, L. A. M.; ROOS, A. L.; RECHETELO, J. Habitat Associations of Land Birds in. **Ornitologia Neotropical**, [s. l.], n. July, p. 26–34, 2016.
- MIÑO, C. I. et al. Use of noninvasive ‘bug-eggs’ to enable comparative inferences on

- genetic mating system with and without parental information: A study in a cattle egret colony. **Plos One**, [s. l.], v. 12, n. 8, p. e0183153, 2017.
- MOHANTY, M. K.; BHUNYA, S. P. Karyological studies in four species of ardeid birds (Ardelidae, Ciconiiformes). **Genetica**, [s. l.], v. 81, n. 3, p. 211–214, 1990.
- MORALEZ-SILVA, E.; DEL LAMA, S. N. Colonization of Brazil by the cattle egret (*Bubulcus ibis*) revealed by mitochondrial DNA. **NeoBiota**, [s. l.], v. 21, n. November, p. 49–63, 2014.
- MORALEZ-SILVA, E.; DEL LAMA, S. N. Multiple Females Laying in Cattle Egret (*Bubulcus ibis*) Clutches: Behavioral and Genetic Evidence. **Waterbirds**, [s. l.], v. 40, n. 4, p. 403–409, 2017.
- MORALEZ-SILVA, E.; NASSIF DEL LAMA, S. Colonization of Brazil by the cattle egret (*Bubulcus ibis*) revealed by mitochondrial DNA. **NeoBiota**, [s. l.], v. 21, p. 49–63, 2014.
- MORIN, P. A.; LUIKART, G.; WAYNE, R. K. SNPs in ecology, evolution and conservation. **Trends in Ecology and Evolution**, [s. l.], v. 19, n. 4, p. 208–216, 2004.
- MORIN, P. A.; MARTIEN, K. K.; TAYLOR, B. L. Assessing statistical power of SNPs for population structure and conservation studies. **Molecular Ecology Resources**, [s. l.], v. 9, n. 1, p. 66–73, 2009.
- MURPHY, M. A. et al. Representing genetic variation as continuous surfaces: An approach for identifying spatial dependency in landscape genetic studies. **Ecography**, [s. l.], v. 31, n. 6, p. 685–697, 2008.
- NACINOVIC, J. B.; TEIXEIRA, D. M. As aves de Fernando de Noronha: uma lista sistemática anotada. **Revista Brasileira de Biologia**, n. 49, p. 709-729, 1989.
- NARUM, S. R. et al. Population differentiation determined from putative neutral and divergent adaptive genetic markers in Eulachon (*Thaleichthys pacificus*, Osmeridae), an anadromous Pacific smelt. **Molecular Ecology Resources**, [s. l.], v. 15, n. 6, p. 1421–1434, 2015.
- NEI, M.; MARUYAMA, T.; CHAKRABORTY, R. The bottleneck effect and genetic variability in populations. **Evolution**, [s. l.], v. 29, n. 1, p. 1–10, 1975.
- NEI, M.; TAJIMA, F. Genetic drift and estimation of effective population size. **Genetics**, [s. l.], v. 98, n. 3, p. 625–640, 1981.
- NICHOLS, R. A.; HEWITT, G. M. The genetic consequences of long distance dispersal during colonization. **Heredity**, [s. l.], v. 72, n. 3, p. 312–317, 1994.
- NIETLISBACH, P. et al. Pedigree-based inbreeding coefficient explains more variation in fitness than heterozygosity at 160 microsatellites in a wild bird population. **Proceedings of the Royal Society B: Biological Sciences**, [s. l.], v. 284, n. 1850, 2017.
- NUNES, M. F. C. et al. The Cattle Egret (*Bubulcus ibis*) on Fernando de Noronha Archipelago: history and population trends. **Revista Brasileira de Ornitologia**, [s. l.], v. 18, n. 4, p. 315–327, 2010.
- OJEDA, G. N. et al. Mating system of *Caiman yacare* (Reptilia: Alligatoridae) described from microsatellite genotypes. **Journal of Heredity**, [s. l.], v. 108, n. 2, p. 135–141, 2017.
- ORR, H. A. Testing Natural Selection vs. Genetic Drift in Phenotypic Evolution Using

- Quantitative Trait Locus Data. **Genetics**, [s. l.], v. 149, p. 2099–2104, 1998.
- ORSINI, L. et al. Drivers of population genetic differentiation in the wild: Isolation by dispersal limitation, isolation by adaptation and isolation by colonization. **Molecular Ecology**, [s. l.], v. 22, n. 24, p. 5983–5999, 2013.
- PAETKAU D. et al. Microsatellite analysis of population structure in Canadian polar bears. **Molecular Ecology**, [s. l.], v. 4, p. 347–354, 1995.
- PASCUAL, M. et al. Microsatellite variation in colonizing and palearctic populations of *Drosophila subobscura*. **Molecular Biology and Evolution**, [s. l.], v. 18, n. 5, p. 731–740, 2001.
- PEAKALL, R. O. D.; SMOUSE, P. E. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. **Molecular Ecology Notes**, v. 6, n. 1, p. 288–295, 2006
- PETKOVA, D.; NOVEMBRE, J.; STEPHENS, M. Visualizing spatial population structure with estimated effective migration surfaces. **Nature Genetics**, [s. l.], v. 48, n. 1, p. 94–100, 2015.
- PHILLIPS, R. B.; WIEDENFELD, D. A.; SNELL, H. L. Current status of alien vertebrates in the Galápagos Islands: Invasion history, distribution, and potential impacts. **Biological Invasions**, [s. l.], v. 14, n. 2, p. 461–480, 2012.
- PICKRELL, J. K.; PRITCHARD, J. K. Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. **PLoS Genetics**, [s. l.], v. 8, n. 11, 2012.
- PIERTNEY, S. B. et al. Resolving patterns of population genetic and phylogeographic structure to inform control and eradication initiatives for brown rats *Rattus norvegicus* on South Georgia. **Journal of Applied Ecology**, [s. l.], v. 53, n. 2, p. 332–339, 2016.
- PIRY, S. et al. GENECLASS2: A Software for Genetic Assignment and First-Generation Migrant Detection. **Journal of Heredity**, [s. l.], v. 95, n. 6, p. 536–539, 2004.
- POLAND, J. A.; RIFE, T. W. Genotyping-by-Sequencing for Plant Breeding and Genetics. **The Plant Genome Journal**, [s. l.], v. 5, n. 3, p. 92, 2012.
- POOTAKHAM, W. et al. Genome-wide SNP discovery and identification of QTL associated with agronomic traits in oil palm using genotyping-by-sequencing (GBS). **Genomics**, [s. l.], v. 105, n. 5–6, p. 288–295, 2015.
- POST, P. W. First report of the Cattle Egret in Chile and range expansion in Peru. **Auk**, n. 87, p. 361, 1970.
- PRIMMER, C. R. et al. Low frequency of microsatellites in the avian genome. **Genome Research**, [s. l.], v. 7, n. 5, p. 471–482, 1997.
- PRITCHARD, J. K.; STEPHENS, M.; DONNELLY, P. Inference of Population Structure Using Multilocus Genotype Data. **Genetics**, [s. l.], v. 155, n. 2, p. 945–959, 2000.
- PRUETT, C. L.; WINKER, K. Northwestern song sparrow populations show genetic effects of sequential colonization. **Molecular Ecology**, [s. l.], v. 14, n. 5, p. 1421–1434, 2005.
- PURCELL, S. et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. **The American Journal of Human Genetics**, [s.

l.], v. 81, n. 3, p. 559–575, 2007.

PYŠEK, P.; RICHARDSON, D. M. The biogeography of naturalization in alien plants. **Journal of Biogeography**, [s. l.], v. 33, n. 12, p. 2040–2050, 2006.

R DEVELOPMENT CORE TEAM v3.4.4. 2018. R: A language and environment for statistical computing. Viena: R Foundation for Statistical Computing. Disponível em: <<http://www.R-project.org/>>. Acesso em: 01 mar. 2018.

RABBI, I. et al. Genetic mapping using genotyping-by-sequencing in the clonally propagated cassava. **Crop Science**, [s. l.], v. 54, n. 4, p. 1384–1396, 2014.

RAMBAUT, A. et al. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. **Systematic Biology**, [s. l.], v. 67, n. 5, p. 901–904, 2018.

RANDI, E. et al. Development of SNP markers for population structure and phylogeography characterization in little owl (*Athene noctua*) using a genotyping-by-sequencing approach. **Conservation Genetics Resources**, [s. l.], v. 8, n. 1, p. 13–16, 2015.

RANNALA B. BayesAss edition 3.0 user's manual. Davis (CA): University of California. Disponível em: <<http://ftp.ie.freshrpms.net/pub/sourceforge/b/ba/bayesass/BA3/3.0.3/docs/BA3Manual.pdf>>. Acessado em 10 de janeiro de 2019.

RANNALA, B.; MOUNTAIN, J. Detecting immigration by using multilocus genotypes. **Proceedings of the National Academy of Sciences**, [s. l.], v. 94, n. 17, p. 9197–9201, 1997.

ROLLINS, L. A. et al. Invasive species can't cover their tracks: Using microsatellites to assist management of starling (*Sturnus vulgaris*) populations in Western Australia. **Molecular Ecology**, [s. l.], v. 18, n. 8, p. 1560–1573, 2009.

ROUSSET, F. GENEPOP'007: A complete re-implementation of the GENEPOP software for Windows and Linux. **Molecular Ecology Resources**, [s. l.], v. 8, n. 1, p. 103–106, 2008.

SAHU, B. P. et al. Isolation and characterization of polymorphic microsatellite loci in Indian major carp, *Catla catla* using next-generation sequencing platform. **Biochemical Systematics and Ecology**, [s. l.], v. 57, p. 357–362, 2014.

SAMRAOUI, F.; MENAI, R.; SAMRAOUI, B. Reproductive ecology of the Cattle Egret (*Bubulcus ibis*) at Sidi Achour, north-eastern Algeria. **Ostrich**, [s. l.], v. 78, n. 2, p. 481–487, 2007.

SCHUELKE, M. An economic method for the fluorescent labeling of PCR fragments - A poor man's approach to genotyping for research and high-throughput diagnostics. **Nature Biotechnology**, [s. l.], v. 18, n. February, p. 233–234, 2000.

SCHULZ, M. H. et al. Oases: Robust de novo RNA-seq assembly across the dynamic range of expression levels. **Bioinformatics**, [s. l.], v. 28, n. 8, p. 1086–1092, 2012.

SEXTON, J. P.; HANGARTNER, S. B.; HOFFMANN, A. A. Genetic isolation by environment or distance: Which pattern of gene flow is most common? **Evolution**, [s. l.], v. 68, n. 1, p. 1–15, 2014.

SICK, H. *Bubulcus ibis* (L.) na ilha de Marajó, Pará: Garça ainda não registrada no Brasil. Anais da Academia Brasileira de Ciências, Rio de Janeiro. v. 37, p. 567-570. 1965.

SIEGFRIED, W.R. The status of the Cattle Egret in the Cape Province. **The Ostrich**, v. 36, n. 3, p. 109-116, 1965.

SILVA, R. S.; SILVA, J. R. **Reprodução e status da Garça-vaqueira (*Bulbucus ibis*) no Arquipélago de Fernando de Noronha**. Em: Resumos do XI Congresso Brasileiro de Ornitologia. Feira de Santana: Universidade Estadual Feira de Santana, 2003.

SPRUNT, A., JR. The spread of the Cattle Egret. Smithsonian Institution Annual Report for 1954, n. 4198, p. 259-276, 1955.

RASMUSSEN, P. C.; ANDERTON, J. C. **Birds of South Asia: the Ripley guide 1 & 2**. Barcelona, 2005.

SHAFFER, S. A. et al. Migratory shearwaters integrate oceanic resources across the Pacific Ocean in an endless summer. **Proceedings of the National Academy of Sciences**, [s. l.], v. 103, n. 34, p. 12799–12802, 2006.

SHARAH, A. H.; ALI, A. E.; MOHAMMED, D. I. The Feeding Behavior of the Cattle Egrets, (*Bulbucus ibis* L.) in Northeastern Arid Zone of Nigeria. **Journal of Agriculture & Social Sciences**, [s. l.], v. 4, n. 1983, p. 6–12, 2008.

SI BACHIR, A. et al. The recent expansion of an avian invasive species (the Cattle Egret *Ardea ibis*) in Algeria. **Journal of Arid Environments**, [s. l.], v. 75, n. 11, p. 1232–1236, 2011.

SIEGFRIED, W. R. COMMUNAL ROOSTING OF THE CATTLE EGRET. **Transactions of the Royal Society of South Africa**, [s. l.], v. 39, n. 4, p. 419–443, 1971.

SIEGFRIED, W. R. Energy Metabolism of the Cattle Egret. **Zoologica Africana**, [s. l.], v. 4, n. 2, p. 265–273, 2015.

SILVA-JR, J. M.; PERES-JR, A. K.; SAZIMA, I. Natural History Notes. *Euprepis atlanticus* (Noronha Skink). Predation. **Herpetological Review**, [s. l.], n. 36, p. 62–63, 2005.

SIMBERLOFF, D. et al. Impacts of biological invasions: What's what and the way forward. **Trends in Ecology and Evolution**, [s. l.], v. 28, n. 1, p. 58–66, 2013.

SLARKIN, M. Gene Flow in Natural Populations. **Annual Review of Ecology and Systematics**, [s. l.], v. 16, n. 1, p. 393–430, 1985.

SPURGIN, L. G. et al. Genetic and phenotypic divergence in an island bird: Isolation by distance, by colonization or by adaptation? **Molecular Ecology**, [s. l.], v. 23, n. 5, p. 1028–1039, 2014.

TELFAIR., R.C. **The Cattle Egret: a Texas focus and world view**. Texas: Texas A&M Univ. Press. 144 p., 1983.

TELFAIR, R.C. "Cattle egret (*Bulbucus ibis*)." **The Birds of North America**, n.113, POOLE, A.; GILL, F. eds. The Academy of Natural Sciences. The American Ornithologists' Union, Philadelphia, Washington, D.C. p. 877. 1994.

TIEDEMANN, R. et al. Mitochondrial DNA and microsatellite variation in the eider duck (*Somateria mollissima*) indicate stepwise postglacial colonization of Europe and limited current long-distance dispersal. **Molecular Ecology**, [s. l.], v. 13, n. 6, p. 1481–1494, 2004.

TIGANO, A. et al. Outlier analyses to test for local adaptation to breeding grounds in a

- migratory arctic seabird. **Ecology and Evolution**, [s. l.], v. 7, n. 7, p. 2370–2381, 2017.
- TOEWS, D. P. L. et al. Genomic approaches to understanding population divergence and speciation in birds. **The Auk**, [s. l.], v. 133, n. 1, p. 13–30, 2015.
- UNTERGASSER, A. et al. Primer3-new capabilities and interfaces. **Nucleic Acids Research**, [s. l.], v. 40, n. 15, p. 1–12, 2012.
- VALVERDE, J. A.; WEICKERT, P. **Sobre la migración de varias garzas españolas**, 1956.
- VALVERDE, J.A. **Memorias de un biólogo heterodoxo**. España: Quercus Ed., 2003.
- VAN DEN BOSCH, F.; HENGEVELD, R.; METZ, J. A. J. Analysing the Velocity of Animal Range Expansion. **Journal of Biogeography**, [s. l.], v. 19, n. 2, p. 135, 2006.
- VAN DEN BUSSCHE, R. A.; LONGMIRE, J. L.; BAKER, R. J. How bats achieve a small C-value: frequency of repetitive DNA in *Macrotus*. **Mammalian Genome**, [s. l.], v. 6, n. 8, p. 521–525, 1995.
- VAN OOSTERHOUT, C. et al. MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. **Molecular Ecology Notes**, [s. l.], v. 4, n. 3, p. 535–538, 2004.
- VAURIE, C. Systematic notes on the Cattle Egret (*Bubulcus ibis*). **Bulletin of the British Ornithologists' Club**, n. 83, p. 164–166, 1963.
- VITALI, M. **Conservação da biodiversidade e uso dos recursos naturais em Fernando de Noronha: sustentabilidade em ambientes sensíveis**. 2009. Universidade de Brasília, [s. l.], 2009.
- VOISIN, C. **The herons of Europe**. London, 1991.
- WALTARI; EDWARDS. Evolutionary Dynamics of Intron Size, Genome Size, and Physiological Correlates in Archosaurs. **The American Naturalist**, [s. l.], v. 160, n. 5, p. 539, 2002.
- WANG, I. J. Examining the full effects of landscape heterogeneity on spatial genetic variation: A multiple matrix regression approach for quantifying geographic and ecological isolation. **Evolution**, [s. l.], v. 67, n. 12, p. 3403–3411, 2013.
- WANG, J. Informativeness of genetic markers for pairwise relationship and relatedness inference. **Theoretical Population Biology**, [s. l.], v. 70, n. 3, p. 300–321, 2006.
- WATTERSON, G. A. On the Number of Segregating Sites in Genetical Models without Recombination. **Theoretical Population Biology**, [s. l.], v. 276, n. 7, p. 256–276, 1975.
- WEIR, B. S.; COCKERHAM, C. C. Covariances of Relatives Stemming from a Population Undergoing Mixed Self and Random Mating. **Biometrics**, [s. l.], v. 40, n. 1, p. 157, 1984.
- WESTERSKOV, K. E. Probable first breeding of the cattle egret (*Bubulcus ibis*) in the New Zealand. **Notornis**, [s. l.], v. 21, n. 3, p. 205–218, 1974.
- WILSON, G.; RANNALA, B. Bayesian inference of recent migration rates using multilocus genotypes. **Genetics**, [s. l.], v. 163, n. 3, p. 1177–1191, 2003.
- WILSON, J. R. U. et al. Something in the way you move: dispersal pathways affect invasion success. **Trends in Ecology and Evolution**, [s. l.], v. 24, n. 3, p. 136–144, 2009.

- WOLD, J. R. **Phylogenetic relationships, population connectivity, and the development of genetic assignment testing in Buller's Albatross (*Thalassarche bulleri*)**. 2017. Victoria University of Wellington, [s. l.], 2017.
- WRIGHT, S. Evolution in Mendelian Populations. **Genetics**, [s. l.], v. 16, n. 2, p. 97–159, 1931.
- WRIGHT, S. Isolation by distance. **Genetics**, [s. l.], v. 28, p. 114–118, 1943.
- WRIGHT, S. Genetical structure of populations. **Nature**, [s. l.], v. 166, n. 4215, p. 247–249, 1950.
- WRIGHT, S. Character change, speciation, and the higher taxa. **Evolution**, [s. l.], v. 36, n. 3, p. 427–443, 1982.
- WYNNE-EDWARDS, V. C. **Animal dispersion**. Edinburgh and London, Oliver and Boyd, 1962.
- XING, C. et al. Comparison of microsatellites, single-nucleotide polymorphisms (SNPs) and composite markers derived from SNPs in linkage analysis. **BMC Genetics**, [s. l.], v. 6, n. Suppl 1, p. S29, 2006.
- YANG, C. C. et al. Propagule pressure and colony social organization are associated with the successful invasion and rapid range expansion of fire ants in China. **Molecular Ecology**, [s. l.], v. 21, n. 4, p. 817–833, 2012.
- YANNIC, G. et al. High connectivity in a long-lived high-Arctic seabird, the ivory gull *Pagophila eburnea*. **Polar Biology**, [s. l.], v. 39, n. 2, p. 221–236, 2016.
- YU, J. N. et al. Fast and cost-effective mining of microsatellite markers using NGS technology: An example of a Korean water deer *Hydropotes inermis argyropus*. **PLoS ONE**, [s. l.], v. 6, n. 10, 2011.
- ZANE, L.; BARGELLONI, L.; PATARNELLO, T. Strategies for microsatellite isolation: a review. **Molecular ecology**, [s. l.], v. 11, n. 1, p. 1–16, 2002.
- ZERBINO, D. R.; BIRNEY, E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. **Genome Research**, [s. l.], v. 18, n. 5, p. 821–829, 2008.
- ZHANG, G. et al. Comparative genomics reveals insights into avian genome evolution and adaptation. **Science**, [s. l.], v. 346, n. 6215, p. 1311–1321, 2014.
- ZHBANNIKOV, I. Y. et al. SeqyClean: A Pipeline for High-throughput Sequence Data Preprocessing. In: PROCEEDINGS OF THE 8TH ACM INTERNATIONAL CONFERENCE ON BIOINFORMATICS, COMPUTATIONAL BIOLOGY, AND HEALTH INFORMATICS 2017, Boston. **Anais...** Boston
- ZHOU, X. et al. The complete mitochondrial genomes of sixteen ardeid birds revealing the evolutionary process of the gene rearrangements. **BMC Genomics**, [s. l.], v. 15, n. 1, p. 1–9, 2014.

## V - ANEXO I

Manuscrito “Development and characterisation of 22 microsatellite markers,  
*de novo* assembly and preliminary gene annotation for the cattle egret  
(*Bubulcus ibis*) using next-generation sequencing”



1 **Development and characterisation of 22 microsatellite markers, *de novo***  
2 **assembly and preliminary gene annotation for the cattle egret (*Bubulcus***  
3 ***ibis*) using next-generation sequencing**

4

5 Fagner M. da Silva<sup>1\*</sup>, Talita A. Valdes<sup>1</sup>, Bruno A. Laroca<sup>1</sup>, Silvia N. Del Lama<sup>1</sup>

6

7 <sup>1</sup> Departamento de Genética e Evolução, Universidade Federal de São Carlos, CEP 13565-  
8 905, São Carlos, São Paulo, Brazil.

9 \*Corresponding author: Fagner Miguel da Silva; Address: Laboratório de Genética de Aves,  
10 Departamento de Genética e Evolução, Universidade Federal de São Carlos, Rodovia  
11 Washington Luís, km 235, CEP 13565-905, São Carlos, SP, Brazil; Tel: +55 16 3351 8391;  
12 E-mail: fagner.miguel.silva@gmail.com

13

14 **ABSTRACT**

15 The cattle egret (*Bubulcus ibis*) is a cosmopolitan bird that has attracted the attention of  
16 conservationists in new colonised areas due to its potential as an invasive species. In this  
17 study, we assembled the draft of the first reference genome for the cattle egret and isolated 22  
18 polymorphic microsatellite markers for the species using Illumina paired-end sequencing  
19 data. We also characterised the novel set of loci in two cattle egret populations to calculate  
20 allele frequencies and demonstrate the informativeness of the markers by estimating genetic  
21 diversity indices and the power of the loci in discriminating different kinship relationships. A  
22 total of 32,479 homologous genes were annotated in the reference genome. The number of  
23 alleles per locus varied between two and fourteen and expected heterozygosity ranged from  
24 0.03 to 0.85. Thirteen loci were highly informative (polymorphic informative content  $\geq 0.5$

25 and power of discrimination  $\geq 0.7$  in at least one population). The prospected panel of loci  
26 enabled estimating genetic diversity for the populations sampled that was in agreement with  
27 previously reported levels of genetic diversity for aquatic bird populations. The loci also  
28 exhibited greater power with regard to inferring relatedness than previous sets of  
29 microsatellite markers characterised in cattle egret populations. These 22 microsatellites are  
30 novel powerful molecular tools for future family and population genetic studies that can also  
31 assist in management of this species.

32 *Subjects: Molecular Biology, Conservation Biology, Genetics, Evolutionary Studies,*  
33 *Genomics*

34 *Keywords: Birds, simple sequence repeats, polymorphic loci, genetic diversity, reference*  
35 *genome, neutral nuclear marker*

36

## 37 **1. INTRODUCTION**

38 The cattle egret (*Bubulcus ibis*, Linnaeus, 1758; Pelecaniformes: Ardeidae) is a  
39 cosmopolitan bird species with high colonisation capacity and invasive potential (*Kushlan &*  
40 *Hafner, 2000; GISD, 2018*). Three subspecies are currently recognised as belonging to the  
41 monotypic genus *Bubulcus*: *B. ibis ibis* (Linnaeus, 1758), *B. ibis coromandus* (Boddaert,  
42 1783) and *B. ibis seychellarum* (Salomonsen, 1934) (*Vaurie, 1963; Kushlan & Hafner, 2000*).  
43 A rapid range expansion of both *B. ibis ibis* and *B. ibis coromandus* during the twentieth  
44 century led the cattle egret to occupy all continents except Antarctica (*Maddock & Geering,*  
45 *1994; Kushlan & Hancock, 2005*). This range expansion associated with the high interspecific  
46 competitiveness and predatory behaviour of the cattle egret has drawn the interest of  
47 researchers to the possible negative impacts generated by the species in newly colonised areas  
48 (*Burger, 1978; Avery & Tillman, 2005; Nunes et al., 2010; Meese, 2012*).

49 Molecular marker-based studies can contribute valuable information on the biology

50 of the cattle egret and the dynamics of its populations as well as lend support to control and  
51 management efforts in non-native areas occupied by the species. Previous studies on  
52 population genetics inferred the historical range expansion of cattle egret populations using  
53 sequences from both mitochondrial DNA and a single nuclear intron (*Moralez-Silva & Del*  
54 *Lama, 2014, Congrains et al., 2016*). However, due to the low levels of polymorphisms in the  
55 first microsatellite markers isolated for the cattle egret (*Campanini et al., 2012*), a single  
56 microsatellite marker-based study was published with this set of loci, focusing on the  
57 reproductive biology of the species (*Moralez-Silva & Del Lama, 2017*). To investigate the  
58 contemporary genetic structure of cattle egret populations and enable accurate inferences on  
59 the genetics of the mating system in this species, a greater number of informative  
60 microsatellite loci proved necessary (*Miño et al., 2017; Geronimo et al., submitted*).

61         The primary aim of the present study was to develop a novel set of microsatellite  
62 markers for the cattle egret using next-generation sequencing and characterise polymorphic  
63 loci in two populations of the species to report respective allele frequencies. We proved the  
64 usefulness of the novel set of loci for genetic population and parentage analyses by estimating  
65 genetic diversity indices in the populations sampled. We also proved the power of the loci in  
66 discriminating different kinship relationships. Furthermore, we presented a reference genome  
67 for the cattle egret obtained by *de novo* assembly and annotated in a public protein database.  
68 These novel microsatellite loci and the reference genome will enable future genetic studies on  
69 cattle egret as well as microsatellite transferability tests with other bird species.

70

## 71 **2. MATERIAL AND METHODS**

72         Four sets of blood samples ( $n = 96$ ) were collected from cattle egret nestlings in  
73 different breeding colonies and used in this study (Appendix 1, Supplemental Information).  
74 Genomic DNA from all samples was isolated using the DNeasy® Blood & Tissue kit

75 (Qiagen), following the manufacturer's instructions.

76 Genomic DNA (5 µg) from four individuals collected in different locations  
77 (Appendix 1, Supplemental Information) were used in the preparation of paired-end genomic  
78 libraries using the Nextera® DNA sample preparation kit (Illumina) and following the  
79 manufacturer's guide (Part # 15027987, Rev. B, October 2012, Illumina). Paired-end libraries  
80 with insert sizes ranging from 300 to 350 bp were sequenced in two lanes on an Illumina  
81 HiScan-SQ platform (Illumina, Inc., San Diego, CA, USA). Base-calling of the sequenced  
82 libraries and FASTQ file generation were performed with the CASAVA v1.8.2 pipeline  
83 (Illumina Inc., San Diego, CA, USA). Read quality was first evaluated using FastQC v0.11.5  
84 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Reads with more than 50% of  
85 quality bases  $\leq 24$  on a Phred scale, reads with adaptor and vector sequences detected by  
86 blasting in the NCBI UniVec database  
87 (<http://www.ncbi.nlm.nih.gov/VecScreen/UniVec.html>) and reads shorter than 65 bp were  
88 trimmed using SeqyClean v1.10.06 (<https://github.com/ibest/seqyclean>).

89 *De novo* assemblies were performed for the samples independently and with all  
90 samples together to obtain a reference genome using the Velvet v1.2.0.9 program (Zerbino &  
91 Birney, 2008; Schulz *et al.*, 2012). The assemblies were performed by first evaluating the  
92 distribution of the number of contigs by depth coverage using a customised script of the  
93 Bioconductor project (Gentleman *et al.*, 2014) and testing K-mer lengths between 31 and 71 bp  
94 at 10-bp intervals. According to the depth coverage that showed the largest number of contigs,  
95 the minimum *cov\_cutoff* value was set to three and a new round of assemblies was performed.  
96 Comparing the values of N50 length, number of contigs > 1000 bp and the maximum length of  
97 a contig, the best assemblies were obtained with k-mer values equal to 61 and 71. Preliminary  
98 gene annotation was carried out comparing the best assembly of the reference genome against  
99 the Swiss-Prot protein database (Boeckmann *et al.*, 2003).

100           Microsatellite motif detection in the assembled reference genome was performed using  
101 the QDD v3.0 pipeline (*Megléczy et al., 2014*). Pairs of primers flanking microsatellite sequences  
102 were designed using PRIMER3 v0.4.0 (*Koressaar & Remm, 2007; Untergasser et al., 2012*).  
103 Criteria were used for selecting target sequences with a maximum length of 250 bp, perfect  
104 microsatellites and at least ten repeats of dinucleotide motifs. Moreover, strict criteria were  
105 used to design primers with GC content less than 60%, length between 18 and 25 bp, hairpin  
106 loop probability of less than 2%, melting temperatures above 53 °C and maximum difference  
107 of 2 °C between melting temperatures of the forward and reverse sequences of each pair.  
108 Following the established criteria for the selection of target sequences and primer design, 28  
109 candidate loci were selected for initial amplification and confirmatory polymorphism tests  
110 (Table 1). In a previous study on the mating system of the cattle egret, six out of the 28  
111 candidate loci were characterised as polymorphic in one Brazilian population (*Miño et al.,*  
112 *2017*). In the present study, the remaining 22 loci were submitted to amplification and  
113 confirmatory polymorphism tests (Table 1).

114           Polymorphism tests of the candidate loci were performed using a pooled sample ( $n =$   
115  $31$ ) with individuals collected from four locations in Brazil and four locations on the African  
116 continent (Appendix 1, Supplemental Information). Frequencies in polymorphic loci were  
117 characterised in two populations of cattle egret sampled in Rio Claro ( $n = 32$ ; state of São  
118 Paulo, Brazil, BR, S 22° 20' 29" and W 47° 35' 37") and Cape Town ( $n = 29$ ; Western Cape  
119 province, South Africa, ZA, S 34° 03' 45.54" and E 18° 29' 43.5") (Appendix 1,  
120 Supplemental Information). Microsatellite loci were amplified by polymerase chain reaction  
121 (PCR) in mixes of 10  $\mu$ L containing 36 to 40 ng of DNA template, 0.2  $\mu$ M of the forward  
122 primers with a 5'-M13 (TGTAACGACGGCCAGT) tail (*Schuelke, 2000*), 0.8  $\mu$ M of the  
123 universal M13(-21) primer labelled with 6-FAM dye, 0.8  $\mu$ M of the reverse primers, 2.5 mM  
124 of MgCl<sub>2</sub>, 0.37 mM of each dNTP, 1x reaction buffer BD2 (Solis BioDyne) and 0.5 U of HOT

125 FIREPol® *Taq* DNA polymerase (Solis BioDyne). PCRs for amplification of locus Bi50 were  
126 optimised using 40 ng of DNA template. PCRs for amplification of loci Bi39 and Bi40 were  
127 optimised using 2.75 mM and 3.0 mM of MgCl<sub>2</sub>, respectively. PCR amplifications were  
128 performed in a Veriti™ 96-Well Thermal Cycler (Applied Biosystems®, Foster City, CA,  
129 USA), initially using a standard cycling profile: preheating at 95 °C for 15 min, thirty initial  
130 cycles (95 °C for 30 s, annealing temperatures for 45 s, and 72 °C for 30 s), ten final cycles  
131 (95 °C for 30 s, 53 °C for 45 s, and 72 °C for 30 s), and final polymerisation at 72 °C for 10  
132 min (Appendix 2, Supplemental Information). Standard cycling profiles were modified to  
133 increase the amplification success rate at some loci (Appendix 2, Supplemental Information).  
134 Amplification success was examined using standard gel electrophoresis with 2% agarose gels.  
135 Genotyping was performed in a MegaBACE®1000 automated sequencer (GE Healthcare,  
136 Bucks, UK) using a 500-bp Rox ladder size standard (The Gel Company, San Francisco, CA,  
137 USA) and genotypes were determined using FRAGMENT PROFILER® v1.2 (GE Healthcare,  
138 Bucks, UK). All microsatellite loci were genotyped in duplicates from independent  
139 amplifications in 10% of the individuals to minimize genotyping and scoring errors by  
140 confirming the scored genotypes.

141 Genotypes were checked for evidence of null alleles, scoring errors caused by stutter  
142 peaks and large allele dropout using MICRO-CHECKER v2.2.3 (*van Oosterhout et al., 2004*).  
143 Probabilities of linkage disequilibrium between loci ( $10^4$  MCMC dememorizations,  $10^3$   
144 batches and  $10^4$  iterations per batch) and departures from Hardy-Weinberg (HW) expectations  
145 ( $10^5$  MCMC dememorizations and  $10^3$  batches of length equal to  $10^3$ ) were determined in the  
146 *genepop* package v4.7 of R language (*Rousset, 2008; R Development Core Team, 2018*).  
147 Selective neutrality was verified using the  $F_{ST}$ -outlier method in BayeScan v2.1 (*Foll &*  
148 *Gaggiotti, 2008; Foll et al., 2010; Fischer et al., 2011*). The Holm-Bonferroni correction  
149 (Holm, 1979) was used to adjust critical  $p$ -values ( $\alpha = 0.05$ ) for multiple comparisons.

150 Polymorphic informative content (PIC) and allelic frequencies for each locus across all  
151 samples were computed using Cervus v3.0.7 (*Kalinowski et al., 2007*). The probability of  
152 identity (PI) for each locus was estimated using GENALEX v6.5.03 (*Peakall & Smouse,*  
153 *2012*) and the power of discrimination for each locus was calculated as  $1 - PI$ . Genetic  
154 diversity for each locus and population was estimated by computing the number of observed  
155 alleles ( $A_O$ ), observed heterozygosity ( $H_O$ ) and expected heterozygosity ( $H_E$ ) in the *diveRsity*  
156 package of R language (*Keenan et al., 2013; R Development Core Team, 2018*) and allelic  
157 richness ( $A_R$ ) was estimated in HP-RARE v1.0 (*Kalinowski, 2005*). Pairwise comparisons  
158 between genetic diversity values determined for the two cattle egret populations sampled were  
159 performed using the Wilcoxon signed-rank test in the R environment (*R Development Core*  
160 *Team, 2018*). The multilocus power (PWR) for the inference of relatedness was computed to  
161 estimate the power of the set of novel loci to distinguish full-siblings and half-siblings from  
162 unrelated individuals as well as full-siblings from half-siblings. PWR values were estimated  
163 in the KinInfor v2 program (*Wang, 2006*) based on the allelic frequencies of the Brazilian  
164 population sampled using  $10^3$  simulated pairs of genotypes for each candidate relationship  
165 tested, with the statistical significance level set to 0.05.

166

### 167 3. RESULTS AND DISCUSSION

168       Among the 1,457,233 raw reads produced using the Illumina sequencing platform,  
169 approximately 574,727 paired-end cleaned sequences were generated and deposited in the  
170 International Nucleotide Sequence Databases under accession numbers XYZ. The reference  
171 genome was composed of 462,510 paired-end cleaned sequences with lengths ranging from  
172 140 bp to 24,205 bp. Using the reference genome of the cattle egret, a total of 32,479  
173 homologous genes were annotated and a total of 242 microsatellite loci were successfully  
174 identified (Appendix 3, Supplemental Information). In the present study, all 22 loci selected

175 were successfully amplified and were polymorphic among the individuals of the pooled  
176 sample as well as between the populations sampled in Brazil and South Africa. Overall, 138  
177 different alleles were detected and the number of alleles per locus ranged from two (Bi40) to  
178 fourteen (Bi56) (Table 1). The total amplification success rate and detected levels of  
179 polymorphisms demonstrated that combining next-generation sequencing and strict criteria  
180 for primer design led to an improvement in the isolation of microsatellite loci compared to  
181 traditional methods. In a previous isolation of microsatellite loci for the cattle egret using  
182 enriched genomic libraries with biotinylated probes, 32 pairs of primers were drawn from  
183 sequences of 432 recombinant clones and only 11 polymorphic loci were found, with a total  
184 of 28 different alleles (*Campanini et al., 2012*).

185         No evidence was found for significant linkage disequilibrium (corrected  $p$ -values  
186 ranged from 0.17 to 1.00) or deviations from either HW expectations (corrected  $p$ -values  
187 ranged from 0.20 to 1.00) or selective neutrality (corrected  $p$ -values ranged from 0.78 to 0.91)  
188 at any locus in the sampled populations. Moreover, no loci showed evidence of null alleles,  
189 scoring errors due to stuttering or large allele dropout. According to the polymorphic  
190 informative content and the power of discrimination (Table 2), the set of 22 loci included  
191 thirteen highly informative loci with PIC values  $\geq 0.5$  and PD values  $\geq 0.7$  in at least one  
192 population sampled (Bi31, Bi39, Bi44, Bi46, Bi47, Bi49, Bi50, Bi53, Bi54, Bi55, Bi56, Bi57,  
193 and Bi60). Six loci (Bi37, Bi40, Bi42, Bi48, Bi51, and Bi59) had frequency  $> 0.7$  in a single  
194 allele in both the Brazilian and South African populations (Appendix 4, Supplemental  
195 Information). Despite displaying one highly frequent allele, loci Bi48 and Bi59 exhibited  
196 moderate informativeness (PIC ranged from 0.25 to 0.35 and PD ranged from 0.45 to 0.59).  
197 Overall mean PIC and PD values (Table 2) demonstrated that the set of 22 loci described  
198 herein were more informative both in comparison to the set of 11 loci reported by *Campanini*  
199 *et al.* (2012) (mean PIC value =  $0.27 \pm 0.14$ ) and the set of six novel loci reported by *Miño et*



200 *al.* (2017) (mean PD value =  $0.56 \pm 0.06$ ).

201 Genetic diversity indices did not differ significantly among the populations studied  
202 ( $p$ -values  $> 0.05$  for all comparisons by using the Wilcoxon signed-rank test). Average  
203 observed heterozygosity ( $H_O$ ) over all loci was equal to  $0.56 (\pm 0.23)$ , the average level of  
204 expected heterozygosity ( $H_E$ ) ranged from  $0.54 (\pm 0.23)$  to  $0.56 (\pm 0.20)$  and the average  
205 allele richness ( $A_R$ ) ranged from  $4.59 (\pm 2.51)$  to  $5.07 (\pm 2.24)$  (Table 2). The genetic diversity  
206 indices determined herein were significantly higher than those calculated using the values  
207 reported by *Campanini et al.* (2012) (mean  $H_O = 0.30 \pm 0.18$  and mean  $H_E = 0.33 \pm 0.18$ ,  $p$ -  
208 values  $< 0.05$  for all comparisons by using the Wilcoxon signed-rank test). In contrast, the  
209 genetic diversity indices determined herein did not differ significantly from those reported by  
210 *Miño et al.* (2017) (mean  $H_O = 0.53 \pm 0.07$  and mean  $H_E = 0.51 \pm 0.06$ ,  $p$ -values  $> 0.05$  for all  
211 comparisons by using the Wilcoxon signed-rank test). The expected heterozygosity values  
212 determined herein were in agreement with the levels of genetic diversity compiled by *Eo et al.*  
213 (2011) for aquatic bird populations using microsatellite information (mean  $H_E$  per locus per  
214 population =  $0.56 \pm 0.21$ ). Moreover, 28 private alleles were detected and all fit the mutation  
215 pattern expected for dinucleotide loci: 13 alleles were detected only in the pooled sample,  
216 while the number of exclusive alleles in the South African and the Brazilian populations were  
217 12 and 3, respectively (Appendix 5, Supplemental Information). The larger number of private  
218 alleles in the South African population than the Brazilian population is in line with  
219 expectations, since the Americas were occupied by the cattle egret during a more recent  
220 colonisation, possibly with fewer migration events (*Moralez-Silva & Del Lama, 2014*;  
221 *Congrains et al., 2016*).

222 Regarding the multilocus power (PWR) for the inference of relatedness calculated  
223 using simulations based on the allelic frequencies of the Brazilian population sampled, the full  
224 power to discriminate between full siblings and unrelated individuals was achieved by

225 employing eight loci (Fig. 1). To discriminate half siblings from full siblings, the full power  
226 of inference was achieved by employing 19 out of the 22 novel loci and the highest power of  
227 inference for the comparisons between half siblings and unrelated individuals was 0.88 using  
228 the total set of novel loci (Fig. 1). In an earlier investigation, lower PWR values were  
229 computed by employing 11 loci both for comparisons between full siblings and half siblings  
230 (multilocus PWR between 0.45 and 0.50) as well as between half siblings and unrelated  
231 individuals (multilocus PWR between 0.40 and 0.45) (*Miño et al., 2017*).

232         The panel of 22 novel microsatellite loci for the cattle egret described in this study are  
233 powerful molecular tools that can assist future investigations on population genetics and  
234 increase the resolution of relatedness inferences. Moreover, this novel set of loci can be used  
235 for future transferability tests in other closely related bird species. Microsatellite loci  
236 previously described for the cattle egret have demonstrated a high success rate of cross-  
237 amplification among other species of Pelecaniformes (*Campanini et al., 2012; Souza, Miño &*  
238 *Del Lama, 2012*). Here, we proved the usefulness of this highly polymorphic set of loci for  
239 genetic diversity evaluations. The PWR values reported herein also demonstrate the potential  
240 of the novel microsatellite markers to support family studies by improving genetic  
241 discrimination between different kinship relationships. These loci also can contribute to the  
242 genetic characterisation of cattle egret populations in colonised areas and assist in decision  
243 making for the management of this invasive species, particularly on oceanic islands, where its  
244 negative impact on populations of native species has been documented (*Silva Jr., Peres Jr. &*  
245 *Sazima, 2005; Barbosa-Filho et al., 2009*).

246

## 247 **ACKNOWLEDGMENTS**

248         The authors are grateful to former students and staff of the Birds Genetics Laboratory  
249 for collecting samples. We thank Dr. Sónia Andrade for clarifications regarding the assembly

250 of the reference genome.

251

## 252 **ADDITIONAL INFORMATION AND DECLARATIONS**

### 253 **Funding**

254 This study was supported by the Brazilian fostering agency *Fundação de Amparo à Pesquisa*  
255 *do Estado de São Paulo* (FAPESP) through research grants to S.N.D.L. (2010/52315-7,  
256 2016/01673-7). F.M.S. was supported by a scholarship from *Conselho Nacional de*  
257 *Desenvolvimento Científico e Tecnológico* (CNPq, Brazil, 142000/2015-9).

### 258 **Grant Disclosures**

259 The following grant information was disclosed by the authors:

260 *Fundação de Amparo à Pesquisa do Estado de São Paulo* (FAPESP, Brazil): 2010/52315-7,  
261 2016/01673-7.

262 *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq, Brazil):  
263 142000/2015-9.

### 264 **Competing Interests**

265 The authors declare there are no competing interests.

### 266 **Author Contributions**

267 All authors designed the study and worked on sample collection, selection of candidate loci,  
268 and primer design. F.M.S. and S.N.D.L. wrote the manuscript. F.M.S. conducted the genetic  
269 and statistical analyses.

### 270 **Field Study Permissions**

271 The following information was supplied relating to field study approvals (i.e., approving  
272 body and any reference numbers):

273 All blood samples from Brazil were collected under legal permission from the Brazilian  
274 environmental protection agency (IBAMA); permit numbers: 20295-2/2010, 30023-3/2011,

275 12437-2/2013, 30023-4/2013, and 30023-5/2013.

276 All blood samples from Senegal, Ghana, Nigeria and South Africa were collected under legal  
 277 permissions and transported to Brazil under permission of the Brazilian environmental  
 278 protection agency (IBAMA); import permit codes: 11BR007258/DF, 11BR007010/DF,  
 279 11BR007199/DF, and DV.0890/2011-SP.

## 280 **DNA Deposition**

281 The following information was supplied regarding the deposition of related data:

282 Sequence data is archived at NCBI Sequence Read Archive; accession number: XYZ.

283 Microsatellite primer sequences are archived in NCBI GenBank; accession numbers: XYZ.

## 284 **Data Availability**

285 The following information was supplied regarding data availability:

286 The dataset used in this study is available as Dataset S1.

## 287 **Supplemental Information**

288 Supplemental information for this article can be found online at XYZ

289

## 290 **REFERENCES**

291

292

293 Avery M, Tillman E. 2005. Alien birds in North America – challenges for wildlife managers. In: *Wildlife*  
 294 *Damage Management Conferences -Proceedings*. 81–89.

295 Beaumont MA, Nichols RA. 1996. Evaluating loci for use in the genetic analysis of population  
 296 structure. *Proceedings of the Royal Society B: Biological Sciences* 263:1619–1626. DOI:  
 297 10.1098/rspb.1996.0237.

298 Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E, Martin MJ, Michoud K,  
 299 O’Donovan C, Phan I, Pilbout S, Schneider M. 2003. The SWISS-PROT protein knowledgebase  
 300 and its supplement TrEMBL in 2003. *Nucleic Acids Research* 31:365–370. DOI:  
 301 10.1093/nar/gkg095.

302 Burger J. 1978. Competition between Cattle Egrets and Native North American Herons, Egrets, and  
 303 Ibises. *The Condor* 80:15. DOI: 10.2307/1367786.

304 Campanini EB, Sanches A, Hatanaka T, Del Lama SN. 2012. Isolation and characterization of 11  
 305 microsatellite loci from cattle egret (*Bubulcus ibis*) and cross-amplification in other Ardeidae  
 306 species. *Conservation Genetics Resources* 4:707–709. DOI: 10.1007/s12686-012-9627-4.

307 Congrains C, Miranda EA, Henry DAW, Abalaka J, Manu SA, Sá J, Holbech LH, Gbogbo F, Cumming GS,  
 308 Monteiro H, Carvalho AF, Diop MS, Rocha CD, Lama SN Del. 2016. Genetic and paleomodelling  
 309 evidence of the population expansion of the cattle egret *Bubulcus ibis* in Africa during the

- 310 climatic oscillations of the Late Pleistocene. *Journal of Avian Biology* 47:846–857. DOI:  
 311 10.1111/jav.00972.
- 312 Eo SH, Doyle JM, DeWoody JA. 2011. Genetic diversity in birds is associated with body mass and  
 313 habitat type. *Journal of Zoology* 283:220–226. DOI: 10.1111/j.1469-7998.2010.00773.x.
- 314 Foll M, Gaggiotti O. 2008. A genome-scan method to identify selected loci appropriate for both  
 315 dominant and codominant markers: A Bayesian perspective. *Genetics* 180:977–993. DOI:  
 316 10.1534/genetics.108.092221.
- 317 Gentleman RC, Carey VJ, Bates DM, Bolstad BM, Dettling M, Dudoit S, Ellis B, Gautier L, Ge YC, Gentry  
 318 J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry RA, Leisch F, Li C, Maechler M, Rossini AJ,  
 319 Sawitzki G, Smith C, Smyth G, Tierney L, Yang JYH, Zhang JH. 2004. Bioconductor: open software  
 320 development for computational biology and bioinformatics. *Genome Biology* 5:R80. DOI:  
 321 10.1186/gb-2004-5-10-r80.
- 322 Hafner H, Lansdown R V, Kushlan JA, Butler RW, Custer TW, Davidson IJ, Erwin RM, Hancock JA, Lyles  
 323 AM, Maddock M, Marion L, Morales G, Mundkur T, Perennou C, Pineau O, Turner D, Ulenaers P,  
 324 VanVessem J, Young L. 2000. Conservation of herons. In: Kushlan JA, Hafner H orgs. *Heron*  
 325 *Conservation*. San Diego, CA: Academic Press,.
- 326 Holm S. 1979. A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of*  
 327 *Statistics* 6:65–70.
- 328 Kalinowski ST. 2005. HP-RARE 1.0: A computer program for performing rarefaction on measures of  
 329 allelic richness. *Molecular Ecology Notes* 5:187–189. DOI: 10.1111/j.1471-8286.2004.00845.x.
- 330 Kalinowski ST, Taper ML, Marshall TC. 2007. Revising how the computer program CERVUS  
 331 accommodates genotyping error increases success in paternity assignment. *Molecular Ecology*  
 332 16:1099–1106. DOI: 10.1111/j.1365-294X.2007.03089.x.
- 333 Keenan K, Mcginnity P, Cross TF, Crozier WW, Prodöhl PA. 2013. DiveRsity: An R package for the  
 334 estimation and exploration of population genetics parameters and their associated errors.  
 335 *Methods in Ecology and Evolution* 4:782–788. DOI: 10.1111/2041-210X.12067.
- 336 Koressaar T, Remm M. 2007. Enhancements and modifications of primer design program Primer3.  
 337 *Bioinformatics* 23:1289–1291. DOI: 10.1093/bioinformatics/btm091.
- 338 Kushlan JA, Hancock JA. 2005. *14. The Herons (Ardeidae)*. New York: Oxford University Press.
- 339 Maddock M, Geerin D. 1994. Range expansion and migration of the cattle egret. *Ostrich* 25:191–203.
- 340 Meese RJ. 2012. Cattle egret predation causing reproductive failures of nesting tricolored blackbirds.  
 341 *California Fish and Game* 98:47–50.
- 342 Meglécz E, Costedoat C, Pech N, Dubut V, Martin J-F, Malausa T, Gilles A. 2009. QDD: a user-friendly  
 343 program to select microsatellite markers and design primers from large sequencing projects.  
 344 *Bioinformatics* 26:403–404. DOI: 10.1093/bioinformatics/btp670.
- 345 Miño CI, Valdes TA, Morales-Silva E, de Souza ED, Cortiço Corrêa Rodrigues VL, Del Lama SN. 2017.  
 346 Use of noninvasive ‘bug-eggs’ to enable comparative inferences on genetic mating system with  
 347 and without parental information: A study in a cattle egret colony. *Plos One* 12:e0183153. DOI:  
 348 10.1371/journal.pone.0183153.
- 349 Morales-Silva E, Del Lama SN. 2014. Colonization of Brazil by the cattle egret (*Bubulcus ibis*) revealed  
 350 by mitochondrial DNA. *NeoBiota* 21:49–63. DOI: 10.3897/neobiota.21.4966.
- 351 Morales-Silva E, Del Lama SN. 2017. Multiple Females Laying in Cattle Egret ( *Bubulcus ibis* ) Clutches:  
 352 Behavioral and Genetic Evidence. *Waterbirds* 40:403–409. DOI: 10.1675/063.040.0412.
- 353 Nunes MFC, Barbosa-filho RC, Roos AL, Mestre ALM. 2010. The Cattle Egret ( *Bubulcus ibis* ) on  
 354 Fernando de Noronha Archipelago : history and population trends. *Revista Brasileira de*  
 355 *Ornitologia* 18:315–327.
- 356 Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P. 2004. MICRO-CHECKER: Software for  
 357 identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*  
 358 4:535–538. DOI: 10.1111/j.1471-8286.2004.00684.x.
- 359 Peakall R, Smouse PE. 2012. GenALEx 6.5: Genetic analysis in Excel. Population genetic software for  
 360 teaching and research-an update. *Bioinformatics* 28:2537–2539. DOI:  
 361 10.1093/bioinformatics/bts460.

- 362 Rousset F. 2008. GENEPOP'007: A complete re-implementation of the GENEPOP software for  
363 Windows and Linux. *Molecular Ecology Resources* 8:103–106. DOI: 10.1111/j.1471-  
364 8286.2007.01931.x.
- 365 Schuelke M. 2000. An economic method for the fluorescent labeling of PCR fragments - A poor man's  
366 approach to genotyping for research and high-throughput diagnostics. *Nature Biotechnology*  
367 18:233–234.
- 368 Schulz MH, Zerbino DR, Vingron M, Birney E. 2012. Oases: Robust de novo RNA-seq assembly across  
369 the dynamic range of expression levels. *Bioinformatics* 28:1086–1092. DOI:  
370 10.1093/bioinformatics/bts094.
- 371 Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. 2012. Primer3-new  
372 capabilities and interfaces. *Nucleic Acids Research* 40:1–12. DOI: 10.1093/nar/gks596.
- 373 Wang J. 2006. Informativeness of genetic markers for pairwise relationship and relatedness  
374 inference. *Theoretical Population Biology* 70:300–321. DOI: 10.1016/j.tpb.2005.11.003.
- 375 Zerbino DR, Birney E. 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn  
376 graphs. *Genome Research* 18:821–829. DOI: 10.1101/gr.074492.107.
- 377 GISD. Global Invasive Species Database. 2017. Disponível em  
378 <<http://www.iucngisd.org/gisd/search.php> on 09-03-2019> Acesso em: 05 de junho de 2017.
- 379 Silva-Jr JM, Péres-Jr AK, Sazima I. 2005. *Euprepis atlanticus* (Noronha Skink) predation. *Herpetological*  
380 *Review* 36:62–63.
- 381 Barbosa-Filho RCB, Souza AEAB, Freitas GL, Nunes MFC, Souza EA, Zappelini-Filho D. 2009. A  
382 garça-vaqueira (*Bubulcus ibis* Linnaeus, 1758) e o atobá-de-pé-vermelho (*Sula sula* Linnaeus,  
383 1766) no Arquipélago de Fernando de Noronha: uma abordagem ecológica comparativa.  
384 *Ornithologia*, 3:101–114.
- 385 R Development Core Team v3.4.4. 2018. R: A language and environment for statistical computing.  
386 Viena: R Foundation for Statistical Computing. Disponível em:<<http://www.R-project.org/>>.  
387 Acesso em: 01 mar. 2018.

388 **Table 1** Microsatellite loci characteristics. Forward (F) and reverse (R) primer sequences, repeat motif, annealing temperature (AT), number of  
 389 observed alleles ( $A_0$ ), size range of amplified fragments (Range), GenBank Accession Number (GenBank) and the source of the first use were  
 390 shown for the 28 microsatellite loci selected for the cattle egret.

Locus	Primer sequence (5'-3')	Repeat motif	AT (°C)	$A_0$	Range (bp)	GenBank	Source
Bi31	F: CCTTTCAGTGTAATACCAACCG R: AAAGCAGCATCCGCAGTAAT	(AC) <sub>10</sub>	55	6	166-176		This study
Bi32	F: GCTAAATTGCAGCTGTGGGA R: GGCTGTGGTCAGAGGAGTGA	(AG) <sub>12</sub>	59	6	164-198		<i>Miño et al., 2017</i>
Bi33	F: GGCAATGAAGACTGAAATGTGC R: GGACTCGATGATCTCAAAGGG	(AC) <sub>10</sub>	57	10	175-203		<i>Miño et al., 2017</i>
Bi34	F: AAActCTGCAACTCGCACAAA R: TGACACGATACCAACACCCA	(TG) <sub>11</sub>	55	7	93-111		<i>Miño et al., 2017</i>
Bi36	F: GTGTGGCCGCTCTTCTTTC R: AGCCTGTGGGTTTACTGCAC	(AC) <sub>13</sub>	56	7	82-102		<i>Miño et al., 2017</i>
Bi37	F: TGAATTCAGGAGGAAAGCAGG R: GGTGTGGCCTGCATGTTTAC	(GT) <sub>10</sub>	57	3	126-130		This study
Bi38	F: TGCTGTTTAAATGCTGGTGC R: GTACGAGATCCGAACCCGA	(GA) <sub>10</sub>	55	4	182-190		<i>Miño et al., 2017</i>
Bi39	F: ACCACCTGATTCCACTTCCC R: GGCGCCCTGAACAGATTTAC	(TC) <sub>10</sub>	56	7	202-220		This study
Bi40	F: GCTGAATCCTCGTGTGTTGGT R: CCTCCTTCACCTTGCGATTT	(AG) <sub>10</sub>	56	2	96-98		This study
Bi42	F: CCTGGAAGCCAAGGTGATTT R: TCCTTTGCACTAACAAGCCC	(TG) <sub>11</sub>	58	5	114-126		This study
Bi43	F: AAATTGGTCAGCAGTGGGTG R: TCCACCACATGCCTCTGTTA	(CA) <sub>10</sub>	54	5	122-130		<i>Miño et al., 2017</i>
Bi44	F: GAGCATTTCCACCTCCAGC R: GTGGATGGAAGCATCGAAA	(TA) <sub>10</sub>	55	3	142-146		This study
Bi45	F: GGTTTAATTTGCAGAGCAATGAG R: TTCGTATTTGCTGCCCAAAC	(CA) <sub>12</sub>	57	3	134-138		This study
Bi46	F: GGAGACTCCGGGAAGCCT R: CCGTCGGGGTTGTGCAAT	(TG) <sub>10</sub>	60	12	136-168		This study

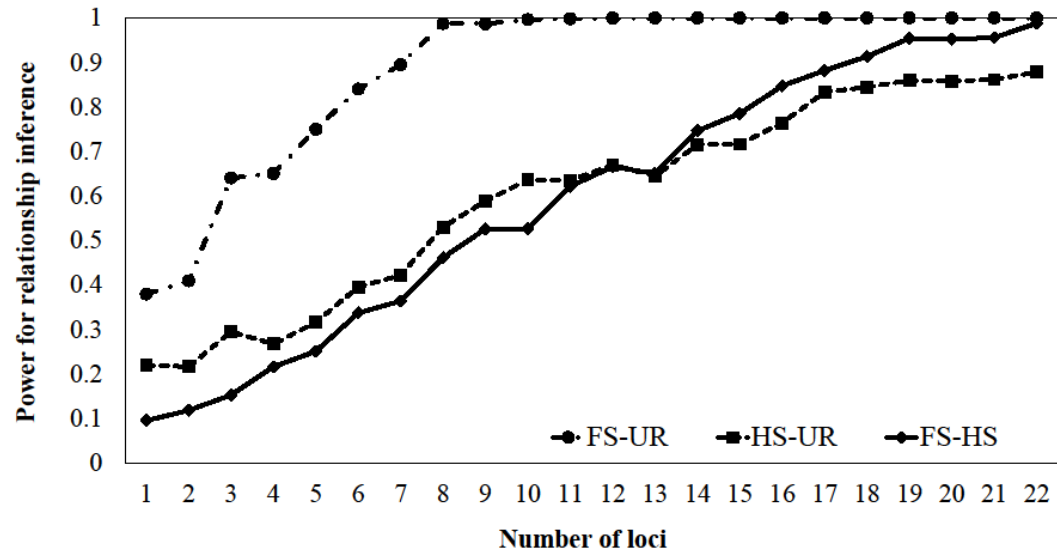
392 **Cont. Table 1** Microsatellite loci characteristics. Forward (F) and reverse (R) primer sequences, repeat motif, annealing temperature (AT),  
 393 number of observed alleles ( $A_0$ ), size range of amplified fragments (Range), GenBank Accession Number (GenBank) and the source of the first  
 394 use were shown for the 28 microsatellite loci selected for the cattle egret.

Locus	Primer sequence (5'-3')	Repeat motif	AT (°C)	Na	Size range (bp)	GenBank	Source
Bi47	F: ATCCCGCCTAGAGATGCCTT R: TCCAGGCTAACTTCTGCATGA	(GT) <sub>11</sub>	60	5	146-154		This study
Bi48	F: ATGGAAAAGTGCTTAGCCTGC R: TGACAGAGGAACCAAGAGCT	(AG) <sub>11</sub>	60	4	170-180		This study
Bi49	F: GACCAATAATCGCCTTTGGA R: GCACTGAGAGCGATTACAAA	(GA) <sub>14</sub>	58	8	158-172		This study
Bi50	F: AGCGCTAACCCACGGATG R: CTGCAAGCGTTGGGTTCTC	(CA) <sub>11</sub>	61-60	12	116-154		This study
Bi51	F: TTGCATACACGCTACTGTGC R: ACATAGGAGCAGGGCGTAAA	(TG) <sub>10</sub>	60	4	174-184		This study
Bi52	F: AGCTTGTGAGATATTTGTGGCA R: ATTTGCTGGGTGTGGAGGTT	(TG) <sub>10</sub>	60	5	150-158		This study
Bi53	F: GCAAAGTTGTAAATACGAAGCG R: ACGCTCACCCACCTACAAG	(GT) <sub>12</sub>	58	6	156-166		This study
Bi54	F: CGCGCCACTGATTTGATACA R: ACTTACAGTGAACGCCTAGC	(TG) <sub>13</sub>	60	11	112-132		This study
Bi55	F: AGTCTGAATCTGGGTGTGGG R: ATCTTCTGGCTGAGCGTGT	(AC) <sub>13</sub>	61-60	11	180-206		This study
Bi56	F: CCCAGCGGCAGATATCTACA R: ACAAGTGCCCCTTCTCCAC	(TC) <sub>15</sub>	61-60	14	150-190		This study
Bi57	F: GGTGACCCAAGATGTTAGCAC R: ACAAACCTCTACTTCTTCCTGTG	(AG) <sub>15</sub>	60	6	162-172		This study
Bi58	F: TTGTCCCCTTAGCCTGTCTG R: GGTCACGTGGAAATGATGGAG	(TG) <sub>10</sub>	60	4	148-156		This study
Bi59	F: GAGCATCGATCCATGGGG R: CAGAGTTTGTGTCCCCTCCA	(CA) <sub>10</sub>	61-60	3	150-154		This study
Bi60	F: ACACCTCACACTGCATTAAGT R: CCACAGCGGGTAAATTAGGATT	(AC) <sub>13</sub>	57-56	4	108-118		This study



396 **Table 2** Genetic diversity and informativeness indices for cattle egret populations sampled. Polymorphic information content (PIC), power of  
 397 discrimination (PD), observed heterozygosity ( $H_O$ ), expected heterozygosity ( $H_E$ ) and allelic richness ( $A_R$ ) were determined for the South African  
 398 and Brazilian cattle egret populations. Mean values and standard errors (SD) are given for each population.

Locus	PIC		PD		$H_O$		$H_E$		$A_R$	
	ZA	BR	ZA	BR	ZA	BR	ZA	BR	ZA	BR
Bi31	0.54	0.57	0.78	0.80	0.52	0.81	0.59	0.65	4.00	5.00
Bi37	0.03	0.06	0.07	0.12	0.03	0.06	0.03	0.06	1.97	2.00
Bi39	0.77	0.66	0.93	0.87	0.79	0.78	0.82	0.72	6.97	5.00
Bi40	0.12	0.31	0.23	0.55	0.14	0.34	0.13	0.40	2.00	2.00
Bi42	0.24	0.30	0.43	0.52	0.28	0.25	0.26	0.33	4.93	3.00
Bi44	0.48	0.50	0.72	0.74	0.45	0.56	0.58	0.60	3.00	3.00
Bi45	0.46	0.33	0.71	0.57	0.45	0.38	0.53	0.40	3.00	3.00
Bi46	0.70	0.81	0.89	0.95	0.79	0.84	0.75	0.85	7.90	9.00
Bi47	0.51	0.44	0.75	0.69	0.72	0.47	0.61	0.55	3.00	4.00
Bi48	0.34	0.34	0.57	0.57	0.38	0.44	0.38	0.37	4.00	4.00
Bi49	0.65	0.46	0.87	0.71	0.79	0.50	0.70	0.51	6.00	4.00
Bi50	0.67	0.73	0.88	0.91	0.79	0.91	0.73	0.78	10.00	7.00
Bi51	0.22	0.21	0.40	0.40	0.28	0.22	0.25	0.25	2.97	2.00
Bi52	0.45	0.38	0.70	0.63	0.59	0.56	0.56	0.49	4.90	3.00
Bi53	0.70	0.59	0.89	0.81	0.76	0.84	0.75	0.66	5.00	4.00
Bi54	0.63	0.67	0.85	0.88	0.59	0.72	0.67	0.71	8.90	9.00
Bi55	0.77	0.62	0.93	0.85	0.86	0.69	0.80	0.67	7.96	7.00
Bi56	0.77	0.74	0.93	0.92	0.79	0.81	0.81	0.79	10.00	9.00
Bi57	0.71	0.66	0.90	0.87	0.72	0.75	0.76	0.71	5.00	5.00
Bi58	0.45	0.42	0.70	0.67	0.55	0.47	0.52	0.47	3.00	4.00
Bi59	0.35	0.25	0.59	0.45	0.48	0.31	0.40	0.28	3.00	3.00
Bi60	0.66	0.56	0.87	0.79	0.66	0.56	0.73	0.62	4.00	4.00
Mean	0.51	0.48	0.71	0.69	0.56	0.56	0.56	0.54	5.07	4.59
SD ±	0.22	0.20	0.24	0.21	0.23	0.23	0.23	0.20	2.51	2.24



**Fig. 1** Informativeness of the novel set of microsatellites isolated for the cattle egret in inferring pairwise relationships. Power index for relatedness inference (PWR) calculated using simulations based on the allelic frequencies of the Brazilian cattle egret population were shown as a function of the number of loci used to discriminate between

## VI. ANEXO II - SUPPLEMENTAL INFORMATION FOR ONLINE PUBLICATION

### **Development and characterisation of 22 microsatellite markers, *de novo* assembly and preliminary gene annotation for the cattle egret (*Bubulcus ibis*) using next-generation sequencing**

Fagner M. da Silva<sup>1\*</sup>, Talita A. Valdes<sup>1</sup>, Bruno A. Laroca<sup>1</sup>, Silvia N. Del Lama<sup>1</sup>

<sup>1</sup> Departamento de Genética e Evolução, Universidade Federal de São Carlos, CEP 13565-905, São Carlos, São Paulo, Brazil.

\**Corresponding author*: Fagner Miguel da Silva; Address: Laboratório de Genética de Aves, Departamento de Genética e Evolução, Universidade Federal de São Carlos, Rodovia Washington Luís, km 235, CEP 13565-905, São Carlos, SP, Brazil; Tel: +55 16 3351 8391; E-mail: fagner.miguel.silva@gmail.com

**Appendix 1** Cattle egret blood sampling information used in this study. The information provided were country and region, colony names, geographical coordinates, number of individuals sampled (*n*), collection years and stage of the study in which the sample was used.

Country (Region)	Colony	Geographical Coordinates	<i>n</i>	Collection year	Stage	
Senegal (Thiès Region)	Somone Lagoon Reserve	N 14° 29' 51.66" W 17° 4' 55.86"	1	2011	Illumina paired-end sequencing	
South Africa (Western Cape Province)	Paarl	S 33° 41' 3.06" E 18° 59' 1.86"	1	2011		
Brazil (Pará State)	Marajó Island	S 0° 30' 19.5" W 49° 13' 46.08"	1	2012		
Brazil (Rio Grande do Sul State)	Eldorado do Sul	S 30° 1' 50.84" W 51° 31' 47.78"	1	2009		
Brazil (São Paulo State)	Lago Azul Park	S 22° 23' 30.83" W 47° 33' 51.02"	4	2014	Polymorphism tests	
Brazil (Pernambuco State)	Garanhuns	S 8° 52' 0" W 36° 28' 0"	4	2010		
Brazil (Rio Grande do Sul State)	Eldorado do Sul	S 30° 1' 50.84" W 51° 31' 47.78"	3	2009		
Brazil (Pará State)	Marajó Island	S 0° 30' 19.5" W 49° 13' 46.08"	4	2012		
Senegal (Dakar Region)	Hann Park and Zoo	N 14° 43' 39.18" W 17° 25' 53.70"	4	2011		
Ghana (Greater Accra Region)	Korle Lagoon	N 5° 32' 17.51" W 0° 13' 12.14"	4	2011		
South Africa (Western Cape Province)	Rondevlei Nature Reserve	S 34°3' 45.54" E 18° 29' 43.5"	4	2011		
Nigeria (Plateau State)	Fobur kasa	N 9° 51' 51.62" E 9° 2' 44.95"	4	2011		
South Africa (Western Cape Province)	Rondevlei Nature Reserve	S 34°3' 45.54" E 18° 29' 43.5"	29	2011		Loci characterization
Brazil (São Paulo State)	Lago Azul Park	S 22° 23' 30.83" W 47° 33' 51.02"	32	2014		

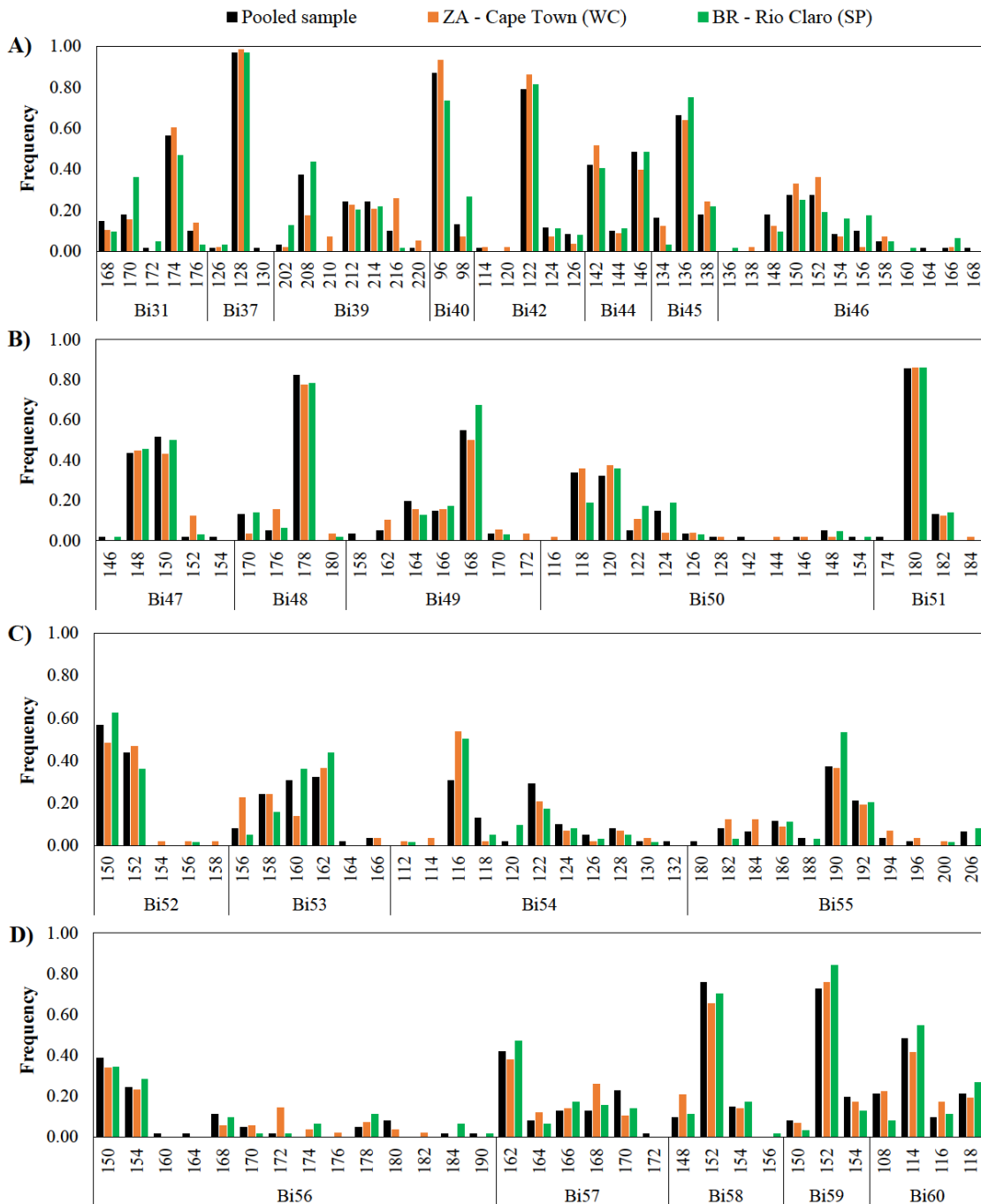
**Appendix 2** Additional information for polymerase chain reactions (PCR) of microsatellites amplified in cattle egret samples.

---

Locus	Modification from the standard cycling profile
Bi31	The number of initial cycles was changed from 30 to 35
Bi37	The number of initial cycles was changed from 30 to 37
Bi39	None
Bi40	The number of initial cycles was changed from 30 to 35
Bi42	The number of initial cycles was changed from 30 to 35
Bi44	The number of initial cycles was changed from 30 to 35
Bi45	The number of initial cycles was changed from 30 to 37
Bi46	None
Bi47	None
Bi48	None
Bi49	None
Bi50	The number of initial cycles was changed from 30 to 20, and ten cycles were added as an earlier step with decreasing annealing temperatures from 61°C to 60°C by touchdown approach
Bi51	None
Bi52	None
Bi53	None
Bi54	None
Bi55	The number of initial cycles was changed from 30 to 20, and ten cycles were added as an earlier step with decreasing annealing temperatures from 61°C to 60°C by touchdown approach
Bi56	The number of initial cycles was changed from 30 to 20, and ten cycles were added as an earlier step with decreasing annealing temperatures from 61°C to 60°C by touchdown approach
Bi57	None
Bi58	None
Bi59	The number of initial cycles was changed from 30 to 20, and ten cycles were added as an earlier step with decreasing annealing temperatures from 61°C to 60°C by touchdown approach
Bi60	The number of initial cycles was changed from 30 to 20, and ten cycles were added as an earlier step with decreasing annealing temperatures from 57°C to 56°C by touchdown approach

---

**Appendix 4** Allele frequencies for the 22 novel microsatellite markers isolated for the cattle egret. Baseline allele frequencies are shown for the loci characterized in the group of pooled samples ( $n = 31$ ) used in polymorphism tests and the natural populations of cattle egret sampled in South Africa (ZA, Cape Town, Western Cape province,  $n = 29$ ) and Brazil (BR, Rio Claro, São Paulo State,  $n = 32$ ). In order, the loci shown are: A) Bi31, Bi37, Bi39, Bi40, Bi42, Bi44, Bi45, B) Bi47, Bi48, Bi49, Bi50, Bi51, C) Bi52, Bi53, Bi54, Bi55, D) Bi56, Bi57, Bi58, Bi59 and Bi60.



**Appendix 5** Private alleles detected in cattle egret samples analysed in this study. Private alleles and their respective locus and frequencies are given for the pooled sample (PS) used in the polymorphism tests, as well as for the cattle egret populations of South Africa (ZA) and Brazil (BR) used in the characterization of the loci.

Group of samples	Locus	Allele	Frequency
PS	Bi46	164	0.016
PS	Bi46	168	0.016
PS	Bi47	154	0.016
PS	Bi49	158	0.032
PS	Bi50	142	0.016
PS	Bi51	174	0.016
PS	Bi53	164	0.016
PS	Bi54	132	0.016
PS	Bi55	180	0.016
PS	Bi56	160	0.016
PS	Bi56	164	0.016
PS	Bi57	172	0.016
PS	Bi37	130	0.016
ZA	Bi46	138	0.017
ZA	Bi49	172	0.034
ZA	Bi50	116	0.018
ZA	Bi50	144	0.018
ZA	Bi51	184	0.017
ZA	Bi52	154	0.017
ZA	Bi52	158	0.017
ZA	Bi54	114	0.034
ZA	Bi56	176	0.018
ZA	Bi56	182	0.018
ZA	Bi39	210	0.069
ZA	Bi42	120	0.017
BR	Bi46	136	0.016
BR	Bi46	160	0.016
BR	Bi58	156	0.016