

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
DEPARTAMENTO DE BIOLOGIA GERAL
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA



DISSERTAÇÃO DE MESTRADO

**GENÉTICA DA CONSERVAÇÃO DO PATO-MERGULHÃO
(*Mergus octosetaceus*), UMA ESPÉCIE DE AVE
CRITICAMENTE AMEAÇADA DO CERRADO BRASILEIRO**

ORIENTADO: Davidson Pinheiro Campos

ORIENTADOR: Dr. Fabrício Rodrigues dos Santos

CO-ORIENTADORA: Dr. Eloisa Helena Reis Sari

BELO HORIZONTE

AGOSTO – 2017

Davidson Pinheiro Campos

Dissertação de Mestrado

**GENÉTICA DA CONSERVAÇÃO DO PATO-MERGULHÃO
(*Mergus octosetaceus*), UMA ESPÉCIE DE AVE
CRITICAMENTE AMEAÇADA DO CERRADO BRASILEIRO**

Dissertação apresentada ao curso de Pós-graduação em Genética do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais como requisito parcial para obtenção de Título de Mestre em Genética.

Orientador: Fabrício Rodrigues dos Santos

Belo Horizonte

Agosto – 2017

À Patrícia e Cleide

Agradecimentos

Agradeço a Fundação Grupo Boticário pelo financiamento deste projeto, a Capes pela concessão da bolsa de mestrado que me possibilitou fazer esta pós-graduação.

Ao meu orientador Fabrício Rodrigues dos Santos pela oportunidade de ingressar em seu laboratório, ainda na graduação, por me aceitar na pós-graduação e por me confiar este projeto.

Aos parceiros e amigos que fiz nas reuniões do PAN do Pato-Mergulhão, o trabalho de vocês para a conservação desta espécie é incrível.

As ONGs e instituições: Instituto Terra Brasilis, Zooparque de Itatiba, Funatura, Naturatins e Cervivo pelo envio das amostras.

Aos membros da banca pela disponibilidade, boa vontade, pela avaliação e sugestões.

Ao Pierre, que na sua breve permanência aqui no Brasil, me ensinou muito a respeito da bioinformática, me mostrando um mundo de possibilidades que até o momento eram desconhecidas por mim, sem sua ajuda dificilmente eu conseguiria defender este mestrado.

Ao Zé por se dispor a me ajudar ao longo de todo meu período no laboratório, sempre buscando me ensinar algo que não sabia, pelas grandes contribuições nesta dissertação, discutindo sobre o tema inclusive nos fins de semana e pela amizade ao longo destes anos.

Ao Marco, que ao longo destes anos sempre me ajudou, inclusive nesta dissertação, por me aconselhar em assuntos pessoais e profissionais, pelos botecos pós laboratório que tronavam os dias mais leves. Você foi uma grande amizade que fiz nestes anos.

A Camilla pela companhia de laboratório, pela ajuda neste projeto, pelas risadas e pela amizade ao longo destes anos.

Ao Raphael que sempre que lia um artigo ou descobrisse um programa que me poderia ser útil sempre buscava me mostrar.

A minha família, em especial a minha mãe, que sempre me apoiou e entendeu minha ausência na fábrica, fazendo tudo o que foi possível para encaminhar as coisas da melhor maneira.

A minha noiva e futura esposa, Patrícia: obrigado por tudo, pelo apoio incondicional, por me incentivar, puxar a orelha em alguns momentos e por ser um exemplo a ser seguido. Te amo!

A todos os amigos do LBEM e fora dele, que tornaram os dias mais leves.

A todos que contribuíram para este projeto, de maneira direta ou indireta, e que não foram citados aqui, muito obrigado.

A Deus que me deu forças para superar cada obstáculo que surgiu neste trajeto.

Índice

Lista de Figuras	VI
Lista de Tabelas	IX
Lista de Abreviaturas	X
Resumo	1
Abstract	2
Introdução.....	3
Objetivos.....	9
Objetivos Específicos	9
Materiais e Métodos.....	10
Amostragem e Preparação do DNA	10
Protocolo de genotype by sequencing (GBS)	10
Pré filtragem dos dados gerados.....	12
Bioinformática de filtragem e genotipagem dos SNPs com as pipelines dos dados GBS.	12
Análises de Genética Populacional	14
Resultados.....	15
Discussão	27
Conclusão.....	30
Referências	31

Lista de Figuras

- Figura 1: *Mergus octosetaceus*, Vieillot, 1817. A esquerda detalhe para o bico serrilhado com a extremidade recurvada. A esquerda um casal de *M. octosetaceus* em um trecho de rio no PARNA da Serra da Canastra.....6
- Figura 2: Mínimo polígono convexo compreendendo todos os pontos de distribuição histórica de *Mergus octosetaceus* Vieillot, 1817.....7
- Figura 3: Pontos de Ocorrência de *Mergus octosetaceus* com o número de amostras utilizadas neste trabalho.....8
- Figura 4: Proporção de *reads* com P maior do que 95% (azul) e a média de P em todos os *reads* (vermelho). No eixo X estão representados todos os 39 indivíduos amostrados de *Mergus octosetaceus* Vieillot, 1817 e no eixo Y a probabilidade, em porcentagem, dos pares de bases (pb) com valores de $P \geq 0.95$15
- Figura 5: Número de *reads* por indivíduo da espécie *Mergus octosetaceus* Vieillot, 1817, com a proporção de cada um dos *reads* por classe de Phred Score (P). Azul escuro – P menor do que 85%; azul claro – P entre 85% e 90%; verde – P entre 90 % e 95%; laranja – P maior do que 95%; e vermelho – *reads* rejeitados por não conterem o sítio de restrição da enzima (TGCA).....16
- Figura 6: Análise de PCA, utilizando 1328 SNPs, com todos os indivíduos da espécie *Mergus octosetaceus* Vieillot, 1817. Azul – indivíduos do Parque Estadual do Jalapão; marrom – indivíduos do Parque Nacional da Chapada dos Veadeiros; verde escuro – indivíduos da região do Alto Paranaíba; verde claro – indivíduos do Parque Nacional da Serra da Canastra.....17
- Figura 7: Análise de PCA, utilizando 1328 SNPs, com os indivíduos não aparentados para a espécies *Mergus octosetaceus* Vieillot, 1817. Azul – indivíduos do Parque Estadual do Jalapão; marrom – indivíduos do Parque Nacional da Chapada dos Veadeiros; verde escuro – indivíduos da região do Alto Paranaíba; verde claro – indivíduos do Parque Nacional da Serra da Canastra.....18
- Figura 8: Gráfico do Geneland mostrando a distribuição espacial dos agrupamentos formados pelos indivíduos da espécie *Mergus octosetaceus* Vieillot, 1817. Verde – população do Parque Estadual do Jalapão; branco – população do Parque Nacional da

Chapada dos Veadeiros; amarelo – população do Parque Nacional da Serra da Canastra e da região do Alto Paranaíba.....19

Figura 9: Gráfico do Geneland mostrando a distribuição espacial dos agrupamentos formados pelos indivíduos da espécie *Mergus octosetaceus* Vieillot, 1817. Laranja – população do Parque Estadual do Jalapão; verde escuro – população do Parque Nacional da Chapada dos Veadeiros; verde clara – população da região do Alto Paranaíba; branco – população do Parque Nacional da Serra da Canastra e um indivíduo da região do Alto Paranaíba.....20

Figura 10: Resultado do Structure no modelo não correlacionado, incluindo todos os indivíduos amostrados para a espécie *Mergus octosetaceus* Vieillot, 1817. Azul – alelos comuns no agrupamento do Parque Nacional da Serra da Canastra; laranja – alelos comuns no agrupamento formado pelo Parque Nacional da Chapada dos Veadeiros + Parque Estadual do Jalapão; preto – alelos mais comuns no agrupamento da região do Alto Paranaíba.....21

Figura 11: Resultado do Structure no modelo correlacionado, incluindo todos os indivíduos amostrados para a espécie *Mergus octosetaceus* Vieillot, 1817. Azul – alelos comuns no agrupamento do Parque Nacional da Serra da Canastra; verde – alelos comuns no agrupamento formado pelo Parque Nacional da Chapada dos Veadeiros; laranja – alelos comuns no agrupamento Parque Estadual do Jalapão; preto – alelos mais comuns no agrupamento da região do Alto Paranaíba.....21

Figura 12: Resultado do Structure, no modelo não correlacionado, excluindo os indivíduos aparentados em primeiro grau para a espécie *Mergus octosetaceus* Vieillot, 1817. Azul – alelos comuns no agrupamento do Parque Nacional da Serra da Canastra; laranja – alelos comuns no agrupamento formado pelo Parque Nacional da Chapada dos Veadeiros + Parque Estadual do Jalapão; preto – alelos mais comuns no agrupamento da região do Alto Paranaíba.....22

Figura 13: Resultado do Structure, no modelo correlacionado, excluindo os indivíduos aparentados em primeiro grau para a espécie *Mergus octosetaceus* Vieillot, 1817. Azul – alelos comuns no agrupamento do Parque Nacional da Serra da Canastra; laranja– alelos comuns no agrupamento formado pelo Parque Nacional da Chapada dos Veadeiros + Parque Estadual do Jalapão; preto – alelos mais comuns no agrupamento da região do Alto Paranaíba.....22

Figura 14: Frequências alélicas das populações de *Mergus octosetaceus*. A – Alto Paranaíba; B - PARNA Serra da Canastra; C – PARES do Jalapão; D – PARNA da Chapada dos Veadeiros.....26

Lista de Tabelas

Tabela 1: Identificação dos barcodes associados a cada amostra utilizadas neste trabalho.....	11
Tabela 2: Mudanças de bases mais frequentes e índices de diversidade nucleotídicas de cada população de <i>M. octosetaceus</i>	23
Tabela 3: Análise de FST par a par das populações de <i>M. octosetaceus</i> , definidas com base na localidade geográfica.....	23
Tabela 4: AMOVA - Análise de Variância Molecular de 3 níveis paras diversos agrupamentos populacionais de <i>M. octosetaceus</i>	24

Lista de Abreviaturas

µl – Microlitro

AMOVA – Análise de Variância Molecular

CR – Criticamente ameaçado

CR – Região Controle

CytB – Citocromo B

ddRADseq - Double digest restriction-site associated DNA sequencing

ESALQ - Escola Superior de Agricultura Luiz de Queiroz

GBS - Genotype by Sequencing

gDNA – DNA genômico

IUCN - International Union for Conservation of Nature and Natural Resources

Km – Kilômetros

LBEM - Laboratório de Biodiversidade e Evolução Molecular

MCMC - Cadeia de Markov Monte Carlo

min – Minutos

MMA – Ministério do Meio Ambiente

mtDNA – DNA Mitocondrial

MUSK 13 - Íntron 13 do gene Músculo Esquelético Receptor da Tirosina Quinase

Ne – Tamanho efetivo populacional

ng – Nanogramas

NGS - Next Generation Sequencing

PARES – Parque Estadual

PARNA – Parque Nacional

pb – Pares de base

PCA - Análise de Componente Principal

PCR – Reação em cadeia da polimerase

qPCR – PCR quantitativa

QTL - Loci de características quantitativas

RADseq - Restriction-site associated DNA sequencing

RRG – Reduced Representation Genomics

seg – Segundos

SNP – Polimorfismo de Nucleotídeo Único

Tb – Terabases

UCs – Unidades de Conservação

Resumo

Com o aumento do número de espécies em risco de extinção, estudos voltados para a conservação são cada vez mais necessários. Estudos de diversidade genética remanescente de populações ameaçadas podem nos fornecer subsídios para buscar melhores estratégias de manejo para estas populações. O pato-mergulhão (*Mergus octosetaceus*) é uma das espécies de aves mais ameaçadas da América do Sul, apresentando tamanho populacional bastante restrito. Por ser extremamente sensível a perturbações ambientais, foi classificado como criticamente ameaçado na lista vermelha da IUCN desde 1994. Compreender a diversidade genética da população e promover estratégias para aumentar esta variabilidade é fundamental para a conservação da espécie em seu ambiente natural. Para compreender a diversidade genética remanescente do pato-mergulhão estudamos a genômica populacional da espécie. Utilizando “*Genotyping by Sequencing – GBS*” identificamos 48.517 SNPs pertencentes a 39 indivíduos de todas as áreas de ocorrência, 5 indivíduos foram descartados por baixa cobertura, 1 não teve SNPs identificados e 2 foram excluídos por estarem amostrados em duplicidade. Após as filtrações obtivemos 1.328 SNPs pertencentes a 31 indivíduos. As populações remanescentes do pato-mergulhão apresentam baixa variabilidade genética. O PARNA Serra da Canastra e a região do Alto Paranaíba se comportam como só uma população, apresentando alto compartilhamento de alelos. Eventos de *bottleneck* podem ter ocorrido nas populações da espécie e podem ainda estar causando danos como altas taxas de endogamia. Os resultados mostram que ações para a conservação da espécie são urgentes e devem ser iniciadas o quanto antes para que a diversidade genética seja mantida e/ou aumentada, propiciando que ela sobreviva em seu ambiente natural.

Palavras chave: genômica, genética da conservação, sequenciamento de nova geração, SNP

Abstract

With the increasing number of endangered species, conservation studies are increasingly needed. Studies of the reminiscent genetic diversity of threatened populations may provide us with subsidies to seek better management strategies for these populations. The brazilian merganser (*Mergus octosetaceus*) is one of the most endangered bird species in South America, presenting a very restricted population size. For being extremely sensitive to environmental disturbances, it was classified as critically endangered on the IUCN Red List since 1994. Understanding the population's genetic diversity and promoting strategies to increase this variability is critical to the conservation of the species in its natural environment. To understand the remaining genetic diversity of the brazilian merganser we studied the population genomics of the specie. Using "Genotyping by Sequencing - GBS" we identified 48,517 SNPs belonging to 39 individuals from all areas of occurrence, 5 individuals were discarded for low coverage, 1 had no identified SNPs and 2 were excluded because they were duplicate samples. After the filtering procedure, we obtained 1.328 SNPs belonging to 31 individuals. The remaining populations of the brazilian merganser presents low genetic variability, the Serra da Canastra PARNA and the Alto Paranaíba region behave as only one population, presenting high allele sharing. Bottleneck events may have occurred in species populations and may still cause damage such as high rates of inbreeding. The results indicate that actions for the conservation of the species are urgent and must be started as soon as possible so that genetic diversity is maintained and/or increased, allowing it to survive in its natural environment.

Keywords: genomics, conservation genetics, next generation sequencing, SNP

Introdução

Devido a constantes ameaças como caça e pesca excessivas, introdução de espécies exóticas, fragmentação e destruição de habitats, o número de espécies em risco de extinção cresceu consideravelmente (BirdLife International 2016). Para as aves, esse número aumentou cerca de 30%, passando de 1.107 para 1.460 espécies, somente nos últimos vinte anos (BirdLife International 2016). Em decorrência disto, estudos populacionais em aves são cada vez mais necessários para compreender ou inferir os níveis de ameaça e buscar estratégias de conservação (Hughes et al. 2002, Johnson et al. 2005, Schiegg et al. 2006).

Com o advento da genética e o desenvolvimento de novas ferramentas e análises sobre dinâmica populacional com dados de marcadores moleculares houve um aumento no número de estudos voltados para a conservação das espécies (Frankham, 1995; Witzemberger & Hochkirch, 2011). Esse campo emergente ficou conhecido pelo termo “genética da conservação”, uma nova disciplina que busca compreender a dinâmica microevolutiva e os processos que levaram uma determinada espécie às condições atuais (Meffe, 1986; Hedrick, 2001, Frankham et al. 2010). A partir dessas inferências torna-se possível propor estratégias e métodos de conservação para a manutenção das espécies a curto, médio e longo prazos podendo assim traçar estratégias que visam amenizar os efeitos dos eventos estocásticos e antropogênicos nas populações (Frankham et al. 2010).

Até a década passada, sequências de DNA utilizadas em estudos de conservação eram obtidas majoritariamente pela metodologia de sequenciamento tradicional (Sanger), utilizando regiões restritas do genoma (Liu, et al. 2012, Li et al. 2015). Com o surgimento das metodologias de sequenciamento de nova geração (*Next Generation Sequencing* – NGS), houve um aumento no número de sequências de DNA e novos marcadores moleculares através da densa amostragem do genoma (Dierickx et al. 2015, Lozier, 2014). Essa metodologia tornou possível estudar questões biológicas com maior precisão, como por exemplo, o mapeamento de *loci* de características quantitativas (QTL), a identificação de regiões genômicas diferentes entre populações, sinais de seleção natural ao longo do genoma etc. (Allendorf et al. 2010, Davey et al. 2011).

Apesar da redução recente dos custos das análises genômicas, sequenciar o genoma inteiro de um indivíduo de uma espécie ainda é relativamente caro, complexo, demorado e acessível a poucos grupos de pesquisa, principalmente para espécies não modelo (Luca et al. 2011). Com a necessidade de técnicas que permitissem abordagens

populacionais (de múltiplos indivíduos) surgiram vários métodos de redução da complexidade ou representação genômica (RRG) de baixo custo (Elshire et al. 2011, Peterson et al. 2012). Basicamente, dois modos diferentes de se obter esta redução foram descritos. O primeiro envolve o enriquecimento de regiões alvo através de múltiplas reações em cadeia da polimerase (PCR) ou hibridação com oligonucleotídeos complementares (Albert et al. 2007). O segundo envolve enzimas de restrição que fragmentam o DNA genômico em regiões específicas (Baird et al. 2008, Elshire et al. 2011, Peterson et al. 2012). Esta última abordagem reduz a complexidade do genoma através da fragmentação deste com enzimas de restrição durante a construção de bibliotecas genômicas que são submetidas ao sequenciamento de nova geração, que foi utilizada em diversos estudos voltados para a conservação de espécies (Davey et al. 2011, Allendorf et al. 2010, Narum et al. 2013).

Por exemplo, Lozier (2014) utilizou uma metodologia denominada “*Restriction-site associated DNA sequencing – RADseq*” (Baird et al 2008), para avaliar a diversidade genética de populações de abelhas estáveis e em declínio. Nesta metodologia, o DNA genômico é digerido com uma enzima de restrição e os fragmentos são ligados a adaptadores que contém um *barcode* (oligonucleotídeos de 4 ou 5 pares de bases – pb) e sítios de ligação para o primer de sequenciamento. Em seguida, os fragmentos ligados aos adaptadores são cortados aleatoriamente, o que acarreta em fragmentos de diferentes tamanhos, estes fragmentos são ligados a um segundo adaptador que possui uma sequência específica que impede que fragmentos que não possuem o primeiro adaptador sejam amplificados. Após a purificação os fragmentos são enriquecidos através de reações em cadeia de polimerase – PCR (Baird et al. 2008).

Em outros estudos com espécies ameaçadas como o Albatroz-de-pés-negros (*Phoebastria nigripes*) (Dierickx et al. 2015) e a toupeira-da-água (*Galemys pyrenaicus*) (Querejeta et al. 2016), foi utilizada uma metodologia denominada “*double digest Restriction-site associated DNA sequencing – ddRADseq*” (Peterson et al. 2012) para estudos de conservação destas espécies. Esta metodologia é uma adaptação da metodologia descrita por Baird e colaboradores (2008) que, ao invés de uma, utiliza duas enzimas de restrição para a digestão do DNA e faz uma seleção de tamanho mais precisa, excluindo regiões muito próximas ou muito distantes do sítio de restrição da enzima (Peterson et al. 2012).

Outra metodologia de NGS que utiliza enzimas de restrição é denominada “*Genotyping by Sequencing – GBS*” (Elshire et al. 2011). Esta metodologia foi inicialmente descrita em um estudo com genoma de milho (*Zea mays* Linnaeus, 1758), que é de origem

poliploide e de estrutura complexa devido a rearranjos mediados por transposons. Quando comparada à metodologia descrita por Baird e colaboradores (2008), GBS mostrou ser mais simples, com geração dos fragmentos de maneira mais direta, usando digestão do DNA genômico em uma única etapa implicando menos manuseio da amostra, e com diminuição de erros de sequenciamento pela modulação dos *barcodes* proposta nesta metodologia (Elshire et al 2011). O GBS vem sendo utilizado em diversos grupos taxonômicos como mamíferos (White et al. 2013, Johnson, et al.2015, De Donato et al, 2013), plantas (Velmurugan et al. 2016) e aves (Pellegrino et al. 2016).

Com o grande volume de dados gerados pelos métodos de NGS (tecnologia Illumina SBS pode gerar terabases – Tb de dados), programas capazes de processar adequadamente esses dados são cada vez mais necessários (Narum et al. 2013). Diversos algoritmos ordenados em série ou *pipelines* são utilizados para realizar várias etapas destas análises, desde a preparação das sequências de boa qualidade até as análises populacionais e filogenéticas. Por exemplo, alguns algoritmos permitem a filtragem dos dados que pode ser feita com a exclusão de *reads* de baixa cobertura e classificação dos *reads* por indivíduo com base no *barcode*. Posteriormente, estes podem ser alinhados com um genoma de referência ou submetidos a um alinhamento *de novo*, que finalmente permitem identificar o genótipo ou SNPs de cada indivíduo (Glaubitz et al. 2014, Catchen et al. 2011, Puritz et al. 2014). Mesmo sendo desenvolvidos para trabalhar com grupos específicos de dados (genoma de uma espécie), essas *pipelines*, podem ser aplicadas aos mais diversos grupos taxonômicos. Dentre as *pipelines* mais utilizadas, destacam-se Stacks (Catchen et al. 2011), Tassel-GBS (Glaubitz et al. 2014) e Uneak (Lu et al. 2012, White et al. 2013, Matthews et al. 2013). Enquanto o Stacks (Catchen et al. 2011) permite trabalhar com espécies que possuem um genoma de referência (Velmurgan et al. 2016) ou não, utilizando alinhamento *de novo*, o Tassel-GBS (Glaubitz et al. 2014) permite trabalhar somente com espécies que possuem um genoma de referência. O Uneak foi uma *pipeline* desenvolvida para possibilitar que o programa Tassel 3.0 trabalhasse com espécies que não possuem genoma de referência (Lu et al. 2012, White et al. 2013, Matthews et al. 2013).

O pato-mergulhão

O pato-mergulhão, *Mergus octosetaceus* Vieillot, 1817 é o único anseriforme representante da tribo Mergini Livezey, 1986 encontrado no hemisfério sul. Ele tem entre 48 e 55 centímetros de comprimento, peso de aproximadamente 800 gramas, e apresenta um penacho na parte posterior da cabeça que, juntamente com o pescoço, são mais escuros que o restante do corpo, devido à presença de penas com reflexo verde metálico. O peito e

o ventre são cinza com algumas manchas brancas, o dorso, asa e cauda acinzentados, os pés são vermelhos, o bico é negro, fino, estreito, serrilhado e com extremidade recurvada (Sick, 1985; Bruno, 2004) (Figura 1). Sua alimentação baseia-se em peixes (entre 6 a 19 cm de comprimento) e macroinvertebrados aquáticos, com o hábito de forragear tanto em remansos quanto em corredeiras (Partridge,1956; Bartmann, 1988; Silveira e Bartmann, 2001). A espécie apresenta senso visual acurado e é capaz de mergulhar (média de 15 a 20 segundos por mergulho com profundidade de até 50 centímetros), o que permite a sua subsistência alimentando-se de peixes (Hughes et al. 2006).

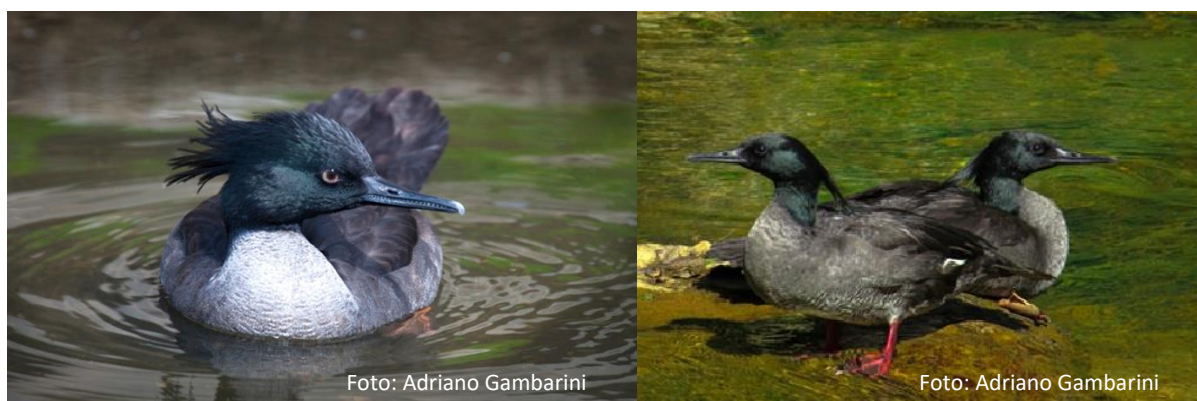


Figura 1: *Mergus octosetaceus*. Na foto à esquerda está em detalhe o bico serrilhado com a extremidade recurvada. Na foto à direita está um casal de *M. octosetaceus* em um trecho de rio no PARNA da Serra da Canastra

Os patos-mergulhões passam boa parte da vida na água, são territorialistas e um casal requer aproximadamente 5 a 12 km de rio de águas cristalinas, com áreas de remanso e corredeiras para forrageamento e reprodução (Partridge,1956; Bartmann, 1988; Silveira e Bartmann, 2001, Hughes et al. 2006). Apesar da sua territorialidade, são raras as brigas entre casais estabelecidos em territórios adjacentes (Silveira e Bartmann 2001), cujo pareamento ocorre por um longo período ou a vida toda, sendo registrados casos de troca de parceiros somente após a morte de um deles (Hughes et al. 2006, Ribeiro 2016).

O status de conservação de *M. octosetaceus* é preocupante, pois a espécie encontra-se como criticamente ameaçada na lista vermelha da IUCN (“*International Union for Conservation of Nature and Natural Resources*”) desde 1994 (BirdLife International 2016). Sua distribuição histórica compreendia três países - Brasil, Argentina e Paraguai - e incluía as bacias hidrográficas dos rios São Francisco, Paraná e Tocantins, com ocorrência em

várias áreas disjuntas, sempre apresentando baixa densidade populacional (Yamashita e Valle, 1990) (Figura 2).

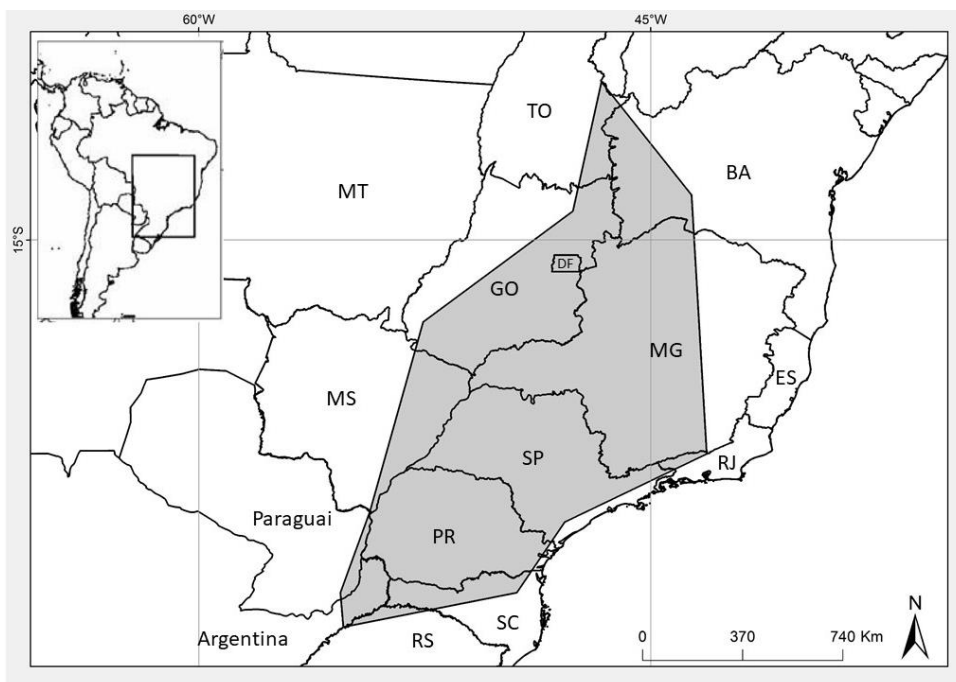


Figura 2: Mínimo polígono convexo compreendendo todos os pontos de distribuição histórica de *Mergus octosetaceus*.

Os últimos registros da espécie na Argentina e no Paraguai datam de 2002 e 1984, respectivamente (BirdLife International 2016; Benstead et al. 1994, Hearn 1994, Fragano et al. 2003). A falta de registros nos últimos anos indica que as populações do pato-mergulhão nesses países podem ser extremamente raras ou ter sido extinta. Por isso acredita-se que a distribuição da espécie esteja limitada ao Brasil. Embora *M. octosetaceus* seja encontrado principalmente em áreas protegidas (Unidades de Conservação – UCs), como o Parque Nacional das Emas (PARNA das Emas), Parque Nacional da Chapada dos Veadeiros (PARNA da Chapada dos Veadeiros), Parque Estadual do Jalapão (PARES do Jalapão), e o Parque Nacional da Serra da Canastra (PARNA da Serra da Canastra) (Braz et al. 2003; Barbosa e Almeida 2010; Bartmann 1988; Silveira e Bartmann 2001; Lamas 2006; Yamashita e Valle 1990; Collar et al. 1992; Bianchi et al. 2005), essas populações encontram-se disjuntas por um mosaico de regiões com forte pressão antrópica (Hughes et al. 2006). Além das UCs, registros recentes no estado de Minas Gerais, Brasil mostram que existem populações da espécie nos municípios de Serra do Salitre, Coromandel e Patrocínio (Região do Alto Paranaíba) em áreas não protegidas e cercadas por forte ação antrópica, como mineração e monoculturas (Hughes et al. 2006) (Figura 3).

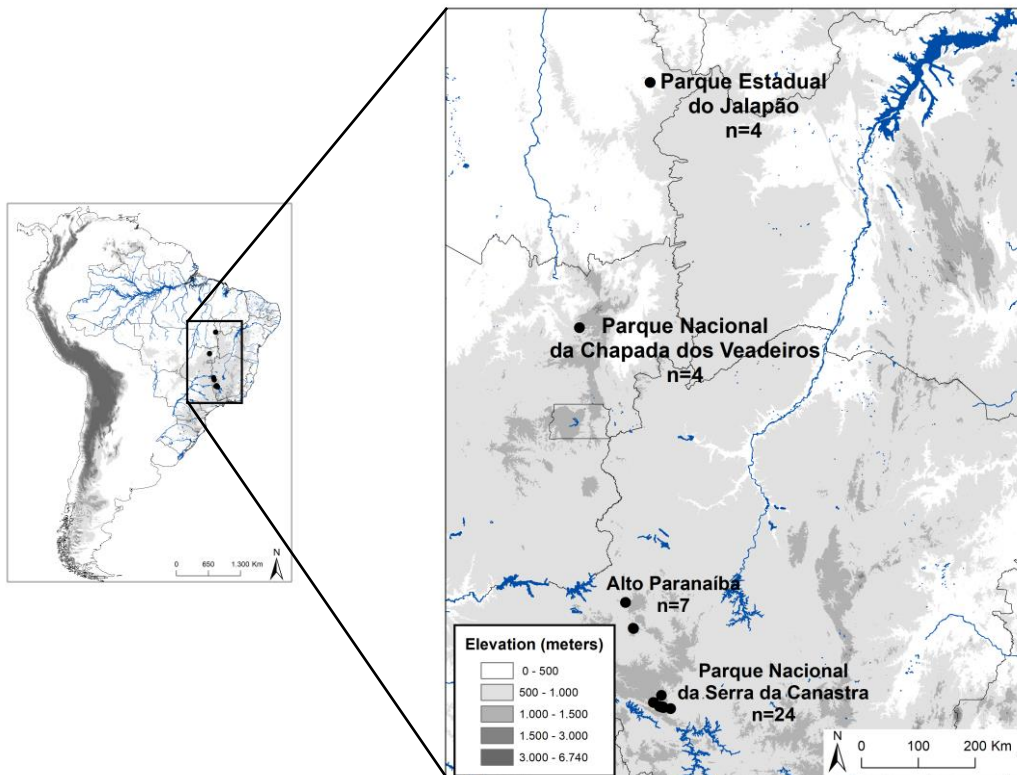


Figura 3: Pontos de ocorrência de *Mergus octosetaceus* com o número de amostras utilizadas neste trabalho.

Embora os estudos sobre o tamanho efetivo populacional (N_e) sejam um pouco especulativos, estima-se que existam apenas 250 indivíduos na natureza e cerca de 140 a 200 indivíduos aptos para a reprodução no PARNA Serra da Canastra (BirdLife International 2016) e 10 indivíduos em cativeiro. Esta estreita relação com a ambiente aquático faz com que os patos-mergulhões sejam bastante sensíveis a perturbações ambientais, tais como poluição das águas e o assoreamento dos rios, que diminuem a disponibilidade de alimento (Hughes et al. 2006). Eles também são sensíveis a perturbações causadas pela realização de esportes aquáticos, como canoagem e *rafting*, que fazem com que as fêmeas abandonem seus ninhos (Bruno, 2004). A baixa tolerância a ambientes antropizados contribui consideravelmente para o declínio populacional da espécie, que foi considerada como uma das aves mais raras e ameaçadas de extinção da América do Sul (Collar et al. 1992).

Apesar de *M. octosetaceus* estar classificado como criticamente ameaçado, tanto na lista mundial da IUCN quanto na lista de espécies ameaçadas de extinção da fauna brasileira (Portaria MMA nº 003/2003), apenas dois artigos sobre a diversidade genética do

pato-mergulhão foram publicados. Vilaça et al. (2012) utilizaram dois marcadores mitocondriais, Citocromo B (CytB) e Região Controle (CR), com um tamanho amostral de 37 indivíduos da Serra da Canastra e 2 indivíduos da Chapada dos Veadeiros, encontraram sinais de estruturação populacional, apesar da baixa variabilidade genética, nos animais localizados na Serra da Canastra, relatando dois agrupamentos distintos, separados por 20 passos mutacionais. Um destes agrupamentos continha indivíduos da Serra da Canastra e os dois indivíduos da Chapada dos Veadeiros e o outro somente indivíduos da Serra da Canastra. Já com a região controle, utilizando 17 indivíduos, encontrou dois haplótipos distintos, separados por 24 passos mutacionais, um referente a indivíduos da Serra da Canastra e outro referente a indivíduos da Chapada dos Veadeiros. Maia et al. (2015) utilizando 7 loci de microssatélites, marcadores mitocondriais (citocromo b e região controle) e um marcador nuclear (íntron 13 do gene Músculo Esquelético Receptor da Tirosina Quinase - MUSK 13) encontrou baixa variabilidade genética nas populações de *M. octosetaceus*, apresentando a população do PARNA Chapada dos Veadeiros como a mais diferenciada.

As diversas ameaças ao ambiente em que o pato-mergulhão vive e o baixo número populacional influenciam fortemente no grau de ameaça em que ele se encontra. Devido ao estado crítico de conservação para esta espécie endêmica do Brasil e aos poucos estudos genéticos, propusemos um projeto com uma abordagem genômica para o estudo das populações remanescentes de *Mergus octosetaceus*.

Objetivos

Estudar a diversidade genética remanescente das populações de pato-mergulhão (*M. octosetaceus*) com uma metodologia de sequenciamento de Nova Geração (NGS).

Objetivos Específicos

1. Obter dados de variantes genéticas de pato-mergulhão através da metodologia *genotype by sequencing* (GBS) e procedimentos de bioinformática.

2. Avaliar a diversidade entre e dentro de populações da espécie ao longo de sua área de ocorrência.
3. Testar a presença e identificar as possíveis barreiras ao fluxo gênico.
4. Identificar quantos possíveis agrupamentos populacionais existem na área de ocorrência da espécie.
5. Estimar a demografia histórica das populações (ex.: bottleneck, expansão).

Materiais e Métodos

Amostragem e Preparação do DNA

Das 211 amostras utilizadas neste trabalho, 201 foram coletadas nas áreas naturais de ocorrência da espécie pelos colaboradores do Instituto Terra Brasilis, Funatura e Cer Vivo, e 10 delas coletadas em cativeiro (Zooparque de Itatiba, Itatiba, SP) pela nossa equipe. O DNA genômico dessas amostras foi extraído a partir de diversos tipos de tecidos, como ovos (embriões, casca), penas encontradas em ninhos abandonados e sangue de animais capturados de acordo com o protocolo fenol-clorofórmio (Sambrook et al., 2001). As amostras de DNA extraídas foram tombadas no banco de DNA do Laboratório de Biodiversidade e Evolução Molecular – LBEM.

Protocolo de genotype by sequencing (GBS)

Os dados genômicos GBS foram produzidos por um serviço da Ecomol, empresa incubada na ESALQ de Piracicaba, com um protocolo adaptado de Elshire et al. (2011). Para a preparação da biblioteca genômica, todas as amostras tiveram sua integridade avaliada através de eletroforese em gel de agarose (2%), onde foi verificado se havia alguma degradação no DNAg. Nos aparelhos Nanodrop 2000 e Qubit® 2.0 as amostras foram avaliadas quanto a sua pureza e quantidade, seguindo as recomendações dos fabricantes. Do total de 211 amostras pertencentes ao banco de DNA do LBEM, somente 39 se mostraram aptas para a aplicação do protocolo. Após a escolha das amostras que estavam mais apropriadas, todas foram diluídas para a concentração de 10 ng/μl. Para a

digestão foram adicionadas separadamente 10 µl de DNA de cada amostra com 16 µl de água, 1 µl da enzima PstI (20 U/µl) (New England Biolabs) e 3 µl de Buffer. Esta mistura foi incubada a 37°C por 60 minutos e a confirmação da digestão foi realizada em gel de agarose (2%). A enzima *PstI* foi escolhida pois apresentou melhores resultados nos testes *in vitro* e também foi utilizada em outros estudos com aves (Toews et al. 2016) e outros animais (De Donato et al. 2013, White, et al. 2013). Após a digestão foram utilizados *barcodes* diferentes (Tabela 1), que variavam de 7 a 9 pares de base, para identificar as amostras. O DNA foi então liofilizado na centrífuga a vácuo (45°C por 120 minutos), e foi realizada a ligação do DNA aos adaptadores de acordo com o protocolo de Elshire e colaboradores (2011), adicionando 1µl da enzima T4 Ligase (New England Biolabs), 5µl do Buffer 10x e 19 µl de água, incubando a 22°C por 120 minutos e inativando a enzima por incubação a 65°C por 30 minutos. Após a ligação as amostras foram purificadas com (QIAquick PCR Purification Kit; Qiagen, Valencia, CA), seguindo as instruções do fabricante. Depois foi feita a amplificação dos produtos purificados através da inserção de *primers* com sequencias complementares aos fragmentos de restrição e adaptadores, que ligarão os produtos de PCR aos oligonucleotídeos que cobrem a *flow cell*. As temperaturas e os tempos das PCRs são dados a seguir: 1x(72°C – 5min; 98°C – 30s); 18x(98°C – 10seg; 65°C – 30seg; 72°C – 30seg); 1x(72°C – 5min; 4°C – infinito). O produto da amplificação foi conferido no gel de agarose (2%) e purificado com AMPure XP beads (Beckman-Coulter) para retirada de pequenos fragmentos de DNA. A biblioteca gerada foi quantificada por qPCR através do Bioanalyzer (Agilent Technologies). As bibliotecas selecionadas foram então sequenciadas no Illumina HiSeq 2000 (*single-end* com *reads* de 100 pb).

Tabela 1: Identificação dos *barcodes* associados a cada amostra utilizadas neste trabalho.

Amostra	Barcode	Sequência Barcode	Amostra	Barcode	Sequência Barcode
MO18	PstI_B65	GAATACCA	MO74	PstI_B85	AACAGAATA
MO69	PstI_B66	ACCGTCCA	MO135	PstI_B86	CCACCTCGT
MO70	PstI_B67	CTTACGCA	MO180	PstI_B87	TTGTTCGAT
MO210	PstI_B68	TAGCGACA	MO121	PstI_B88	GGTTGGTCA
MO206	PstI_B69	GGAACAAT	MO98	PstI_B89	AACGAACAT
MO203	PstI_B70	ACCTACCG	MO151	PstI_B90	CCAACCTGA

MO199	PstI_B71	CTTCTGTA	MO147	PstI_B91	TTGCTTACA
MO207	PstI_B72	CGAGGTGA	MO38	PstI_B92	AGTGCGCTT
MO202	PstI_B73	GAGAACAT	MO66	PstI_B93	GAACAGAAT
MO204	PstI_B74	ACCTCATG	MO100	PstI_B94	CCTACACCG
MO198	PstI_B75	TTGCTTCA	MO155	PstI_B95	TTCTTGTTA
MO139	PstI_B76	CGTGAGGT	MO99	PstI_B96	CGAGGTGGA
MO200	PstI_B77	GAACGACCT	MO23	PstI_B49	GAATACA
MO35	PstI_B78	ACCATGACA	MO148	PstI_B50	CCTGTCA
MO201	PstI_B79	TTGTCCTCA	MO42	PstI_B51	TTCCGCA
MO205	PstI_B80	ACTGGTGGT	MO133	PstI_B52	TAGACCA
MO209	PstI_B81	GGAACAAGT	MO134	PstI_B53	GGCCTCA
MO211	PstI_B82	CACCATGAA	MO43	PstI_B54	ACATGCA
MO71	PstI_B83	TTGTTCTCA	MO159	PstI_B55	CTTGCAA
MO208	PstI_B84	GGTGAGTCT			

Pré filtragem dos dados gerados

As seqüências brutas de DNA (Fastq file) geradas no sequenciador HiSeq 2500 tiveram sua qualidade avaliada através da probabilidade P. Para cada amostra foi calculado a probabilidade P (*Phred score*) através da fórmula abaixo:

$$P = \prod_i^n (1 - 10^{-\frac{Q_i}{10}})$$

Onde, n é o número de pb em um determinado *read* e Qi, o escore de *phred* de pares de base i. Esta probabilidade é uma descrição global do *read* onde, se um único pb tiver um valor muito baixo de Q, o valor de P será fortemente afetado.

Bioinformática de filtragem e genotipagem dos SNPs com as pipelines dos dados GBS.

As sequências de DNA foram analisadas em três *pipelines* distintas.

i) UNEAK – implementado no programa Tassel 3.0 (Lu et al. 2012), utilizado para um alinhamento *denovo*. A pipeline selecionou *reads* de boa qualidade, que não possuíam “N” e os cortou em 64 pb após o barcode. Posteriormente os *reads* foram agrupados em *tags* por amostra e mesclados *tags* únicos por táxon, formando uma *Master tag* (para ser considerado uma *master tag*, o *tag* deveria aparecer pelo menos 10 vezes). Após esta etapa foi feito um alinhamento *pairwise*, onde somente *tags* com no máximo 1pb de diferença foram considerados como candidatos a SNPs, a taxa de tolerância de erro considerada foi o *default*. Por fim, um arquivo HapMap foi gerado com todos os SNPs encontrados para todos os indivíduos.

ii) Stacks (Catchen et al 2011) – utilizado para um alinhamento *de novo*. Essa pipeline tem cinco etapas, na primeira etapa, “*process_radtags*” os *reads* são demultiplexados e é feita uma filtragem pela qualidade do *read*. Posteriormente os *reads* são alinhados usando o programa “*ustacks*”. Um catálogo de loci de consenso foi construído usando o programa “*cstacks*”. Os *reads* foram comparados com o catálogo para determinar quais possuíam quais alelos usando o programa “*sstacks*”. Por último, o programa “genótipos” foi utilizado para exportar os SNPs identificados.

iii) Stacks (Catchen et al 2011) – utilizado para alinhamento com genoma de referência de *Anas platyrhynchos* Linnaeus, 1758), disponível em www.ensembl.org. Esta pipeline também tem cinco etapas, “*process_radtags*”, “*pstacks*” ao invés de “*ustacks*”, seguido de “*cstacks*” e “*sstacks*” com a opção – g para fazer com que seu algoritmo de correspondência coincida com a localização genômica e não com a similaridade dos *reads*, e por último, o programa “genótipos” foi utilizado para exportar os SNPs identificados.

Para uma melhor filtragem dos SNPs encontrados pelas três *pipelines*, utilizamos uma linguagem de programação chamada AWK (Brian, 1992) para selecionar SNPs com no mínimo 5X de cobertura [`awk 'NR>1{mm=1000; for (i=3; i<=33; i++) {if ($i<mm) {mm=$i}}; if (mm>=5) {print$1\t$2}}`]. Posteriormente utilizamos o VCFtools (Danecek et al. 2011), onde selecionamos somente os SNPs que eram comuns a todos os indivíduos (--max-missing-count 0), em que cada alelo deveria aparecer em pelo menos 2 *reads* distintos (--mac 2), sem inserções ou deleções (--remove-indels) e que o *locus* seja bi-alelético (--min-alleles 2 – max-alleles 2).

Análises de Genética Populacional

Análises multivariadas, gerando duas dimensões representadas graficamente, foram feitas através de Análise de Componente Principal (PCA) no programa Octave v.4.2.0 (Eaton et al. 2016), para representar graficamente a distância genética entre os indivíduos e as populações.

O número de possíveis agrupamentos populacionais e seus limites geográficos foram estimados com o programa Geneland v. 3.2.2 (disponível no software estatístico R). Esse programa estatístico usa cálculos de Cadeia de Markov Monte Carlo (MCMC) e a correlação genética espacial para inferir os agrupamentos mais prováveis (Guillot et al. 2008). Foram realizadas três corridas independentes, com 30 milhões de iterações, com um intervalo de 1 a 10 agrupamentos possíveis, em dois modelos diferentes: modelo de frequência alélica não correlacionada (1), que assume que as frequências alélicas entre as populações são diferentes; e correlacionada (2), que assume que as frequências alélicas são similares entre as populações. As corridas foram feitas com um modelo de alelos nulo falso e *thinning* de 300.000 e os resultados foram visualizados após um *burnin* de 100 vezes, onde foi avaliada a convergência de cada modelo.

O programa STRUCTURE v.2.3.4 (Pritchard et al. 2000) também foi utilizado para inferir o número provável de agrupamentos populacionais, porém sem correlação geográfica. Foi utilizado o modelo que prevê a possibilidade de fluxo gênico ou miscigenação entre as populações (*Admixture model*), com *burnin* de 250.000 e 1.000.000 randomizações coletadas MCMC. Foram avaliados valores de K de 1 a 10 com 10 corridas independentes para cada K. A escolha do K apropriado foi feita com o Structure Harvester (Earl et al. 2012) e o gráfico final foi gerado com o Clumpak (Kopelman et al. 2015).

Após a seleção do agrupamento mais provável, foram feitas algumas análises no programa Arlequin 3.5 (Excoffier & Lischer 2010). Análises de variância molecular (AMOVA), para estimar a diferenciação populacional; os índices de estatística de fixação (F_{ST} , F_{CT} , F_{SC}), quantificando a distribuição hierárquica da variabilidade genética em diferentes agrupamentos populacionais; o teste de Mantel (Mantel, 1967) para testar a correlação entre o isolamento genético e a distância geográfica.

O VCFtools foi utilizado para calcular a frequência dos alelos presentes em cada população e averiguar se eles estavam tendendo a fixação.

Resultados

O sequenciamento gerou 192.590.988 *reads*, 6.5% destes não apresentaram os *barcodes* ligados e 0.3% não possuíam o sítio da enzima de restrição (TGCA), ambos foram excluídos das análises. Após esta limpeza obtivemos 179.411.344 *reads*, que tiveram a qualidade inferida através da estatística P (*Phred Score*, Figura 4).

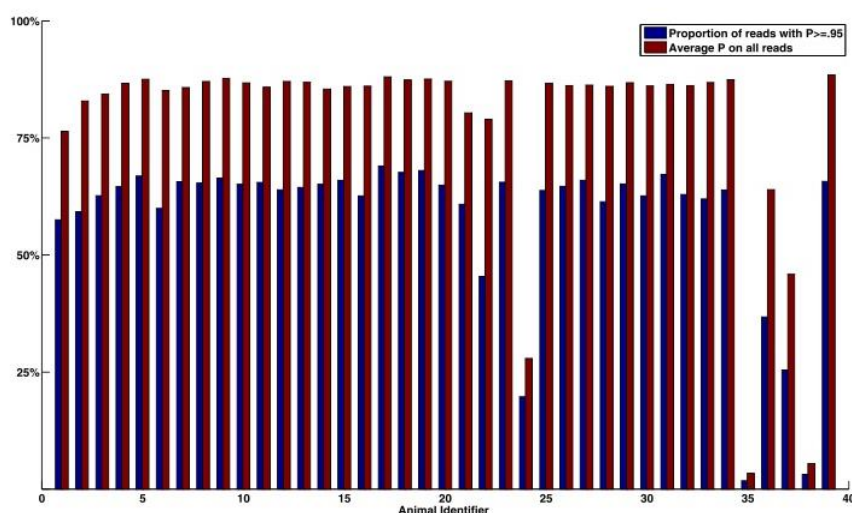


Figura 4: Proporção de *reads* com P maior do que 95% (azul) e a média de P em todos os *reads* (vermelho). No eixo X estão representados todos os 39 indivíduos amostrados de *Mergus octosetaceus* e no eixo Y a probabilidade, em porcentagem, dos pares de bases (pb) com valores de $P \geq 0.95$.

Dos 39 indivíduos analisados seis foram descartados. (MO42, MO43, MO121, MO133, MO134) por apresentar um baixo número de *reads* (Figura 5) e um (MO135) por não ter SNPs identificados ao término das análises, ambos da Serra da Canastra. Através de análises posteriores dos SNPs verificamos a possibilidade de duplicação de dois indivíduos. Após verificação com os parceiros responsáveis pela coleta das amostras, constatamos que, realmente estes indivíduos foram amostrados em coletas distintas (dados obtidos através do número da anilha), em um intervalo de três anos, cuja recaptura não havia sido relatada ao nosso laboratório. Desta forma, as análises posteriores foram feitas

com 31 indivíduos, sendo quatro do PARES do Jalapão, quatro do PARNA da Chapada dos Veadeiros, sete da região do Alto Paranaíba e dezesseis do PARNA da Serra da Canastra.

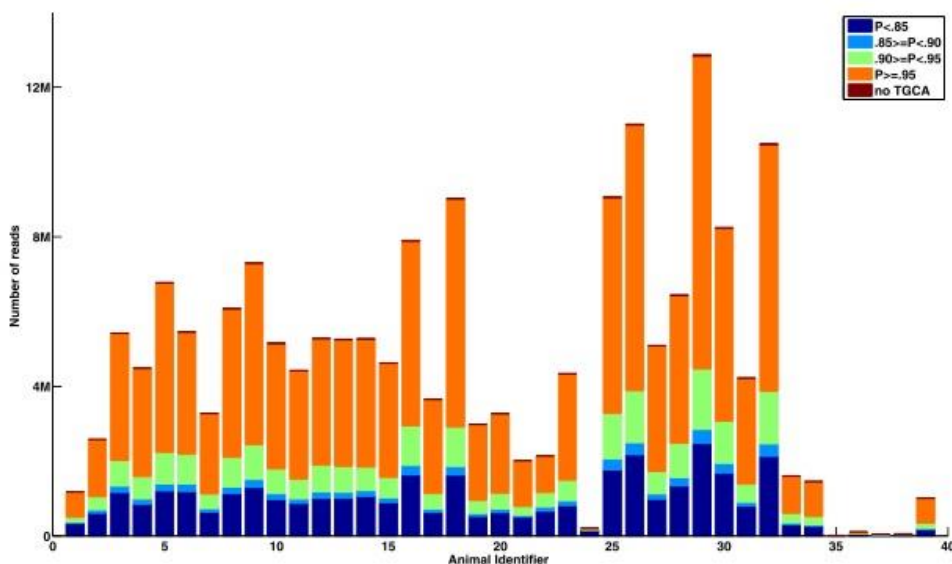


Figura 5: Número de *reads* por indivíduo da espécie *Mergus octosetaceus*, com a proporção de cada um dos *reads* por classe de *Phred Score* (P). Azul escuro – P menor do que 85%; azul claro – P entre 85% e 90%; verde – P entre 90 % e 95%; laranja – P maior do que 95%; e vermelho – *reads* rejeitados por não conterem o sítio de restrição da enzima (TGCA).

As *pipelines* UNEAK, *Stacks denovo* e *Stacks* com genoma de referência, encontraram respectivamente, 21.231, 48.517 e 53.376 SNPs cada um. Após a filtragem com VCFtools obtivemos respectivamente 2.523, 1.328 e 1.069 SNPs, comuns a todos os indivíduos, com no mínimo 5X de cobertura, no qual cada alelo deveria aparecer pelo menos duas vezes, sem *indels* e em *locus* bi-alelício.

Para as análises populacionais foram utilizados os resultados obtidos da *pipeline* do *Stacks de novo*, pelos motivos listados a seguir. A *pipeline* do UNEAK corta os *reads* em 64 pb após o sítio de restrição da enzima e só alinha *tags* com no máximo 1 pb de diferença. A parte descartada pela *pipeline* que vai de 64 pb a 91 pb corresponde a cerca de um terço dos *reads* e pode conter informações úteis. Além do mais, o fato da *pipeline* não alinhar *tags* com mais de 1 pb de diferença faz com que *tags* que possuam um *indel* não sejam reconhecidos, o que acarretaria em um erro de duplicidade. A *pipeline* *Stacks* com o genoma de referência de *A. platyrhynchos* foi descartada, pois a espécie *M. octosetaceus*

apresentou uma alta divergência genética quando comparada com *A. platyrhynchos*. O uso de um genoma de referência com alta divergência genética pode enviesar os resultados obtidos (Paris et al. 2017).

O PCA identificou quatro agrupamentos: O primeiro com todos os indivíduos do PARES do Jalapão, o segundo com todos indivíduos do PARNA da Chapada dos Veadeiros, o terceiro com a maioria dos indivíduos da região do Alto Paranaíba, e o último com todos os indivíduos do PARNA da Serra da Canastra. Alguns indivíduos do Alto Paranaíba agruparam com os indivíduos do PARNA da Serra da Canastra (Figura 6).

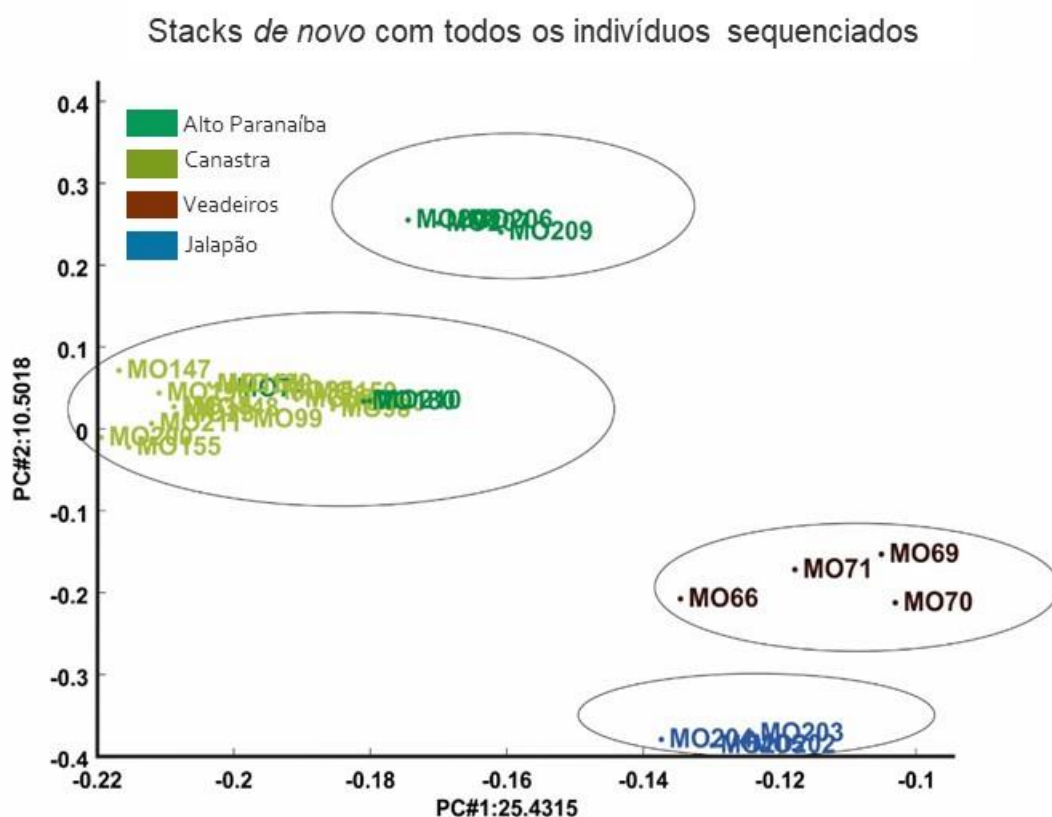


Figura 6: Análise de PCA, utilizando 1.328 SNPs, com todos os indivíduos sequenciados da espécie *Mergus octosetaceus*. Azul – indivíduos do Parque Estadual do Jalapão; marrom – indivíduos do Parque Nacional da Chapada dos Veadeiros; verde escuro – indivíduos da região do Alto Paranaíba; verde claro – indivíduos do Parque Nacional da Serra da Canastra.

Como o PCA utiliza uma matriz de diferenças para representação gráfica bidimensional, indivíduos mais aparentados tendem a ficar mais próximos, o que pode acabar inflando o gráfico, fazendo com que se forme um grupo mais distante dos demais. Por isto, em outras análises resolvemos excluir indivíduos que são estimados como irmãos

completos, pai e filho ou mãe e filho. Após esta etapa encontramos os mesmos quatro agrupamentos da análise que utilizamos todos os indivíduos (Figura 7).

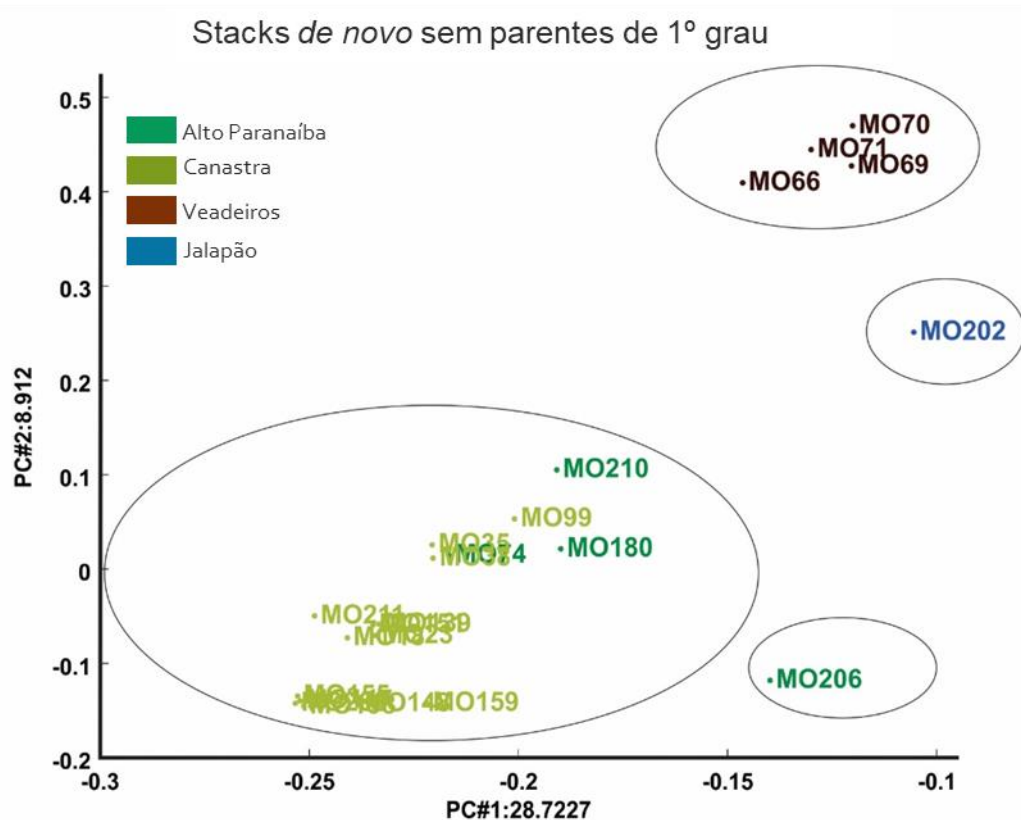


Figura 7: Análise de PCA, utilizando 1.328 SNPs, com os indivíduos não aparentados para a espécie *Mergus octosetaceus*. Azul – indivíduos do Parque Estadual do Jalapão; marrom – indivíduos do Parque Nacional da Chapada dos Veadeiros; verde escuro – indivíduos da região do Alto Paranaíba; verde claro – indivíduos do Parque Nacional da Serra da Canastra.

Três agrupamentos foram identificados, no modelo de frequência alélica não correlacionado, pelo programa Geneland. O primeiro composto por indivíduos do PARES do Jalapão, o segundo por indivíduos do PARNA da Chapada dos Veadeiros e o terceiro composto por indivíduos da região do Alto Paranaíba e do PARNA da Serra da Canastra (Figura 8).

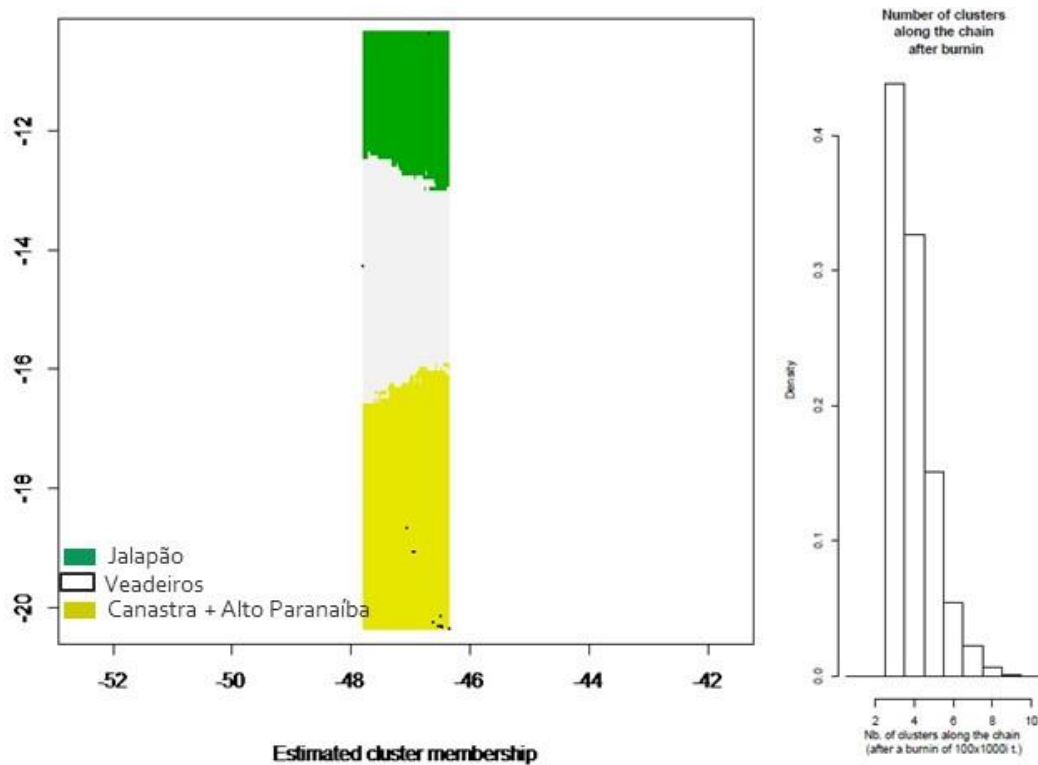


Figura 8: Gráfico do Geneland mostrando a distribuição espacial dos agrupamentos formados pelos indivíduos da espécie *Mergus octosetaceus*. Verde – população do Parque Estadual do Jalapão; branco – população do PARNA da Chapada dos Veadeiros; amarelo – população do PARNA da Serra da Canastra e da região do Alto Paranaíba.

Utilizando o modelo de frequência alélica correlacionado, o Geneland identificou quatro agrupamentos populacionais. O primeiro composto por indivíduos do PARES do Jalapão, o segundo por indivíduos do PARNA da Chapada dos Veadeiros, o terceiro composto por indivíduos da região do Alto Paranaíba e o quarto composto por indivíduos do PARNA da Serra da Canastra, com um indivíduo da região do Alto Paranaíba (Figura 9).

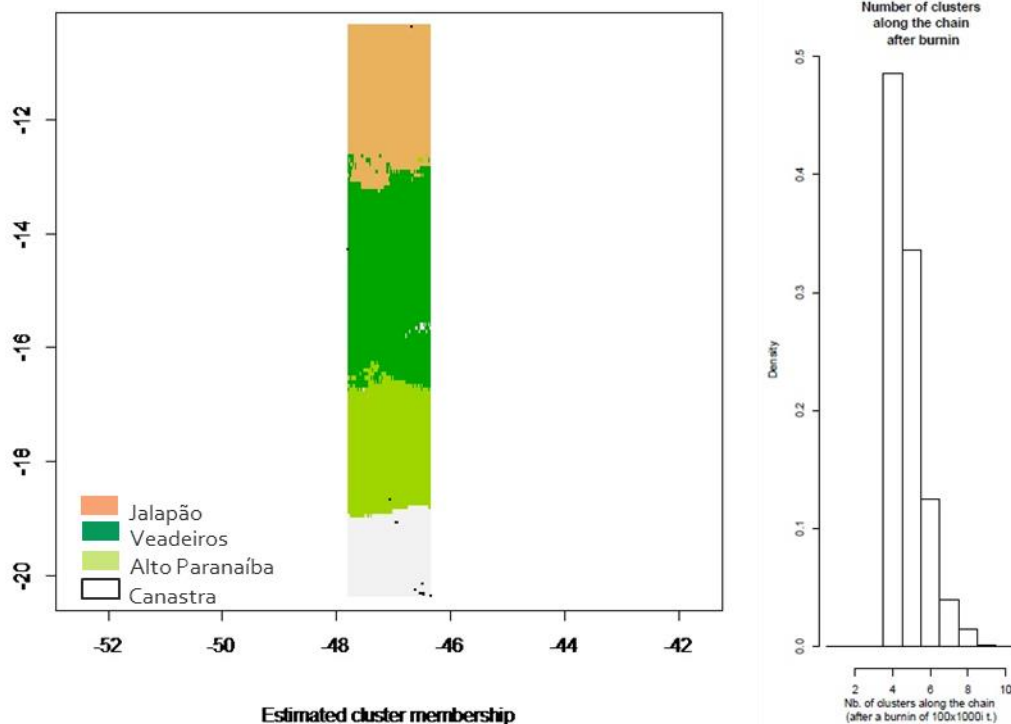


Figura 9: Gráfico do Geneland mostrando a distribuição espacial dos agrupamentos formados pelos indivíduos da espécie *Mergus octosetaceus*. Laranja – população do PARES do Jalapão; verde escuro – população do PARNA da Chapada dos Veadeiros; verde clara – população da região do Alto Paranaíba; branco – população do PARNA da Serra da Canastra e um indivíduo da região do Alto Paranaíba.

Para inferência do número de *clusters* sem correlação geográfica, utilizamos o programa Structure com modelos de frequência alélica correlacionada e não correlacionada. Usando o modelo não correlacionado, três agrupamentos foram encontrados (Figura 10). O agrupamento formado pelos espécimes do PARNA da Serra da Canastra, serve de fonte para as populações dos demais agrupamentos, PARES do Jalapão + PARNA da Chapada dos Veadeiros e Região do Alto Paranaíba. Já com o modelo correlacionado cinco agrupamentos foram encontrados (Figura 11), onde podemos verificar que a região do Alto Paranaíba possui alelos de todas as outras regiões.

Stacks *denovo* todos os indivíduos - frequência. alélica não correlacionada

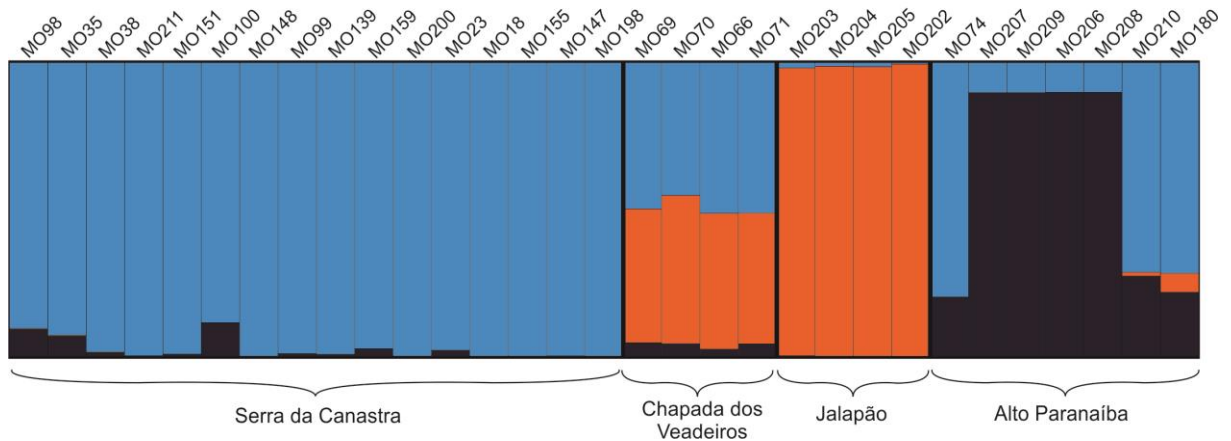


Figura 10: Resultado do Structure no modelo não correlacionado, incluindo todos os indivíduos amostrados para a espécie *Mergus octosetaceus*. Azul – alelos comuns no agrupamento do PARNA da Serra da Canastra; laranja – alelos comuns no agrupamento formado pelo PARNA da Chapada dos Veadeiros + PARES do Jalapão; preto – alelos mais comuns no agrupamento da região do Alto Paranaíba.

Stacks *denovo* todos os indivíduos - frequência. alélica correlacionada

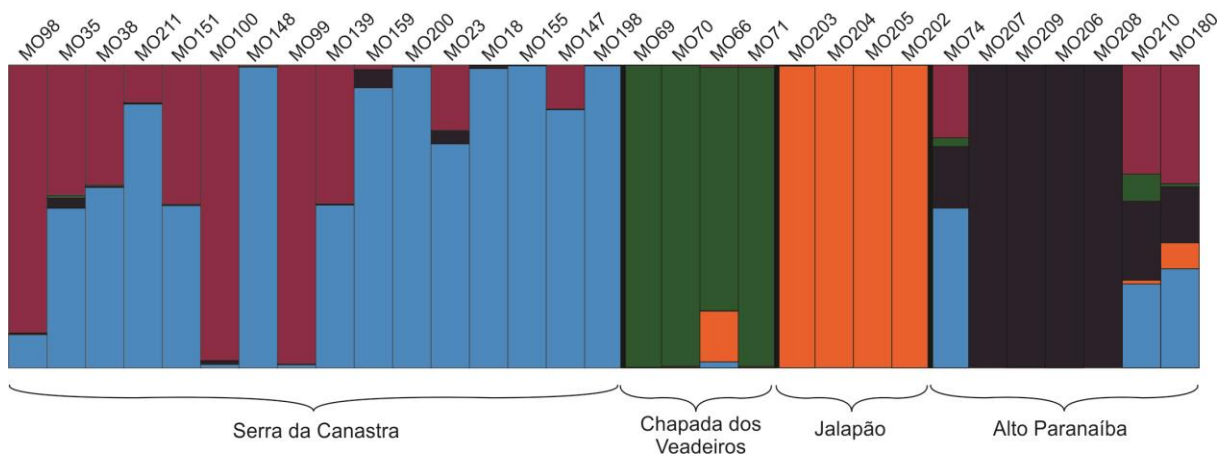


Figura 11: Resultado do Structure no modelo correlacionado, incluindo todos os indivíduos amostrados para a espécie *Mergus octosetaceus*. Azul – alelos comuns no agrupamento do PARNA da Serra da Canastra; verde – alelos comuns no agrupamento formado pelo PARNA da Chapada dos Veadeiros; laranja – alelos comuns no agrupamento PARES do Jalapão; preto – alelos mais comuns no agrupamento da região do Alto Paranaíba.

Como neste programa os indivíduos mais aparentados também podem influenciar o resultado, excluímos os indivíduos aparentados em 1º grau das análises nos dois modelos, em ambos encontramos 3 agrupamentos (Figura 12 e Figura 13).

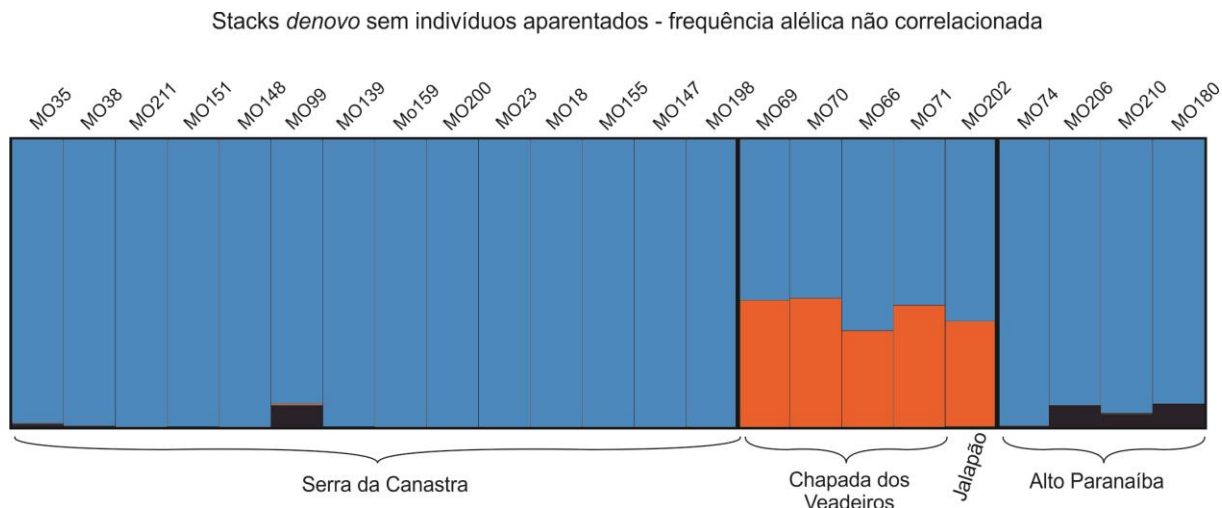


Figura 12: Resultado do Structure, no modelo não correlacionado, excluindo os indivíduos aparentados em primeiro grau para a espécie *Mergus octosetaceus*. Azul – alelos comuns no agrupamento do PARNA da Serra da Canastra; laranja – alelos comuns no agrupamento formado pelo PARNA da Chapada dos Veadeiros + PARES do Jalapão; preto – alelos mais comuns no agrupamento da região do Alto Paranaíba.

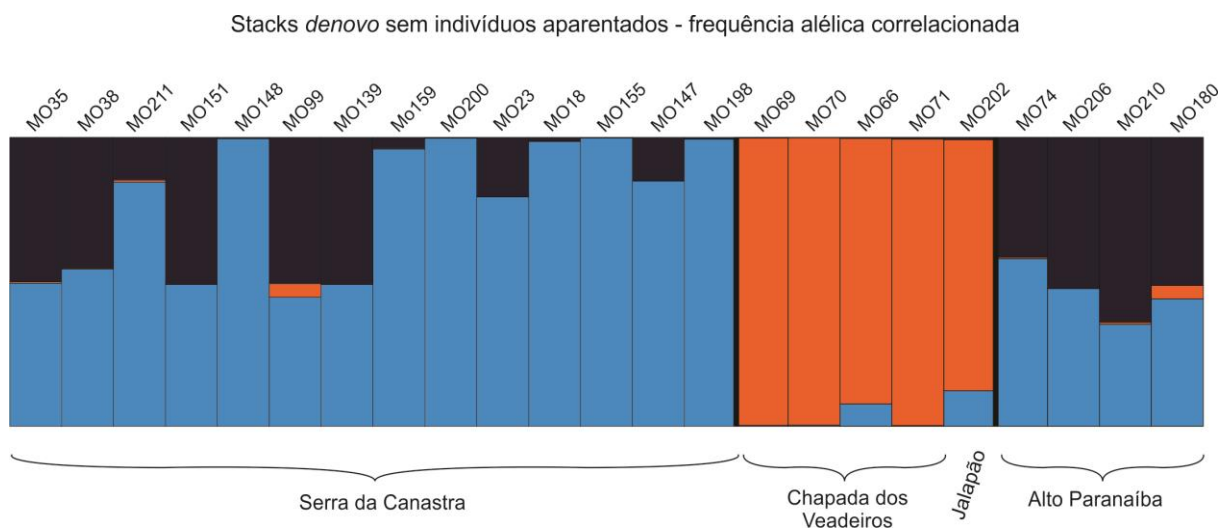


Figura 13: Resultado do Structure, no modelo correlacionado, excluindo os indivíduos aparentados em primeiro grau para a espécie *Mergus octosetaceus*. Azul – alelos comuns no agrupamento do PARNA da Serra da Canastra; laranja – alelos comuns no agrupamento formado pelo PARNA da Chapada dos Veadeiros + PARES do Jalapão; preto – alelos mais comuns no agrupamento da região do Alto Paranaíba.

Os índices de diversidade genética, mostraram que em todas as populações conhecidas, com base na localidade geográfica, a mudança de base mais frequente foi do tipo transição, representando cerca de 79%. A população do PARNA Chapada dos Veadeiros apresenta maior diversidade nucleotídica, a população do PARNA Serra da Canastra apresenta maior número de sítios polimórficos e a população do PARES do Jalapão apresenta menor número de sítios polimórficos e menor diversidade nucleotídica (Tabela 2).

A análise de F_{ST} par a par apresentou valores significativos para todas as comparações ($p < 0,05$) e indicou que a população do PARES do Jalapão é a que mais se diferencia das demais, e que as populações de PARNA Serra da Canastra e da região do Alto Paranaíba são mais semelhantes entre si do que com as demais. (Tabela 3).

Tabela 2: Mudanças de bases mais frequentes e índices de diversidade nucleotídicas de cada população de *M. octosetaceus*.

	Todas as populações N=31	Canastra N=16	Alto paranaíba N=7	Veadeiros N=4	Jalapão N=4
Composição de nucleotídeos	C 32.04% T 17.83% A 18.89% G 31.24%	C 32.23% T 17.55% A 18.72% G 31.50%	C 31.87% T 17.98% A 18.84% G 31.31%	C 31.97% T 18.03% A 18.95% G 31.04%	C 31.64% T 18.48% A 19.63% G 30.26%
Diversidade nucleotídica	0.292556	0.261798	0.243752	0.290609	0.226011
Número de loci	1328	1328	1328	1328	1328
Número de sítios polimórficos	1328	1068	890	889	627
Número de transições	1050	844	708	693	480
Número de transversões	278	224	182	196	147
Número de substituições	1328	1068	890	889	627

Tabela 3: Análise de F_{ST} par a par das populações de *M. octosetaceus*, definidas com base na localidade geográfica.

Population pairwise F_{ST}s				
	Alto Paranaíba	Canastra	Jalapão	Veadeiros
Alto Paranaíba	0.00000			
Canastra	0.10157	0.00000		
Jalapão	0.28612	0.23219	0.00000	
Veadeiros	0.20908	0.14722	0.21373	0.00000

Nível de significância $p < 0,05$

Para fazer a Análise Variância Molecular (AMOVA), testamos vários agrupamentos populacionais, o primeiro conforme determinado pelo Geneland no modelo de frequência alélica não correlacionado, onde as populações do PARNA Serra da Canastra e da região do Alto Paranaíba são agrupadas em um único cluster, o segundo seguindo o resultado do Structure no modelo de frequência alélica não correlacionado onde as populações do PARES do Jalapão e do PARNA Chapada dos Veadeiros pertencem ao mesmo cluster, e um terceiro agrupando as populações do PARNA Serra da Canastra com a região do Alto Paranaíba e PARES do Jalapão com PARNA Chapada dos Veadeiros. Os testes indicaram que a maior parte da variação genética está dentro das populações (79% no agrupamento 1, 82% no agrupamento 2 e 80% no agrupamento 3) (Tabela 4).

O teste de Mantel, não apresentou resultado significativo ($p > 0.05$), o que indica que a distância geográfica por si só não é capaz de explicar a variância molecular.

Tabela 4: AMOVA - Análise de Variância Molecular de 3 níveis para diversos agrupamentos populacionais de *M. octosetaceus*.

AMOVA design and results (average over 1328 loci):				
Agrupamento 1	Source variation	Sum of squares	Variance components	Percentage variation
	Among groups	1410.210	25.93461	12.02006
	Among populations within groups	545.302	19.23751	8.91611
	Within populations	9894.165	170.58905	79.06383
	Total	11849.677	215.76117	
	Average F-Statistics over all loci			
	Fixation Indices			
	FST	0.20936		
	FSC	0.10134		
	FCT	0.12020		
AMOVA design and results (average over 1328 loci):				
Agrupamento 2	Source variation	Sum of squares	Variance components	Percentage variation

Veadeiros +	Among groups	1411.012	-11.46454	-5.56900
	Jalapão			
X	Among populations within groups	544.500	46.73887	22.70383
	Within populations	9894.165	170.58905	82.86517
Canastra	-----			
	Total	11849.677	205.86339	
X	-----			
	Average F-Statistics over all loci			
Alto Paranaíba	Fixation Indices			
	FST	0.17135		
	FSC	0.21506		
	FCT	-0.05569		

AMOVA design and results (average over 1328 loci):

Agrupamento 3	Source variation	Sum of squares	Variance components	Percentage variation

Veadeiros +	Among groups	865.710	14.61317	6.87852
	Jalapão			
X	Among populations within groups	1089.802	27.24423	12.82405
	Within populations	9894.165	170.58905	80.29743
Canastra +	-----			
	Total	11849.677	212.44646	
Alto Paranaíba	-----			
	Average F-Statistics over all loci			
	Fixation Indices			
	FST	0.19703		
	FSC	0.13771		
	FCT	0.06879		

A análise das frequências alélicas indicou que as populações estão apresentando tendência a fixação (alta frequência de alelos em 0 e 1), sendo mais evidente na população do PARES do Jalapão e menos evidente na população do PARES Serra da Canastra (Figura 14).

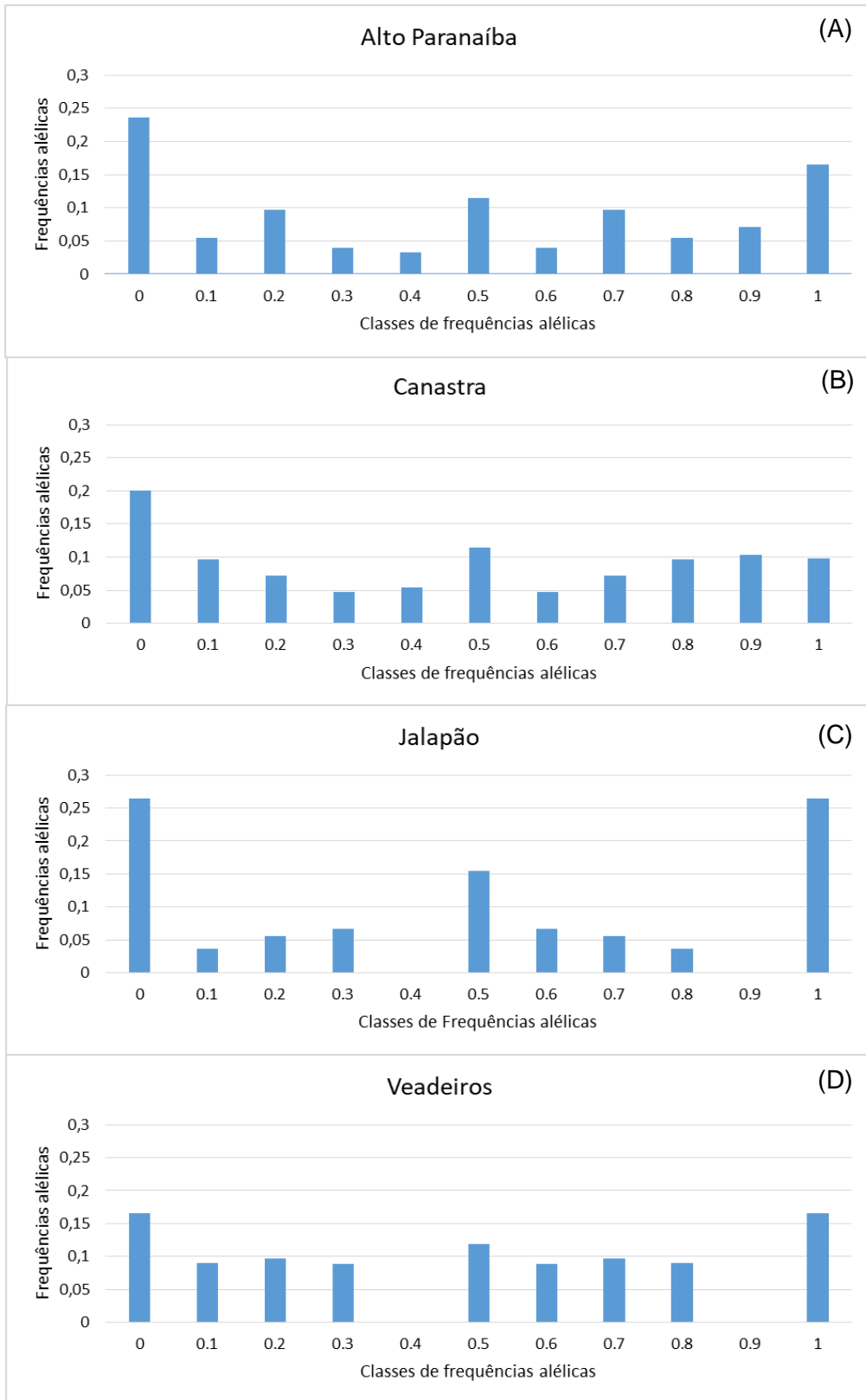


Figura 14: Frequências alélicas das populações de *Mergus octosetaceus*. A – Alto Paranaíba; B - PARNA Serra da Canastra; C – PARES do Jalapão; D – PARNA da Chapada dos Veadeiros.

Discussão

Neste trabalho analisamos 1.328 loci de SNPs gerados a partir de metodologia GBS com sequenciamento de última geração. Delimitando as populações por localidade geográfica identificamos que a população do PARNA Serra da Canastra é a que apresenta a maior quantidade de sítios polimórficos ($n=1.068$), sustentado pela observação de que esta região abriga a maior parte da população do pato-mergulhão (BirdLife International 2016) e também possui a maior parte das amostras utilizadas neste trabalho. No entanto, a população do PARES do Jalapão apresenta o menor número de sítios polimórficos ($n=627$) e também o menor índice de diversidade nucleotídica (0.226011), quando comparado às demais populações, tal resultado pode estar sendo influenciado pelo fato da estimativa populacional do PARES do Jalapão ser baixa ($n=13$ indivíduos, sendo 4 casais e 5 adultos não pareados) (Barbosa et al. 2016) e pode estar também associado a um alto índice de endogamia no local. De qualquer maneira, nossa amostragem para a região é de apenas 4 indivíduos, sendo eles irmãos completos, que não representam totalmente a população.

Comparando a análise de PCA que abrange todos os indivíduos com a que exclui parentes de primeiro grau, não observamos mudanças significativas. Quatro clusters são identificados em ambas as análises, então os indivíduos aparentados em primeiro grau mudaram pouco a conformação do gráfico, mas não de maneira suficiente para que mudasse a interpretação do resultado. A disposição dos *clusters* refletiu a geografia, pois amostras de populações mais próximas geograficamente estão mais próximas também na disposição do gráfico (Figuras 6 e 7). As populações do PARNA Serra da Canastra e da região do Alto Paranaíba possuem uma relação estreita, com pouca diferenciação uma da outra. Tanto na figura 6, quanto na figura 7 identificamos indivíduos pertencentes à região do Alto Paranaíba no *cluster* do PARNA Serra da Canastra. Este resultado pode ser justificado pelo fato de a distância entre as duas áreas ser pequena, de aproximadamente 50 quilômetros em linha reta, entre os registros mais próximos de ambas as áreas. Este resultado também foi encontrado por Maia et al. (2015), que utilizando marcadores mitocondriais, nucleares e microssatélites, encontrou uma maior semelhança entre as populações do PARNA Serra da Canastra e a Região do Alto Paranaíba, quando comparadas à região do PARNA Chapada dos Veadeiros. A pouca distância entre estas áreas de Minas Gerais pode propiciar a dispersão de indivíduos a procura de territórios disponíveis e que não estejam sob pressão antrópica. Apesar dos modos de dispersão do *M. octosetaceus* ainda serem pouco conhecidos (Hugues et al 2006), registros de monitoramento nunca encontraram filhotes nos mesmo rios que nasceram (Ribeiro et al.

2011), e o primeiro estudo sobre a dispersão desta espécie identificou que um indivíduo foi capaz de se deslocar a uma distância de 25,27 km do local de nascimento passando percursos que, inclusive, não possuíam cursos d'água (Ribeiro et al. 2011).

O número de agrupamentos populacionais e seus limites geográficos inferidos pelo Geneland apresentaram divergências quando utilizamos os dois modelos de frequência alélica. Quando utilizamos o modelo de frequência alélica não correlacionado, encontramos que as populações pertencentes ao PARNA Serra da Canastra e da região do Alto Paranaíba agruparam em um único *cluster*, informação que condiz com a análise de PCA, onde constatamos que indivíduos da região do Alto Paranaíba encontravam-se no *cluster* de indivíduos do PARNA Serra da Canastra. Já com o modelo de frequência alélica correlacionado identificamos que a população da região do Alto Paranaíba se divide em duas, uma parte pertencendo o *cluster* do PARNA da Serra da Canastra e outra formando um *cluster* único. Tal discrepância de resultados também se deve ao fato que, quando utilizamos o modelo de frequência alélica correlacionado, o programa é capaz de identificar populações com um F_{ST} muito baixo, que aliado a um tamanho amostral reduzido pode acabar superestimando a inferência do número populacional (Latch et al. 2006, Colonna et al. 2009).

Os agrupamentos populacionais sem correlação geográfica encontrados pelo Structure também apresentaram divergência entre os modelos de frequência alélica, assim como quando utilizamos indivíduos aparentados em primeiro grau nas análises. Quando analisamos todos os indivíduos juntamente com o modelo de frequência alélica correlacionado, a estimativa do valor de K foi inflada ($k=5$), pois indivíduos muito semelhantes geneticamente, tendem a se agrupar, fazendo com que o programa identifique um K exclusivo para eles (Pritchard et al. 2000). Quando retiramos os parentes de 1º grau, encontramos $K=3$ para ambos os modelos de frequência alélica, porém no modelo correlacionado as populações tendem a apresentar uma diferenciação um pouco maior quando comparado ao outro modelo, isto também se deve ao fato que neste modelo o programa tende a maximizar pequenas diferenças encontradas, podendo até atribuí-las a diferenças populacionais (Latch et al. 2006, Colonna et al. 2009).

A similaridade encontrada entre as populações do PARNA Serra da Canastra e da região do Alto Paranaíba pelo Geneland é confirmada também pela análise de F_{ST} par a par, onde encontramos menores índices de fixação entre estas duas populações quando comparadas com as demais, corroborando um resultado obtido por microssatélites e mtDNA

(Maia et al. (2015). Os índices de fixação apresentaram valores moderados, indicando que apesar dos sinais de estruturação populacional, não existe limitação de fluxo gênico.

Os resultados da AMOVA para os três agrupamentos populacionais foram semelhantes, todos indicaram que a maior parte da variação genética está dentro das populações, este fato nos mostra que a população é bastante semelhante geneticamente. Os resultados indicam que o agrupamento 1 (Canastra + Alto Paranaíba x Veadeiros x Jalapão) é o agrupamento de populações mais provável por apresentar maior valor de F_{CT} (0.12020). O valor de F_{ST} para os três agrupamentos foram significativos (0.20 para o agrupamento 1, 0.17 para o agrupamento 2 e 0.19 para o agrupamento 3), porém não apresentaram valores elevados, indicando que apesar de apresentar sinais de estruturação populacional não existe limitação de fluxo gênico entre as populações. Assim como pudemos observar também no resultado do Structure, que indica que as populações compartilham alguns alelos, refutando o resultado encontrado por Vilaça et al. 2012 que utilizando marcadores mitocondriais identificou limitações de fluxo gênico entre a população do PARNA Serra da Canastra e PARNA Chapada dos Veadeiros.

As frequências alélicas em todas as populações apresentaram tendências a fixação de alelos. Esta tendência também foi identificada por Maia et al. (2015) utilizando microsatélites e pode indicar eventos de *bottleneck*, sendo mais evidente na população do PARES do Jalapão. Porém, esta localidade apresenta uma estimativa populacional muito baixa ($n=13$) (Barbosa et al. 2016), o que pode acarretar em uma possível endogamia no local, e além disso, nossa amostragem se resume a 4 indivíduos, sendo eles irmãos completos e este fato pode não representar a real situação da população. O PARNA Serra da Canastra foi o local que apresentou uma tendência mais moderada a fixação de alelos, tal fato pode ser justificado pois neste local encontra-se a maior população conhecida da espécie (BirdLife International 2016), sendo este um fator que pode reduzir a endogamia local, e também esta é a região que possui o maior número de amostras utilizadas neste trabalho ($n=16$). A região do Alto Paranaíba apresentou resultados semelhantes ao PARNA Serra da Canastra, este resultado já era esperado já que as duas localidades são mais próximas, quando comparadas com as demais. Já o PARNA Chapada dos Veadeiros apresentou resultado semelhante ao PARES do Jalapão, apesar da distância entre as duas áreas ser razoavelmente grande (em torno de 500 km), estas duas áreas também apresentaram resultados semelhantes no Structure, outro fator que pode influenciar tal fato é a amostragem reduzida para esta área ($n=4$) que pode não representar a real situação da população.

Conclusão

Nossos resultados indicam que apesar de sinais de estruturação populacional, as populações de *Mergus octosetaceus* apresentam baixa variabilidade genética. As populações do PARNA Serra da Canastra e da região do Alto Paranaíba se comportam como uma única população, compartilhando muitos alelos. A estruturação populacional encontrada não impede cruzamentos entre indivíduos destas áreas, o que pode fornecer um plantel de indivíduos com boa diversidade genética para a população de cativeiro. As regiões do PARES do Jalapão e do PARNA Chapada dos Veadeiros devem ser melhor amostradas, pois o número amostral utilizado aqui foi mais baixo que as demais localidades e isto pode não representar a população como um todo.

Eventos de *bottleneck* podem ter acontecido nas populações e ainda estar causando danos, como altas taxas de endogamia nas áreas remanescentes. Este evento pode reduzir a resiliência da espécie a possíveis mudanças ambientais.

A baixa variabilidade genética das populações de *Mergus octosetaceus* tornam urgentes medidas mais eficazes para a conservação da espécie como a recuperação das áreas degradadas onde a espécie ocorria historicamente, para uma possível reintrodução, redução da influência antrópica nos locais onde a espécie ocorre, cruzamentos *ex-situ* de modo a aumentar a variabilidade genética de indivíduos que serão reintroduzidos e melhores planos de manejo para as áreas remanescentes do pato-mergulhão. Estas ações, previstas no Plano de Ação Nacional do pato-mergulhão, devem ser executadas o quanto antes para que a espécie consiga sobreviver no ambiente natural.

Referências

- Albert, T. J., Molla, M. N., Muzny, D. M., Nazareth, L., Wheeler, D., Song, X., Gibbs, R. a. (2007). Direct selection of human genomic loci by microarray hybridization. *Nature Methods*, 4(11), 903–905. <https://doi.org/10.1038/nmeth1111>
- Allendorf, F. W., Hohenlohe, P. a, & Luikart, G. (2010). Genomics and the future of conservation genetics. *Nature Reviews. Genetics*, 11(10), 697–709. <https://doi.org/10.1038/nrg2844>
- Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, 3(10), 1–7. <https://doi.org/10.1371/journal.pone.0003376>
- Barbosa, M. O., Pinheiro, R. T., & Barbosa, K. V. C. (2016). Population estimate of *Mergus octosetaceus* in the Jalapão region, Tocantins, Brazil. *Revista Brasileira de Ornitologia*, 23(4), 417–427.
- Bartmann, W. (1988). New observations on Brazilian merganser. *Wildfowl*, 39, 7–14.
- Benstead, P. J., Hearn, R. D. & Nedelcoff, A. R. S. (1994) A recent sighting of Brazilian Merganser *Mergus octosetaceus* in Misiones province, Argentina. *Cotinga* 2: 35–36.
- Bianchi, C. A.; Brant, S.; Brandão, R. A. e Brito, B. F. (2005). New records of Brazilian Merganser *Mergus octosetaceus* in the rio das Pedras, Chapada dos Veadeiros, Brazil. *Cotinga*, 24:72-74.
- BirdLife International. 2016. *Mergus octosetaceus*. The IUCN Red List of Threatened Species 2016: e.T22680482A92863947. <http://dx.doi.org/10.2305/IUCN.UK.2016-3.RLTS.T22680482A92863947.en>. Downloaded on 26 January 2017.
- Braz, V. S., Abreu, T. L. S., Lopes, L. E., Leite, L. O., França, F. G. R., Vasconcelos, M. M., Balbino, S. F. (2003). Brazilian Merganser *Mergus octosetaceus* discovered in Jalapão State Park, Tocantins, Brazil. *Cotinga*, 20, 68–71.
- Brian, A. (1992). The awk programming language. *Computers and the Humanities* (Vol. 26). <https://doi.org/10.1007/BF00054275>

Bruno, S. F. Biologia e Conservação do pato-mergulhão (*Mergus octosetaceus* Vieillot, 1817) no parque nacional da Serra da Canastra e entorno, Minas Gerais, Brasil (2004). 45f. Monografia (Especialização em Biologia) – Universidade Federal de Lavras, Minas Gerais,

Bruno, S.F. Biologia e conservação do pato-mergulhão (*Mergus octosetaceus* Vieillot, 1817) no Parque Nacional da Serra da Canastra e Entorno, Minas Gerais, Brasil. Lavras, 2004. 43 p. Monografia (Pós-Graduação em Biologia) – Instituto de Biologia, Universidade Federal de Lavras, MG.

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). Stacks : Building and Genotyping Loci De Novo From Short-Read Sequences. *G3: Genes|Genomes|Genetics*, 1(3), 171–182. <https://doi.org/10.1534/g3.111.000240>

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., Postlethwait, J. H., & De Koning, D.-J. (2011). Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences. *G3: Genes|Genomes|Genetics*, 1(3), 171–182. <https://doi.org/10.1534/g3.111.000240>

Collar, N. J.; Gonzaga, L. P.; Krabbe, N. et al. Threatened Birds of the Americas: The ICBP/IUCN Red Data Book. Cambridge, Uk: ICBP, 1992.

Colonna, V., Natile, T., Ferrucci, R. R., Fardella, G., Aversano, M., Barbujani, G., & Ciullo, M. (2009). Comparing population structure as inferred from genealogical versus genetic information, 1635–1641. <https://doi.org/10.1038/ejhg.2009.97>

Danecek, P., Auton, A., Abecasis, G., Albers, C. a, Banks, E., DePristo, M. A., Durbin, R. (2011). The variant call format and VCF tools. *Bioinformatics (Oxford, England)*, 27(15), 2156–8. <https://doi.org/10.1093/bioinformatics/btr330>

Davey, J., Hohenlohe, P., Etter, P., Boone, J., Catchen, J., & Blaxter, M. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, 12(7), 499–510. <https://doi.org/10.1038/nrg3012>

De Donato, M., Peters, S. O., Mitchell, S. E., Hussain, T., & Imumorin, I. G. (2013). Genotyping-by-Sequencing (GBS): A Novel, Efficient and Cost-Effective Genotyping Method for Cattle Using Next-Generation Sequencing. *PLoS ONE*, 8(5). <https://doi.org/10.1371/journal.pone.0062137>

De Oliveira Barbosa, M., & De Almeida, M. L. (2010). Novas observações e dados reprodutivos do pato-mergulhão *Mergus octosetaceus* na região do Jalapão, Tocantins, Brasil. *Cotinga*, 32(1), 109–113.

Dierickx, E. G., Shultz, A. J., Sato, F., Hiraoka, T., & Edwards, S. V. (2015). Morphological and genomic comparisons of Hawaiian and Japanese Black-footed Albatrosses (*Phoebastria nigripes*) using double digest RADseq: Implications for conservation. *Evolutionary Applications*, 8(7), 662–678. <https://doi.org/10.1111/eva.12274>

Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4(2), 359–361. <https://doi.org/10.1007/s12686-011-9548-7>

Eaton, J. W., Bateman, D., Hauberg, S., Wehbring, R. (2016). GNU Octave version 4.2.0 manual: a high-level interactive language for numerical computations. URL <http://www.gnu.org/software/octave/doc/interpreter/>

Elshire R, Glaubitz J, Sun Q, et al. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6: e19379. doi:10.1371/journal.pone.0019379.

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10(3), 564–567. <https://doi.org/10.1111/j.1755-0998.2010.02847.x>

Fragano, F. and Clay, R. (2003) Biodiversity status of the Interior Atlantic Forest of Paraguay. Chapter 25 in C. Gallindo-Leal and I. G. Câmara, eds. *The Atlantic Forest of South America: biodiversity status, threats, and outlook*. Washington, DC: Island Press.

Frankham, R. (1995). Conservation genetics. *Annu Rev Genet*, 29, 305–327. <https://doi.org/10.1146/annurev.genet.29.1.305>

Frankham, R. (2010). Challenges and opportunities of genetic approaches to biological conservation. *Biological Conservation*. <https://doi.org/10.1016/j.biocon.2010.05.011>

Frankham, R., Ballou, J.D. and Briscoe, D.A. (2002). *Introduction to conservation genetics* Cambridge University Press, Cambridge, UK.

- Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q., & Buckler, E. S. (2014). TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. *PLoS ONE*, 9(2). <https://doi.org/10.1371/journal.pone.0090346>
- Guillot, G., Santos, F., & Estoup, A. (2008). Analysing georeferenced population genetics data with Geneland: A new algorithm to deal with null alleles and a friendly graphical user interface. *Bioinformatics*, 24(11), 1406–1407. <https://doi.org/10.1093/bioinformatics/btn136>
- Hearn, R. (1994). The current status of the Brazilian Merganser (*Mergus octosetaceus*) in Argentina. *Threatened Waterfowl Research Group Newsletter*, WWT, Slimbridge, UK, v. 5, p. 14- 15.
- Hedrick, P. W. (2001). Conservation genetics: where are we now? *Trends in Ecology & Evolution*, 16(11), 629–636. [https://doi.org/10.1016/S0169-5347\(01\)02282-0](https://doi.org/10.1016/S0169-5347(01)02282-0)
- Hughes, B., Dugger, B., Cunha, H. J., Lamas, I. R., Goerck, J. M., Lins, L. V., et al. (2006). Plano de ação para a conservação do pato-mergulhão (*Mergus octosetaceus*). Brasília: IBAMA.
- Hughes, J. B., Daily, G. C., & Ehrlich, P. R. (2002). Conservation of tropical forest birds in countryside habitats. *Ecology Letters*, 5(1), 121-129.
- İpek, A., Yılmaz, K., Sıkıcı, P., Tangu, N. A., Öz, A. T., Bayraktar, M., Gülen, H. (2016). SNP Discovery by GBS in Olive and the Construction of a High-Density Genetic Linkage Map. *Biochemical Genetics*, 54(3), 313–325. <https://doi.org/10.1007/s10528-016-9721-5>
- Johnson J.A., Watson R.T., Mindell D.P. (2005) Prioritizing species conservation: does the Cape Verde kite exist? *Proc. R. Soc. Lond. B* 272, 1365–1371.
- Johnson, J. L., Wittgenstein, H., Mitchell, S. E., Hyma, K. E., Temnykh, S. V., Kharlamova, A. V., Kukekova, A. V. (2015). Genotyping-by-sequencing (GBS) detects genetic structure and confirms behavioral QTL in tame and aggressive foxes (*Vulpes vulpes*). *PLoS ONE*, 10(6), 1–22. <https://doi.org/10.1371/journal.pone.0127013>
- Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, 15(5), 1179–1191. <https://doi.org/10.1111/1755-0998.12387>

- Lamas, I. R. (2006). Census of Brazilian Merganser *Mergus octosetaceus* in the region of Serra da Canastra National Park, Brazil, with discussion of its threats and conservation. *Bird Conservation International*, 16(2), 145. <https://doi.org/10.1017/S0959270906000220>
- Latch, E. K., Dharmarajan, G., Glaubitz, J. C., & Jr, O. E. R. (2006). Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation, 295–302. <https://doi.org/10.1007/s10592-005-9098-1>
- Li, J., Chen, X., Kang, B., & Liu, M. (2015). Mitochondrial DNA Genomes Organization and Phylogenetic Relationships Analysis of Eight Anemonefishes (Pomacentridae: Amphiprioninae). *Plos One*, 10(4), e0123894. <https://doi.org/10.1371/journal.pone.0123894>
- Liu, G., Zhou, L. Z., & Gu, C. M. (2012). Complete sequence and gene organization of the mitochondrial genome of scaly-sided merganser (*Mergus squamatus*) and phylogeny of some Anatidae species. *Molecular Biology Reports*, 39(3), 2139–2145. <https://doi.org/10.1007/s11033-011-0961-5>
- Livezey, B. C. (1986). A phylogenetic Analysis of Recent Anseriform Genera using Morphological Characters. *The Auk*, (October), 737–754.
- Lozier, J. D. (2014). Revisiting comparisons of genetic diversity in stable and declining species: Assessing genome-wide polymorphism in North American bumble bees using RAD sequencing. *Molecular Ecology*, 23(4), 788–801. <https://doi.org/10.1111/mec.12636>
- Lu, F., Glaubitz, J., Harriman, J., Casstevens, T., & Elshire, R. (2012). TASSEL 3.0 Universal Network Enabled Analysis Kit (UNEAK) pipeline documentation. White Paper, 1–12.
- Luca, F., Hudson, R. R., Witonsky, D. B., Luca, F., Hudson, R. R., Witonsky, D. B., & Rienzo, A. Di. (2011). and applications to human evolution A reduced representation approach to population genetic analyses and applications to human evolution, 1087–1098. <https://doi.org/10.1101/gr.119792.110>
- Maia, T. A. (2015). Genética da conservação do pato-mergulhão *Mergus octosetaceus* Viellot,1817. Dissertação de Mestrado. Pontifícia Universidade Católica de Minas Gerais, Belo Horizonte, Minas Gerais, Brasil.

Matthews, P. D., Coles, M. C., & Pitra, N. J. (2013). Next Generation Sequencing for a Plant of Great Tradition: Application of NGS to SNP Detection and Validation in Hops (*Humulus lupulus* L.). *BrewingScience*, 66(11–12), 185–191.

Meffe, G. K. (1986). Conservation Genetics and the Management of Endangered Fishes. *Fisheries*, 11(1), 14–23. [https://doi.org/10.1577/1548-8446\(1986\)011<0014](https://doi.org/10.1577/1548-8446(1986)011<0014)

MMA (2003) Anexo à Instrução Normativa no. 3, de 27 de maio de 2003, do Ministério do Meio Ambiente. Lista das Espécies da Fauna Brasileira Ameaçadas de Extinção. <http://www.ibama.gov.br/fauna/downloads/lista%20spp.pdf>

Narum, S. R., Buerkle, C. A., Davey, J. W., Miller, M. R., & Hohenlohe, P. A. (2013). Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology*, 22(11), 2841–2847. <https://doi.org/10.1111/mec.12350>

Paris, J. R., Stevens, J. R., & Catchen, J. M. (2017). Lost in parameter space: A road map for stacks. *Methods in Ecology and Evolution*. <https://doi.org/10.1111/2041-210X.12775>

Partridge, W. H. (1956). Notes on Brazilian merganser in Argentina. *The Auk*, 73(4), 473–488.

Pellegrino, I., Boatti, L., Cucco, M., Mignone, F., Kristensen, T. N., Mucci, N., ... Pertoldi, C. (2016). Development of SNP markers for population structure and phylogeography characterization in little owl (*Athene noctua*) using a genotyping-by-sequencing approach. *Conservation Genetics Resources*, 8(1), 13–16. <https://doi.org/10.1007/s12686-015-0513-8>

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE*, 7(5). <https://doi.org/10.1371/journal.pone.0037135>

Pritchard, J., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2), 945–959. <https://doi.org/10.1111/j.1471-8286.2007.01758.x>

Puritz, J. B., Hollenbeck, C. M., & Gold, J. R. (2014). dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ*, 2, e431. <https://doi.org/10.7717/peerj.431>

Querejeta, M., González-Esteban, J., Gómez, A., Fernández-González, A., Aymerich, P., Gosálbez, J., Castresana, J. (2016). Genomic diversity and geographical structure of the

Pyrenean desman. *Conservation Genetics*, 17(6), 1333–1344.
<https://doi.org/10.1007/s10592-016-0865-y>

Ribeiro, F. (2016). Biologia reprodutiva do pato-mergulhão, *Mergus octosetaceus* na região do Parque Nacional da Serra da Canastra, Minas Gerais, Brasil. Dissertação de mestrado. Universidade Federal de Minas Gerais. Belo Horizonte, Minas Gerais, Brasil.

Ribeiro, F., Lins, L. V., Gomes, V. M., Nery, F. H., & dos Reis, E. S. (2011). Dispersão e maturidade sexual de *Mergus octosetaceus* Vieillot, 1817 na região da Serra da Canastra, Minas Gerais, Brasil. *Revista Brasileira de Ornitologia*, 19(3), 391–397.

Sambrook, J., & Russell, D. W. (2001). *Molecular cloning: a laboratory manual*. Cold Spring Harbour.

Schiegg, K., Daniels, S. J., Walters, J. R., Priddy, J. A., & Pasinelli, G. (2006). Inbreeding in red-cockaded woodpeckers: effects of natal dispersal distance and territory location. *Biological Conservation*, 131(4), 544-552.

Sick, H. *Ornitologia brasileira, uma introdução*. Brasília: Universidade de Brasília, 1985

Silveira, L. F., & Bartmann, W. D. (2001). Natural history and conservation of Brazilian Merganser *Mergus octosetaceus* at Serra da Canastra National Park, Minas Gerais, Brazil. *Bird Conservation International*, 11(4), 287–300.
<https://doi.org/10.1017/S0959270901000338>

Toews, D. P. L., Brelsford, A., Grossen, C., Milá, B., & Irwin, D. E. (2016). Genomic variation across the Yellow-rumped Warbler species complex. *The Auk*, 133(4), 698–717.
<https://doi.org/10.1642/AUK-16-61.1>

Velmurugan, J., Mollison, E., Barth, S., Marshall, D., Milne, L., Creevey, C. J., Milbourne, D. (2016). An ultra-high density genetic linkage map of perennial ryegrass (*Lolium perenne*) using genotyping by sequencing (GBS) based on a reference shotgun genome assembly. *Annals of Botany*, mcw081. <https://doi.org/10.1093/aob/mcw081>

Vilaça, S. T., Redondo, R. A. F., Lins, L. V., & Santos, F. R. (2012). Remaining genetic diversity in Brazilian Merganser (*Mergus octosetaceus*). *Conservation Genetics*, 13(1), 293–298. <https://doi.org/10.1007/s10592-011-0262-5>

White, T. A., Perkins, S. E., Heckel, G., & Searle, J. B. (2013). Adaptive evolution during an ongoing range expansion: The invasive bank vole (*Myodes glareolus*) in Ireland. *Molecular Ecology*, 22(11), 2971–2985. <https://doi.org/10.1111/mec.12343>

Witzenberger, K. A., & Hochkirch, A. (2011). Ex situ conservation genetics: A review of molecular studies on the genetic consequences of captive breeding programmes for endangered animal species. *Biodiversity and Conservation*, 20(9), 1843–1861. <https://doi.org/10.1007/s10531-011-0074-4>

Yamashita, C., & Valle, M. D. P. (1990). Ocorrência de duas aves raras no Brasil Central: *Mergus octosetaceus* e *Tigrisoma fasciatum fasciatum*. *Ararajuba*.