



UNIVERSIDADE FEDERAL DO AMAZONAS - UFAM
INSTITUTO DE CIÊNCIAS BIOLÓGICAS - ICB
PROGRAMA DE PÓS-GRADUAÇÃO EM ZOOLOGIA - PPGZOO

**Genética da Conservação do gavião-real (*Harpia harpyja*,
Linnaeus, 1758) no Brasil**

MYLENA KAIZER XAVIER FONSECA

Manaus, Amazonas
Julho/2023



UNIVERSIDADE FEDERAL DO AMAZONAS - UFAM
INSTITUTO DE CIÊNCIAS BIOLÓGICAS - ICB
PROGRAMA DE PÓS-GRADUAÇÃO EM ZOOLOGIA - PPGZOO

**Genética da Conservação do gavião-real (*Harpia harpyja*,
Linnaeus, 1758) no Brasil**

Discente: Mylena Kaizer Xavier Fonseca

Orientador: Prof. Dr. Tomas Hrbek

Coorientador: Prof. Dr. Aureo Banhos dos Santos

Dissertação apresentada à Universidade Federal do Amazonas como parte dos requisitos para obtenção do título de Mestre pelo Programa de Pós-Graduação em Zoologia.

Manaus, Amazonas

Julho/2023

Ficha Catalográfica

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

F676g Fonseca, Mylena Kaizer Xavier
Genética da Conservação do gavião-real (*Harpia harpyja*,
Linnaeus, 1758) no Brasil / Mylena Kaizer Xavier Fonseca . 2023
84 f.: il. color; 31 cm.

Orientador: Tomas Hrbek
Coorientador: Aureo Banhos dos Santos
Dissertação (Mestrado em Zoologia) - Universidade Federal do
Amazonas.

1. Harpia. 2. Diversidade genética. 3. Genética da Conservação.
4. Genômica. 5. Marcadores moleculares. I. Hrbek, Tomas. II.
Universidade Federal do Amazonas III. Título

MYLENA KAIZER XAVIER FONSECA

**Genética da Conservação do gavião-real (*Harpia harpyja*, Linnaeus, 1758)
no Brasil**

Orientador: Prof. Dr. Tomas Hrbek

Coorientador: Prof. Dr. Aureo Banhos dos Santos

Dissertação apresentada à Universidade Federal do Amazonas como parte dos requisitos para obtenção do título de Mestre pelo Programa de Pós-Graduação em Zoologia.

APROVADA EM: 21 de Julho de 2023

BANCA EXAMINADORA

Prof. Dr. Tomas Hrbek
Universidade Federal do Amazonas

Prof(a). Dr(a). Cintia Cornelius Frische
Universidade Federal do Amazonas

Prof(a). Dr(a). Cristina Miyaki
Universidade de São Paulo

AGRADECIMENTOS

Ao meu orientador, Dr. Tomas Hrbek, pela valiosa oportunidade de fazer parte de seu grupo de pesquisa, pelos grandes ensinamentos durante a minha pós-graduação e pelo apoio técnico e intelectual essenciais para instruir e auxiliar nas análises genéticas desenvolvidas nesta dissertação.

Ao meu coorientador, Dr. Aureo Banhos, que me acompanha desde a graduação, me incentivando e viabilizando a pesquisa com o gavião-real. Agradeço por me acolher e ajudar a me estabelecer em Manaus quando iniciei o mestrado e por todo apoio e conhecimento para construir esse trabalho.

À professora Dra. Izeni Farias, pelas aulas “particulares” e pelas ótimas conversas em sua sala; por me acompanhar e me ajudar com tanta empolgação a trilhar um belo caminho até aqui, agradeço por todo o conhecimento compartilhado comigo.

Ao Laboratório de Evolução e Genética Animal – LEGAL, onde a maioria das etapas laboratoriais foi desenvolvida. Ao grupo de pesquisa que tive oportunidade de conhecer no LEGAL e que me ajudaram de alguma forma na realização desta pesquisa: Adriano, Carlos, Edvaldo, Fabrício, Fábio, Flávio, Ícaro, Ingrid, Joice, Jose, Maria Augusta, Mário, Mayara, Nasrah, Sandra, Sarah, Shizuka e Valéria. Em especial, agradeço o Pedro Bittencourt e o Érico Polo pela grande colaboração nas análises genéticas e pelas longas discussões sobre os resultados obtidos.

À Universidade Federal do Amazonas e ao Programa de Pós-Graduação em Zoologia pela oportunidade de realizar minha pesquisa na Amazônia e pelo apoio em todas as questões burocráticas. Aos professores da pós-graduação que contribuíram para minha formação profissional compartilhando seus conhecimentos com tanta sabedoria. Aos professores Dr. Diego Sotero e Dra. Doriane Picanço por me aceitarem em sua sala de aula para eu realizar o estágio docente, sem dúvidas, essa experiência agregou muito em minha carreira profissional. Aos meus colegas de mestrado que ingressaram no curso comigo. Em especial agradeço aos meus amigos mais próximos, Bryan, Cianir, Cláudia, Joana, Lucas, Silionamã que contribuíram em meu crescimento profissional e pessoal, obrigada pela amizade de vocês.

À coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES pela concessão da bolsa de estudo e a capacitação que recebi durante o mestrado. Ao Projeto Harpia por disponibilizar amostras do gavião-real para realizar este estudo. Aos Zoológicos e instituições conservacionistas, listados no primeiro capítulo desta dissertação, por disponibilizarem amostras de sangue dos gaviões-reais cativos. Ao apoio do Instituto Chico Mendes de Conservação da Biodiversidade (ICMBio). Agradeço também a todos os apoios financeiros concedidos ao Projeto Harpia que possibilitaram a realização deste trabalho: Vale S.A, Veracel Celulose S.A, Fundação Tiergarten-Nürnberg, Associação Beauval Nature e Instituto Últimos Refúgios. Ao Instituto Tecnológico da Vale pela recepção e treinamento que recebi durante os dias em que estive na instituição em Belém, onde parte deste estudo foi desenvolvida.

À minha mãe, Wania Kaizer, ao meu pai, Nelson Fonseca e a minha irmã, Rayanne Kaizer, por todo apoio emocional desde quando decidi fazer o mestrado em Manaus, vocês estiveram presentes todos os dias mesmo com a distância trazendo calma e um amor incondicional. À minha amiga Isabela Procópio por todo apoio em Manaus, me ajudou inúmeras vezes com paciência e dedicação, costumo dizer que ela foi meu anjo da guarda durante esses anos, obrigada pela sua amizade. Às minhas amigas, fiéis escudeiras, Letícia Miranda e Gabriela Munis que me acompanharam desde o princípio, a minha caminhada até aqui não seria a mesma sem vocês, obrigada por tudo.

Muito obrigada a todos!

EPÍGRAFE

“A persistência é o caminho do êxito.”

Charles Chaplin

RESUMO

O gavião-real, uma ave de rapina também conhecida como harpia ou uiraçu, é a maior águia das Américas, ocorrendo em florestas tropicais da América Central e do Sul e depende desses ambientes para sua sobrevivência. No Brasil, o gavião-real está presente principalmente na Amazônia e na Mata Atlântica, com raros registros na região central do Brasil. A caça e a fragmentação florestal ameaçam sua existência, levando essa espécie ser classificada como ameaçada de extinção, situada na categoria Vulnerável em toda sua distribuição. A Genética da Conservação visa preservar espécies ameaçadas ou em vias de extinção e a coleta das informações genéticas pode ser realizada a partir de marcadores moleculares, como os microssatélites (SSRs), polimorfismos de nucleotídeo único (SNPs) e o DNA mitocondrial. Esses marcadores são comumente usados na Genética da Conservação para compreender a história evolutiva da espécie, avaliar a diversidade genética e a presença de estrutura populacional para desenvolver estratégias de manejo adequadas visando preservar a variabilidade genética existente ao longo do tempo e espaço. O manejo pode ser realizado em habitat natural (*in situ*) e em zoológicos e criadouros conservacionistas (*ex situ*). Para o gavião-real as ações *in situ* focam na proteção dos ninhos da espécie e de seus territórios e as ações *ex situ* focam principalmente no resgate, reabilitação, destinação dos indivíduos feridos e reprodução em cativeiro. Neste estudo desenvolvemos dez novos *loci* microssatélites específicos para o gavião-real contendo alto polimorfismo e com poder de exclusão de parentesco e de identificação individual para auxiliar principalmente no manejo *ex situ* (Capítulo I). Os *loci* desenvolvidos não apresentaram alelos nulos, não houve desequilíbrio de ligação e apenas quatro *loci* desviaram da expectativa do equilíbrio de Hardy-Weinberg. Além disso, demonstraram eficiência em excluir os prováveis pais de um indivíduo, realizar a identificação individual e detectar os níveis de parentesco entre os indivíduos. Assim, o novo conjunto de microssatélites pode ser utilizado para auxiliar no direcionamento de estratégias do manejo *ex situ*, visando melhores pareamentos reprodutivos, mantendo a diversidade genética e evitando a endogamia e exogamia. Neste estudo também foi investigada a presença de estrutura genética para o gavião-real no Brasil a partir de polimorfismos de nucleotídeo único (SNPs) e do mitogenoma gerados por um sequenciamento com baixa cobertura (Capítulo II). Quanto aos SNPs, todos os *pipelines* utilizados resgataram um baixo sinal de estrutura genética para o gavião-real no Brasil. O *pipeline* do *DiscoSnpRad* demonstrou maior eficácia para lidar com o sequenciamento de baixa cobertura resgatando informações significativas e consistentes com os resultados encontrados pelo mitogenoma, com a população localizada ao Norte do Rio Amazonas apresentando maior diferenciação genética e as populações localizadas ao Sul do Rio Amazonas e na Mata Atlântica apresentando semelhanças genéticas. Dessa forma, os resultados obtidos nos dois capítulos desenvolvidos neste estudo fornecem ferramentas moleculares e informações que auxiliam no direcionamento de estratégias de manejo e em tomadas de decisões para a conservação *in situ* e *ex situ* do gavião-real.

Palavras-chave: Harpia. Diversidade genética. Genética da Conservação. Genômica. Marcadores moleculares.

ABSTRACT

The Harpy Eagle, also known as the “harpia” or “uiraçu”, is the largest eagle in the Americas, occurring in tropical forests of Central and South America and depending on these environments for its survival. In Brazil, the Harpy Eagle is mainly found in the Amazon Rainforest and the Atlantic Forest, with rare records in the central region of Brazil. Hunting and forest fragmentation threaten its existence, leading this species to be classified as endangered, listed as Vulnerable throughout its distribution. Conservation Genetics aims to preserve endangered or threatened species, and the collection of genetic information can be carried out using molecular markers such as microsatellites (SSRs), single nucleotide polymorphisms (SNPs), and mitochondrial DNA. These markers are commonly used in Conservation Genetics to understand the species evolutionary history, assess genetic diversity and the presence of population structure, and develop appropriate management strategies to preserve the existing genetic variability over time and space. Management can be carried out in the species natural habitat (*in situ*) as well as in zoos and conservation breeding centers (*ex situ*). For the Harpy Eagle, *in situ* actions focus on protecting the species nests and territories, while *ex situ* actions mainly involve rescuing, rehabilitating, providing care for injured individuals, and breeding in captivity. In this study, we developed ten new microsatellite loci specific to the Harpy Eagle, showing high polymorphism and the ability to exclude parentage and individual identification, primarily aiding *ex situ* management (Chapter I). The developed loci showed no null alleles, no linkage disequilibrium, and only four loci deviated from Hardy-Weinberg equilibrium expectations. Additionally, they demonstrated efficiency in excluding the likely parents of an individual, performing individual identification, and detecting levels of relatedness among individuals. Thus, this new set of microsatellites can be used to assist in guiding *ex situ* management strategies, aiming for better reproductive pairings, maintaining genetic diversity, and avoiding inbreeding and outbreeding depression. In this study, we also investigated the presence of genetic structure for the Harpy Eagle in Brazil using single nucleotide polymorphisms (SNPs) and the mitogenome generated through low-coverage sequencing (Chapter II). Regarding the SNPs, all the pipelines used retrieved a weak signal of genetic structure for the Harpy Eagle in Brazil. The DiscoSnpRad pipeline demonstrated greater effectiveness in handling low-coverage sequencing, retrieving significant and consistent information with the results found through the mitogenome. The population located north of the Amazon River showed greater genetic differentiation, while the populations located south of the Amazon River and in the Atlantic Forest showed genetic similarities. Thus, the results obtained in the two chapters developed in this study provide molecular tools and information that assist in guiding management strategies and decision-making for the *in situ* and *ex situ* conservation of the Harpy Eagle.

Palavras-chave: Harpy Eagle. Genetic diversity. Conservation genetics. Genomics. Molecular markers.

LISTA DE TABELA

Capítulo I

Tabela 1. Caracterização de 10 microssatélites em 24 indivíduos de gavião-real (<i>Harpia harpyja</i>) mantidos em cativeiro no Brasil.....	38
--	----

Capítulo II

Tabela 1. Número de SNPs filtrados após a implementação do <i>hard</i> e <i>soft filters</i> e o número de indivíduos utilizados nas análises subsequentes por <i>pipeline</i>	69
---	----

Tabela 2. Valores do F_{ST} , F_{IS} e F_{IT} e seus respectivos valores de p e desvio padrão para cada <i>pipeline</i> utilizado no estudo.....	70
---	----

Tabela 3. Estimativas da diversidade genética para os grupos de gavião-real, determinadas pelo número de SNPs resgatados pelos diferentes <i>pipelines</i> , número de indivíduos, proporção de SNPs polimórficos e heterozigosidade observada e esperada para cada grupo.....	71
---	----

Tabela 4. Esquemas de particionamento determinados pelo PartitionFinder 2 utilizando o d-loop e 11 genes mitocondriais.....	72
--	----

LISTA DE FIGURA

Capítulo I

Figura 1. Representação esquemática da estratégia de amplificação e preparação da biblioteca para o sequenciamento de nova geração pela plataforma Ion Torrent.....34

Capítulo II

Figura 1. Distribuição geográfica das amostras de gavião-real.....62

Figura 2. Resultados da análise de componentes principais (esquerda) e da análise discriminante de componentes principais (direita) para os diferentes conjuntos de dados. A e B para GATK, C e D para *DiscoSnp++* e E e F para *DiscoSnpRad*.....69

Figura 3. Árvore filogenética para o gavião-real, evidenciando diferentes linhagens para a espécie, com o grupo do Norte da Amazônia (NAM) separado geneticamente dos grupos do Sul da Amazônia (SAM) e Mata Atlântica (ATF).....72

LISTA DE APÊNDICE

Capítulo I

- Apêndice 1.** Instituições doadoras das amostras de sangue dos gaviões-reais e as respectivas informações de cada indivíduo amostrado como a identificação individual, origem, nascimento e o sexo pelo histórico e pela sexagem molecular.....48
- Apêndice 2.** Combinação gradativa dos 10 loci em relação a probabilidade de identidade genética (I) e a probabilidade de exclusão de paternidade (Q).....49
- Apêndice 3.** Combinação gradativa dos quatro loci em relação a probabilidade de identidade genética (I) e a probabilidade de exclusão de paternidade (Q).....49
- Apêndice 4.** Matriz de coeficiente de parentesco (r) e suas respectivas relações com maior probabilidade entre pares dos 25 indivíduos de gavião-real mantidos em cativeiro amostrados neste estudo.....50
- Apêndice 5.** Lista com as estimativas de máxima verossimilhança do coeficiente de parentesco e do relacionamento com maior probabilidade entre pares dos 25 indivíduos de gavião-real mantidos em cativeiro, obtidas por meio do ML-Relate.....51
- Apêndice 6.** Teste de razão de verossimilhança realizado no ML-Related para os pares relacionados e com ambiguidade em seu conjunto de confiança.....56

Capítulo II

- Apêndice 1.** Local de origem e estado das amostras de gavião-real, grupo genético, idade e identificação do sexo.....82
- Apêndice 2.** Comandos utilizados para mapear as sequências individuais contra o genoma de referência do gavião-real (*Genome mapping*), para resgatar dos SNPs utilizando o GATK (*SNP calling*) e os *hard* e *soft filters* para realizar a filtragem dos bancos de dados.....83

SUMÁRIO

RESUMO.....	7
ABSTRACT.....	8
INTRODUÇÃO GERAL	14
1. A espécie estudada: o gavião-real	14
2. Genética da Conservação	15
3. Marcadores moleculares por sequenciamento de nova geração	16
4. Estudos moleculares para gavião-real	18
REFERÊNCIAS.....	22
OBJETIVOS.....	26
CAPÍTULO I.....	27
1. Introdução.....	28
2. Material e Métodos	30
2.1. Desenvolvimento de novos microssatélites	30
2.2. Amostragem, extração de DNA e identificação do sexo	31
2.3. Amplificação e preparação da biblioteca de sequenciamento	31
2.4. Genotipagem e análises estatísticas	32
3. Resultados.....	34
3.1. Obtenção dos novos microssatélites	34
3.2. Caracterização dos novos microssatélites.....	35
3.3. Amostragem e identificação do sexo	36
4. Discussão.....	38
5. Agradecimentos.....	41
6. Financiamento	41
7. Referências	42
8. Apêndices.....	47
CAPÍTULO II	57
1. Introdução.....	59
2. Material e Métodos	60
2.1. Amostragem biológica e extração do DNA	60
2.2. Preparação da biblioteca genômica e sequenciamento	61
2.3. Genoma de referência do gavião-real	62
2.4. Extração e pré-filtragem dos SNPs.....	62
2.5. Filtragem dos SNPs no R (<i>soft filters</i>).....	64
2.6. Análises de estrutura genética	64
2.7. Análise de diversidade genética	66
2.8. Extração do mitogenoma	66
2.9. Análises filogenéticas do mitogenoma.....	66
3. Resultados.....	67

3.1.	Sequenciamento e extração dos SNPs	67
3.2.	Estrutura populacional	68
3.3.	Diversidade genética	69
3.4.	Extração do mitogenoma	70
4.	Discussão	72
5.	Agradecimentos	75
6.	Financiamento	75
7.	Referência	75
8.	Apêndices	81
CONCLUSÃO GERAL		84

INTRODUÇÃO GERAL

1. A espécie estudada: o gavião-real

O gavião-real, também conhecido como harpia ou uiraçu, é uma ave de rapina chamada cientificamente como *Harpia harpyja* (Linnaeus, 1758), pertencente à ordem Accipitriformes, à família Accipitridae e é a única espécie do seu gênero. A espécie é considerada a maior águia das Américas e está entre as maiores espécies de águias existentes no mundo, podendo alcançar uma envergadura de até 2 metros (Sick, 1997). Sua ocorrência limita-se às Américas Central e Sul com uma distribuição original estabelecida desde o sul do México ao leste da Bolívia, leste do Brasil e extremo nordeste da Argentina (Vargas et al., 2006).

Essa espécie é encontrada em florestas tropicais e depende desses ambientes para sua sobrevivência, reprodução e forrageamento (Álvarez-Cordero, 1996). O gavião-real é classificado como um predador de topo, ocupando o nível mais alto na cadeia trófica, se alimentando majoritariamente de mamíferos de médio porte como, preguiças e macacos (Aguiar-Silva et al., 2014). Em relação ao comportamento reprodutivo, o casal de gaviões-reais apresenta um padrão monogâmico e constrói seus ninhos em árvores emergentes, que podem atingir mais de 30 metros de altura (Fowler e Cope, 1964; Luz, 2005; Miranda et al., 2020). O ciclo reprodutivo é longo, ocorrendo a cada dois a três anos e gera apenas um filhote por ninhada (Muñiz-López et al., 2007). Ao final de cada nidificação, o casal retorna à mesma árvore para iniciar um novo ciclo reprodutivo (Rettig, 1978).

O Brasil possui a maior área de ocupação do gavião-real, sendo encontrado nas florestas densas e úmidas, com maior frequência na Amazônia e logo em seguida, na Mata Atlântica (Banhos et al., 2018). No entanto, existem raros registros de nidificação e de deslocamento da espécie no corredor de vegetação aberta no centro do Brasil (Oliveira e Silva, 2006; Silva et al., 2013). Esses registros geralmente estão associados às bordas dos rios com presença de matas ciliares ou de manchas florestais que funcionam como uma conexão de floresta permitindo a dispersão do gavião-real entre a Mata Atlântica e Amazônia, e até mesmo a nidificação (Banhos, 2009).

Entretanto, as florestas que abrigam o gavião-real e suas conexões estão enfrentando uma redução cada vez mais significativa. No bioma Amazônico, aproximadamente 15% da floresta foi desmatada entre 2019 e 2021, com maiores perdas ocorrendo no sul e leste da bacia, região conhecida como arco do desmatamento (MapBiomas, 2022). Na Mata Atlântica, cerca de 88% da sua cobertura original foi perdida, restando apenas alguns remanescentes florestais isolados (Fundação SOS Mata Atlântica, 2021). Além disso, a região central do Brasil, que inclui os biomas Cerrado e Caatinga, também sofre com altas taxas de desmatamento e queimadas, perdendo grande parte de seu território para a agropecuária (MapBiomas, 2022).

Segundo a BirdLife International (2023) o gavião-real é classificado como ameaçado de extinção, situado na categoria Vulnerável. Entretanto, na Mata Atlântica, a situação da espécie é ainda mais preocupante, sendo classificada como Em Perigo ou Criticamente em Perigo dependendo da região encontrada (Alves et al., 2000; Bahia, 2017; Chaves et al., 2019). Essa

vulnerabilidade está associada principalmente à fragmentação e perda do habitat devido ao alto índice de desmatamento vinculado às áreas florestais do Brasil (Banhos et al., 2018). Somado a isso, o gavião-real também enfrenta ameaças como caça, perseguição, tráfico e criação ilegal (Trinca et al., 2008; Miranda et al., 2021). Dessa forma, todos esses fatores combinados representam um sério risco para a sobrevivência da espécie.

2. Genética da Conservação

A Genética da Conservação, uma área de estudo derivada da Biologia da Conservação, busca conservar, principalmente, espécies ameaçadas ou em vias de extinção através de análises genéticas populacionais e evolutivas (Loeschcke et al., 1994). Com esse propósito, a disciplina avalia um conjunto de informações para direcionar as estratégias de conservação, focando nas consequências que surgem da redução de uma população através dos fatores de ameaça que atestam a integridade e que impedem a adaptação da espécie às mudanças ambientais (Frankham et al., 2010).

Espécies que estão em situações de risco e que continuam enfrentando constantes ameaças provenientes das ações antrópicas, costumam ser objeto de estudos voltados à Genética da Conservação devido ao fato de que essas ameaças podem causar alterações genéticas a nível populacional ao longo do tempo e espaço, gerando consequências para a sobrevivência da espécie (Frankham et al., 2008; 2010). Exemplo disso é o desmatamento florestal, que destrói o habitat natural e reduz a conectividade entre as áreas florestais, em consequência, as populações são fragmentadas e o fluxo gênico se torna inadequado entre as populações (Frankham et al., 2017). Logo, uma população originalmente grande e exogâmica, se torna reduzida e com ausência ou pouca troca gênica.

Além disso, a fragmentação populacional e a ausência ou diminuição do fluxo gênico pode conduzir a população original ao isolamento genético, podendo ocasionar a diferenciação e a estruturação genética entre as diferentes populações (Hartl e Clark, 2010; Templeton, 2021). Outro problema que essas populações fragmentadas e com baixo fluxo gênico enfrentam, é com o aumento da endogamia, visto que uma população reduzida possui maior probabilidade de cruzamento entre parentais e com a perda da diversidade genética através da ação da deriva genética, uma vez que eventos ao acaso ocasionam a perda de alelos raros, o aumento de homozigotos e o acúmulo de alelos recessivos prejudiciais (Frankham et al., 2008).

Com intuito de mitigar tais consequências, a Genética da Conservação pode auxiliar e atuar em dois âmbitos: *in situ* e *ex situ*. A conservação *in situ* é realizada em habitat natural, onde é possível a espécie dar continuidade aos processos evolutivos e adaptações ao meio ambiente e suas mudanças (Frankham et al., 2008). Portanto, trabalhos que visam entender a história evolutiva, distribuição geográfica, demografia e a biodiversidade molecular da espécie são primordiais para estabelecer estratégias de conservação de forma correta e eficaz. Além disso, é crucial realizar esforços que assegurem a preservação e recuperação dos ecossistemas, uma vez que isso desempenha um papel fundamental na condução da conservação *in situ*.

No eixo da conservação *ex situ*, a preservação da espécie é realizada fora do ambiente natural e sob supervisão humana, como zoológicos e criadouros conservacionistas (CBD, 1993). Os animais que são mantidos em cativeiro geralmente são indivíduos retirados da natureza por ação antrópica ou incidental e que não são aptos para voltar à natureza (Efe et al., 2006). Esses animais possuem extrema importância para a conservação, pois são espécimes representativos das populações de vida livre que podem ser utilizados como um reservatório genético para reprodução em cativeiro, para conhecimento do comportamento da espécie e para a educação e sensibilização ambiental (Francisco e Silveira, 2013).

Apesar do importante potencial, a conservação *ex situ* possui fatores limitantes como, a falta de estrutura e espaço adequados nos recintos, o elevado custo para manutenção, o número reduzido de indivíduos aptos para a reprodução em cativeiro e a consequente perda de diversidade genética e tendência à depressão endogâmica, além da própria adaptação ao cativeiro, resultando em seleção artificial (Frankham et al., 2010). Portanto, diante de todos esses desafios, as instituições dedicadas à conservação precisam desenvolver estratégias de manejo que vão garantir a manutenção de populações viáveis e assegurar a representação da espécie em termos de variabilidade genética e demográfica (Francisco e Silveira, 2013).

Dado que o gavião-real está em estado de vulnerabilidade e ainda enfrenta a caça e a perda constante do seu habitat, monitorar e atuar de acordo com a sua necessidade é importante para proteger a espécie de possíveis declínios locais, da redução do potencial evolutivo e do elevado risco de extinção. O Projeto Harpia, um projeto nacional de conservação das águias florestais (<https://www.projetohearpia.org/>), vem atuando há mais de vinte anos em território brasileiro na conservação da espécie em seu hábitat natural, executando ações de pesquisa, monitoramento e manejo para a conservação do gavião-real na Amazônia, na Mata Atlântica e ecótonos associados.

O projeto também atua, juntamente com instituições parceiras, na conservação *ex situ* da espécie. Atualmente, as atividades *ex situ* envolvem principalmente o resgate, a reabilitação e, se forem reabilitados com sucesso, na soltura de indivíduos que são removidos da natureza (Efe et al., 2006; Oliveira et al., 2022). Essas atividades têm como objetivo garantir o retorno dos indivíduos reabilitados à natureza de forma segura para que possam continuar cumprindo seu papel ecológico em vida livre ou, se não reabilitados, mantê-los seguros em um ambiente controlado e receber os cuidados necessários para sua sobrevivência (Frankham et al., 2008; Francisco e Silveira 2013). Outra atuação envolve a reprodução em cativeiro para que os indivíduos sejam utilizados em ações de conservação no futuro caso seja necessário, com intuito de reforçar as populações naturais já existentes ou fundar novas populações e introduzi-las à natureza (Itaipu Binacional, 2020; Oliveira et al., 2022).

3. Marcadores moleculares por sequenciamento de nova geração

As técnicas para estudos com marcadores moleculares têm sido aprimoradas com os avanços tecnológicos e com o desenvolvimento de novas *pipelines* e metodologias que permitem a identificação de marcadores em escala genômica com alto desempenho e acurácia. Todo esse avanço foi possibilitado com o advento das diferentes plataformas de

sequenciamento de nova geração (NGS), proporcionando um sequenciamento de DNA em grande escala, automatizado, paralelo e de alto rendimento (Slatko et al., 2018). Com essa tecnologia, é possível utilizar inúmeras amostras para sequenciar e genotipar milhares de regiões e genomas de interesse em um único passo e, conseqüentemente, oferecer maior poder em detectar variantes novas ou raras.

Atualmente a Illumina é a plataforma mais utilizada de sequenciamento de DNA para estudos clínicos e pesquisa, por causa da diversidade de ferramentas e protocolos desenvolvidos para utilizar os dados gerados por NGS (Kulski, 2016). Em segundo lugar está a plataforma Ion Torrent, escolhida muitas vezes por conta da velocidade na geração dos dados, porém apresenta uma taxa de transferência de dados limitada e exige um controle de qualidade mais rigoroso e criterioso para evitar artefatos em regiões homopoliméricas (Liu et al., 2012). Vários são os marcadores genéticos utilizados para estudos que abordam a Genética da Conservação, dentre eles estão os microssatélites, os polimorfismos de nucleotídeo único e o mitogenoma.

Os microssatélites ou *Simple Sequence Repeats* (SSR) são repetições em séries de 1 a 6 nucleotídeos com altas taxas mutacionais (Schlötterer, 2000). Devido às suas características é considerado um marcador altamente polimórfico e codominante, fornecendo um grande número de informações que podem ser utilizadas para diversas aplicações, como a avaliação da diversidade genética das populações naturais e cativas, estrutura populacional e análises forenses (Turchetto-Zolet et al., 2017). A implementação das plataformas de sequenciamento de alto rendimento simplificou o processo de isolamento de microssatélites e da genotipagem das amostras em espécies não modelo, tornando a aquisição desse marcador molecular mais rápida, eficaz e menos custosa.

Além dos SSRs serem multialélicos, esses marcadores são acessíveis para compra, quando já desenvolvidos, e de fácil reprodutibilidade. Em vista disso, com o objetivo de direcionar os esforços para um manejo genético eficaz e assertivo, os zoológicos e criadouros conservacionistas podem fazer uso dos SSRs para avaliar a diversidade genética da população cativa e auxiliar nos programas de reprodução em cativeiro (Witzenberger e Hochkirch, 2011; 2013). Essa estratégia visa realizar identificação individual e exclusão parental dos indivíduos, testes de atribuição e criação de matrizes de parentesco para orientar os melhores pares de acasalamentos, evitando o cruzamento de indivíduos aparentados e preservando a diversidade e estrutura genética existente na espécie (Turchetto-Zolet et al., 2017).

Os polimorfismos de nucleotídeo único ou *Single Nucleotide Polymorphisms* (SNPs) são variações genéticas que ocorrem quando um único nucleotídeo é substituído por outro no genoma, podendo ser originados por mutações pontuais no DNA, como as transições e transversões (Allendorf et al., 2013). Os SNPs são considerados marcadores bialélicos e codominantes, possuem taxas de mutação relativamente baixas e são altamente abundantes no genoma, ocorrendo com maior frequência em regiões não codificantes (Turchetto-Zolet et al., 2017). Com os avanços na área da genômica e, principalmente, da bioinformática, surgiram métodos automatizados que permitem a identificação e genotipagem em grande escala de

marcadores SNPs para espécies que não são modelos, com custo reduzido e menor necessidade de mão de obra (Kulsik, 2016).

Embora os SNPs sejam considerados menos informativos quando comparados com os microssatélites por serem bialélicos, a sua presença é ampla no genoma, permitindo a identificação de milhares de variantes por amostra. Isso significa que o uso dos SNPs fornece uma cobertura genômica mais abrangente, oferecendo dados com mais qualidade, com baixa taxa de erro na genotipagem e alta capacidade de detectar pequenas variações genéticas. Desse modo, os SNPs são empregados com sucesso em diversos estudos para investigar e aprimorar a compreensão da variação genética, estrutura populacional, aplicação em taxonomia integrativa, análise de seleção natural e evolução adaptativa, bem como para testes de paternidade e rastreabilidade (Turchetto-Zolet et al., 2017; Young e Gillung, 2019).

O mitogenoma refere-se ao genoma mitocondrial inteiro do indivíduo, incluindo todos os genes e a região controle. Quando as amostras são sequenciadas utilizando as tecnologias de NGS é possível extrair e realizar a montagem do mitogenoma utilizando as sequências referentes ao DNA mitocondrial (DNAMt). Esse marcador molecular é muito utilizado em conjunto com marcadores nucleares, como SSRs ou SNPs, com objetivo de refinar e obter resultados complementares em estudos que visam a reconstrução da história evolutiva de uma determinada espécie. Isto se deve ao diferente modo de herança desses genomas, no caso do mitogenoma a herança é matrilinear e a taxa de mutação é mais alta quando comparada com genes nucleares que possuem herança biparental (Templeton, 2021). Com isso, os padrões encontrados com DNAMt podem ser mais informativos sobre eventos e processos históricos, enquanto os resultados obtidos com o genoma nuclear podem refletir padrões mais contemporâneos (Avice, 2009).

Como foi visto, o desenvolvimento tecnológico no campo da NGS têm facilitado a aquisição de novos marcadores moleculares e a obtenção de dados genômicos em larga escala. Nesse estudo, foram utilizados em conjunto os três marcadores (SSRs, SNPs e mitogenoma) descritos acima para alcançar uma resolução mais precisa e elucidar perguntas previamente não respondidas, bem como identificar *loci* mais variáveis para estudos que abordam a genética da conservação para o gavião-real. Portanto, utilizando essa abordagem mais abrangente possibilita uma análise mais robusta da diversidade genética, estrutura populacional e história evolutiva da espécie e, com isso, auxilia a direcionar de forma mais eficiente os esforços de conservação tanto *in situ* quanto *ex situ*.

4. Estudos moleculares para gavião-real

Estudar a biologia e ecologia do gavião-real, assim como monitorar a espécie na natureza, apresenta grandes desafios. Isso se deve principalmente a baixa densidade populacional da espécie, a sua capacidade de se deslocar por longas distâncias e por habitar no dossel das florestas tropicais densas, o que dificulta a detecção dos indivíduos e de árvores-ninhos (Gonzalez e Vargas, 2011; Banhos et al., 2018). Outro fator limitante é a captura e marcação dos indivíduos quando avistados, impedindo realizar identificações individuais. Além disso, pela ausência de dimorfismo sexual na plumagem, os machos e as fêmeas da espécie

são difíceis de distinguir, sendo sexualmente distintos pelo tamanho corporal, com fêmeas maiores que os machos, e através da sexagem molecular (Banhos et al., 2008; Banhos, 2009).

Uma forma de minimizar as dificuldades para estudar o gavião-real é coletar amostras em momentos oportunos, por exemplo, quando é solicitado o resgate de algum indivíduo, dos animais caçados e até mesmo de espécimes que estão mantidos em cativeiro ou depositados em coleções, e usá-las para estudos genéticos através dos marcadores moleculares. Poucos foram os estudos para gavião-real utilizando marcadores moleculares, o primeiro deles foi a obtenção, caracterização e otimização de marcadores moleculares para auxiliar no manejo e conservação da espécie (Banhos et al., 2008). Os autores selecionaram *loci* microssatélites heterólogos e aprimoraram um novo conjunto de marcadores sexuais para a espécie.

Nesse mesmo estudo, os 24 *loci* de microssatélites selecionados eram polimórficos apresentando uma diversidade genética baixa, mas demonstram alta robustez quanto aos índices de probabilidade conjunta de identificação individual e de exclusão de parentesco (Banhos et al., 2008). Os marcadores sexuais foram importantes para confirmar o sexo dos indivíduos visto que há dificuldades de diferenciar machos e fêmeas visualmente. Ambos os marcadores se tornaram ferramentas úteis para estudos genéticos com objetivo de auxiliar na conservação e manejo *in situ* e *ex situ* do gavião-real.

O segundo estudo foi realizado com sequências da região controle do DNAm para avaliar a diversidade genética, estrutura populacional e história demográfica do gavião-real da América Central e do Sul (Lerner et al., 2009). Como resultado, os autores encontraram diferenciação genética significativa entre a população da América Central e do Sul e um fluxo gênico restrito e unidirecional da América Central para a América do Sul. Entretanto, no estudo foram utilizadas somente três amostras do Brasil, não representando bem a distribuição da espécie no país e criando um viés amostral. Além disso, apesar dos recentes declínios populacionais evidenciados na região, foram encontrados altos níveis de diversidade haplotípica ($h = 0,906 \pm 0,020$), realçando a importância da conservação da população local (Lerner et al., 2009).

O terceiro trabalho de genética molecular para o gavião-real foi realizado somente em território brasileiro, onde a sua distribuição é mais ampla, com objetivo de esclarecer sobre a diversidade genética, estrutura populacional e demografia da espécie no país (Banhos, 2009). No estudo foram utilizados dois tipos de marcadores moleculares, as sequências da região controle do DNAm e 9 *loci* microssatélites heterólogos disponibilizados em Banhos et al. (2008). Como os dois marcadores utilizados apresentam formas de herança distintas, foi possível observar e comparar os resultados encontrados.

Para o DNAm os valores para diversidade genética foram relativamente mais baixos ($h = 0,688 \pm 0,039$), houve diferenciação genética significativa entre a Amazônia e Mata Atlântica, de forma que quanto maior a distância geográfica, maior a diferenciação entre as populações. O autor também encontrou sinais de redução populacional no passado e eventos migratórios unidirecional da Mata Atlântica para a Amazônia pelo leste do Brasil (Banhos, 2009). Além

disso, utilizando os dados do DNAm_t obtidos por Lerner et al. (2009) em conjunto, verificou o oposto, detectando um forte sinal de migração unidirecional da América do Sul para a Central.

Em relação aos microssatélites, a diversidade genética obteve um valor mais baixo ($H_e = 0.477$; $H_o = 0.456$) quando comparado com o DNAm_t, mas foi similar a outras espécies de águias ameaçadas de extinção. Os SSRs não encontraram diferenciação genética entre as populações e nem redução populacional, mas detectaram uma dinâmica populacional bidirecional entre a Amazônia e Mata Atlântica, com a Amazônia recebendo mais migrantes (Banhos, 2009). Por fim, o autor sugeriu uma dispersão dos indivíduos diferenciada entre os sexos, onde as fêmeas são mais residentes e os machos se dispersam mais, o que indica a existência de um sistema de reprodução poligínico, diferentemente do que foi observado na sua ecologia (Fowler e Cope, 1964).

Até o momento, os microssatélites utilizados para investigar a diversidade genética do gavião-real foram os heterólogos. O uso desse tipo de marcador é vantajoso pelo fato de ser menos custoso quando comparado com o isolamento de microssatélites para a espécie específica. Porém, há desvantagens na transferibilidade desses microssatélites, como subestimar a diversidade genética por causa de possíveis mutações na região de pareamento do *primer*, podendo conter alelos nulos e gerar falsos homozigotos. Com intuito de solucionar esse problema, Coser et al. (2014) desenvolveram 10 *loci* microssatélites polimórficos específicos para o gavião-real, entretanto, resgataram pouco polimorfismo por *locus*, inviabilizando análises populacionais, de parentesco e de identificação individual.

Em Banhos et al. (2016), uma extensão de sua tese (Banhos, 2009), oito *loci* microssatélites heterólogos disponibilizados em Banhos et al. (2008) foram utilizados em amostras históricas e contemporâneas para avaliar os efeitos do desmatamento da floresta tropical brasileira nos níveis e na distribuição da diversidade genética do gavião-real. Os resultados mostraram que a diversidade genética da espécie diminuiu nas regiões onde o desmatamento é intenso, principalmente no Sul da Amazônia e na Mata Atlântica. Em contrapartida, as amostras situadas ao norte do Rio Amazonas, região pouco alterada pela pressão do desmatamento, apresentaram uma diversidade genética semelhante às amostras históricas de todo território brasileiro (Banhos et al., 2016).

No mesmo estudo, a análise de estrutura genética utilizando uma abordagem para descoberta de grupos genéticos por meio dos microssatélites heterólogos, não encontrou estrutura entre as populações (Banhos et al., 2016). Porém, através do F_{st} , foi encontrada uma baixa diferenciação genética entre a população do Norte da Amazônia e as populações do Sul da Amazônia e Mata Atlântica e uma diferenciação não significativa entre elas, indicando que o Sul da Amazônia e a Mata Atlântica abrigam populações de gavião-real geneticamente semelhantes (Banhos et al., 2016). A diferença genética observada na população do Norte da Amazônia pode ser atribuída a diferentes fatores, como o fluxo gênico limitado devido à distância geográfica entre as populações, menor necessidade de deslocamento em busca de recursos devido à região sofrer menos com o desmatamento ou o Rio Amazonas atuando como uma barreira genética (Banhos et al., 2016).

Estudos também têm sido realizados no âmbito *ex situ*, envolvendo principalmente esforços para preservar a diversidade genética existente da espécie. Para isso, Oliveira et al. (2022) compilou informações sobre a população do gavião-real mantido em cativeiro no Brasil, incluindo o número de indivíduos mantidos em cativeiros, o sexo, a identificação individual, o tempo de cativeiro e sua origem, quando conhecida. Entretanto, ainda não foram compilados dados que revelem as características genéticas da população cativa, limitando a formação de pares reprodutivos através da variabilidade genética e o nível de parentesco entre os indivíduos.

Desde 1999 a reprodução em cativeiro do gavião-real no Brasil é realizada pela Sociedade de Pesquisa de Fauna Silvestre (CRAX) (Azeredo, 2005). Atualmente, o Refúgio Biológico Bela Vista (RBV) da Itaipu Binacional é a instituição que mais reproduz e mantém gaviões-reais em cativeiro no mundo. De fato, a reprodução do gavião-real em cativeiro é um sucesso, porém os casais são formados sem nenhum levantamento genético dos indivíduos, apenas com base na compatibilidade comportamental e em idades e origem semelhantes, quando conhecidas (Watson et al., 2016; Oliveira et al., 2022). Dessa forma, a falta de estudos genéticos da população *ex situ* pode comprometer todo o esforço realizado da reprodução em cativeiro gerando indivíduos com baixa diversidade genética e endogâmicos.

Diante dos estudos genéticos realizados até o momento, o gavião-real é uma espécie de grande preocupação para a conservação e sem dúvidas muitos esforços têm sido direcionados para a conservação *in situ* e *ex situ* do gavião-real, no entanto, ainda é preciso realizar novos estudos que esclareçam a composição genética das populações de gavião-real no Brasil. Visto que os avanços do sequenciamento NGS e da bioinformática têm proporcionado grandes melhorias na obtenção de novos marcadores moleculares e de bancos de dados com maior número de variações e melhor qualidade, abre caminho para responder a questões não esclarecidas anteriormente e resolver incertezas.

Por fim, esse trabalho foi desenvolvido com objetivo de gerar ferramentas moleculares mais eficazes para definir unidades de manejo necessárias para minimizar a perda de diversidade genética e, ao mesmo tempo, elucidar a estrutura genética existente do gavião-real no Brasil para preservá-la. O primeiro capítulo tem como foco desenvolver e fornecer um novo conjunto de marcadores moleculares do tipo microssatélites específicos para o gavião-real com intuito de auxiliar as instituições conservacionistas no manejo genético da população *ex situ*. O segundo capítulo aborda a diversidade e estrutura genética do gavião-real no Brasil por meio dos SNPs e do mitogenoma completo, marcadores ainda não utilizados para a espécie e que possuem um grande potencial para investigar e elucidar sinais de estrutura genética. Desse modo, ambos os trabalhos poderão ser aplicados na conservação da espécie no âmbito *in situ* e *ex situ*.

REFERÊNCIAS

- Aguiar-Silva FH, Sanaiotti TM, Luz BB (2014) Food habits of the Harpy Eagle, a top predator from the Amazonian rainforest canopy. *J of Raptor Res* 48:24-35. <https://doi.org/10.3356/JRR-13-00017.1>
- Allendorf FW, Luikart GH, Aitken SN (2013) Conservation and the genetics of populations. John Wiley & Sons, New York
- Álvarez-Cordero E (1996) Biology and conservation of the Harpy Eagle in Venezuela and Panamá. Dissertation, University of Florida
- Alves MAS, Pacheco JF, Gonzaga LAP, Cavalcanti RB, Raposo MA, Yamashita C, Maciel NC, Castanheira M (2000) Aves. In: Bergallo HG, Rocha CFD, Sluys MV, Alves MAS (eds) A fauna ameaçada de extinção do Estado do Rio de Janeiro. Universidade Estadual do Rio de Janeiro, Rio de Janeiro, pp 113-124
- Awise JC (2009) Phylogeography: retrospect and prospect. *J of Biogeogr* 36:3-15. <https://doi.org/10.1111/j.1365-2699.2008.02032.x>
- Azeredo R (2005) Reprodução de *Harpia harpyja* em cativeiro. *Rapina* 1:6-13
- Bahia (2017) Lista Oficial das Espécies da Fauna Ameaçadas de Extinção do Estado da Bahia. Secretaria do Meio Ambiente, Bahia
- Banhos A (2009) Genética, Distribuição e Conservação do Gavião-real (*Harpia harpyja*) no Brasil. Tese de Doutorado, Instituto Nacional de Pesquisas da Amazônia
- Banhos A, Hrbek T, Gravena W, Sanaiotti TM, Farias IP (2008) Genomic resources for the conservation and management of the harpy eagle (*Harpia harpyja*, Falconiformes, Accipitridae). *Genet and Mol Biol* 31:146-154. <https://doi.org/10.1590/S1415-475720080001000255>
- Banhos A, Hrbek T, Sanaiotti TM, Farias IP (2016) Reduction of genetic diversity of the Harpy Eagle in Brazilian tropical forests. *PloS One* 11:1-12. <https://doi.org/10.1371/journal.pone.0148902>
- Banhos A, Sanaiotti TM, Aguiar-Silva FH, Martins FD, Luz BB, Carvalho AS, Ruiz CM (2018) *Harpia harpyja* (Linnaeus, 1758). Livro Vermelho da Fauna Ameaçada de Extinção 3:124-128
- BirdLife International (2023) Species factsheet: *Harpia harpyja*. (<http://datazone.birdlife.org/species/factsheet/harpy-eagle-harpia-harpyja>). Acesso em: 10 fev 2023
- CBD (1992) Convention on biological diversity. (<https://www.cbd.int/doc/legal/cbd-en.pdf>). Acesso em: 15 jan 2023

Chaves FG, Duca C, Pinto GO, Rosa GAB, Magnago GR, Daros HJ, Passamani J, Silva JN, Silva JP, Bissoli LB, Mathias LB, Alves MAS, Ribon R, Betzel RL, Vanstreels RET, Hurtado R (2019) Aves ameaçadas de extinção no estado do Espírito Santo. In: Fraga CN, Formigoni MH, Chaves FG (eds) Fauna e flora ameaçadas de extinção no estado do Espírito Santo, Instituto Nacional da Mata Atlântica, Santa Teresa, pp 294-313

Coser R, Martínez JG, Nunes M, Assunção EN, Astolfi-Filho S, Banhos A, Sanaiotti TM, Hrbek T, Farias IP (2014) Development of microsatellite markers for the near threatened eagles *Harpia harpyja* and *Morphnus guianensis* using next-generation sequencing. Conserv Genet Resour 6:999-1001. <https://doi.org/10.1007/s12686-014-0268-7>

Efe MA, Martins-Ferreira C, Olmos F, Mohr LV, Silveira LF (2006) Diretrizes da Sociedade Brasileira de Ornitologia para a destinação de aves silvestres provenientes do tráfico e cativeiro. Rev Bras de Ornitol 14:67-72

Fowler JM, Cope JB (1964) Notes on the harpy eagle in British Guiana. The Auk 81: 257-273. <https://doi.org/4082683>

Francisco MR, Silveira LF (2013) Conservação Animal *ex situ*. In: Piratelli AJ, Francisco MR (eds) Conservação da biodiversidade dos conceitos às ações. Technical Books, Rio de Janeiro, pp 117-130

Frankham R, Ballou JD, Briscoe DA (2008) Fundamentos de genética da conservação. Sociedade Brasileira de Genética, Ribeirão Preto

Frankham R, Ballou JD, Ralls K, Eldridge M, Dudash MR, Fenster CB, Lacy RC, Sunnucks P (2017) Genetic management of fragmented animal and plant populations. Oxford University Press, Oxford

Frankham R, Ballou JD, Briscoe DA (2010) Introduction to Conservation Genetics. Cambridge University Press, Cambridge

Fundação SOS Mata Atlântica (2021) Relatório Anual 2021. (www.sosma.org.br/sobre/relatorios-e-balancos). Acesso em: 10 jun 2022

Gonzalez JDJV, Vargas FH (2011) Nesting density of Harpy Eagles in Darien with population size estimates for Panama. J of Raptor Res 45:199-210

Hartl DL, Clark AG (2010) Princípios de Genética de Populações. Artmed Editora, Porto Alegre

Itaipu Binacional (2020) Nasce o 50º filhote de harpia no Refúgio Biológico da Usina de Itaipu. (<https://www.itaipu.gov.br/sala-de-imprensa/noticia/nasce-o-50%C2%BA-filhote-de-harpia-no-refugio-biologico-da-usina-de-itaipu>). Acesso em: 02 fev 2022

Kulski JK (2016) Next-generation sequencing—an overview of the history, tools, and “Omic” applications. In: Kulski JK (ed) Next generation sequencing-advances, applications and challenges, IntechOpen, London, 10.5772/61964

Lerner HR, Johnson JA, Lindsay AR, Kiff LF, Mindell DP (2009) It's not too late for the harpy eagle (*Harpia harpyja*): high levels of genetic diversity and differentiation can fuel conservation programs. PLoS One 4:e7336. <https://doi.org/10.1371/journal.pone.0007336>

Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M (2012) Comparison of next-generation sequencing systems. J of Biomed and Biotechnol 2012:2314-6133. <https://doi.org/10.1155/2012/251364>

Loeschcke V, Tomiuk J, Jain SK (1994) Introductory remarks: Genetics and conservation biology. In: Loeschcke V, Jain SK, Tomiuk J (eds) Conservation Genetics, Birkhäuser, Basel, pp 3-8. https://doi.org/10.1007/978-3-0348-8510-2_1

Luz BB (2005) Características de árvores emergentes utilizadas por gavião-real (*Harpia harpyja*) para nidificação no centro e leste da Amazônia Brasileira. Dissertação, Instituto de Nacional de Pesquisas da Amazônia

MapBiomas (2022) Relatório Anual de Desmatamento 2021. (https://s3.amazonaws.com/alerta.mapbiomas.org/rad2021/RAD2021_Completo_FINAL_Rev1.pdf). Acesso em: 22 nov 2022

Miranda EB, Peres CA, Marini MÂ, Downs CT (2020) Harpy Eagle (*Harpia harpyja*) nest tree selection: Selective logging in Amazon forest threatens Earth's largest eagle. Biol Conserv 250:108754. <https://doi.org/10.1016/j.biocon.2020.108754>

Miranda DE, Peres CA, Downs CT (2022) Landowner perceptions of livestock predation: implications for persecution of an Amazonian apex predator. Anim Conserv 25:110-124. <https://doi.org/10.1111/acv.12727>

Muñiz-López R, Criollo O, Mendúa A (2007) Results of five years of the “Harpy Eagle (*Harpia harpyja*) Research Program” in the Ecuadorian tropical forest. In: Bildstein K, Barber D, Zimmerman A (eds) Neotropical raptors, Hawk Mountain Sanctuary, Orwigsburg, pp 22-32

Oliveira AL, Silva RS (2006) Registro de Harpia (*Harpia harpyja*) no Cerrado de Tapira, Minas Gerais, Brasil. Rev Bras de Ornitol 14:433-434

Oliveira MJ, Aguiar-Silva FH, Moraes W, Sanaiotti TM, Banhos A, Moreira N (2022) *Ex situ* population of the Harpy Eagle and its potential for integrated conservation. ZooKeys 1083:109-128. [10.3897/zookeys.1083.69047](https://doi.org/10.3897/zookeys.1083.69047)

Rettig NL (1978) Breeding behavior of the Harpy Eagle (*Harpia harpyja*). The Auk 95:629-643.

Schlötterer C (2000) Evolutionary dynamics of microsatellite DNA. *Chromosoma* 109:365-371. 10.1007/s004120000089

Sick H (1997) *Ornitologia Brasileira*. Nova Fronteira, Rio de Janeiro

Silva DA, Melo FRD, Guimarães IG (2013) Historical and recent records of the Harpy Eagle (*Harpia harpyja*) in the Cerrado biome of the state of Goiás, Brazil. *Rev Bras de Ornitol* 21:260-263.

Slatko BE, Gardner AF, Ausubel FM (2018) Overview of next-generation sequencing technologies. *Curr Protoc in Mol Biol* 122:e59. <https://doi.org/10.1002/cpmb.59>

Templeton AR (2021) *Population Genetics and Microevolutionary Theory*. John Wiley & Sons, Saint Louis

Trinca CT, Ferrari SF, Lees AC (2008) Curiosity killed the bird: arbitrary hunting of Harpy Eagles *Harpia harpyja* on an agricultural frontier in southern Brazilian Amazonia. *Cotinga* 30:12-15

Turchetto-Zolet AC, Turchetto C, Zanella CM, Passaia G (2017) Marcadores moleculares na era genômica: Metodologias e aplicações. Sociedade Brasileira de Genética, Ribeirão Preto

Vargas GJJ, Whitacre D, Mosquera R, Albuquerque J, Piana R, Thiollay JM, Márquez C, Sánchez JE, López ML, Matola S, Aguilar S, Rettig N, Sanaiotti TM (2006) Estado y distribución actual del Águila Arpia (*Harpia harpyja*) en Centro y Sur América. *Ornitol Neotrop* 17:39-55. <https://repositorio.inpa.gov.br/handle/1/22603>

Watson RT, McClure CJ, Vargas FH, Jenny JP (2016) Trial restoration of the Harpy eagle, a large, long-lived, Tropical Forest Raptor, in Panama and Belize. *J of Raptor Res* 50:3-22. <https://doi.org/10.3356/rapt-50-01-3-22.1>

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

Witzenberger KA, Hochkirch A (2013) Evaluating *ex situ* conservation projects: Genetic structure of the captive population of the Arabian sand cat. *Mamm Biol* 78:379-382 <https://doi.org/10.1016/j.mambio.2013.03.001>

Young AD, Gillung JP (2020) Phylogenomics-principles, opportunities and pitfalls of big-data phylogenetics. *Syst Entomol* 45:225-247. <https://doi.org/10.1111/syen.12406>

OBJETIVOS

Capítulo I

Objetivo geral

Desenvolver novos microssatélites específicos para o gavião-real, contendo elevado polimorfismo para auxiliar principalmente no manejo da população *ex situ* da espécie.

Objetivos específicos

- Desenvolver bibliotecas genômicas e selecionar microssatélites altamente polimórficos para o gavião-real;
- Avaliar a eficiência dos novos microssatélites quanto ao poder de exclusão de parentesco e de identificação individual;
- Avaliar a diversidade genética da população do gavião-real *ex situ*;
- Estimar as relações de parentesco entre os indivíduos dos gaviões-reais cativos;
- Realizar a sexagem molecular dos indivíduos cativos;
- Gerar ferramentas genéticas úteis para que estas possam ser utilizadas nas recomendações de manejo da espécie.

Capítulo II

Objetivo geral

Investigar a existência de estrutura genética para o gavião-real no Brasil, a partir de um sequenciamento de baixa cobertura.

Objetivos específicos

- Utilizar diferentes *pipelines* para identificar e extrair variantes a partir de um conjunto de dados gerado por sequenciamento de baixa cobertura;
- Selecionar a melhor estratégia para obter as variantes de acordo com o banco de dados do estudo;
- Realizar análises genéticas para descobrir e testar hipótese para a presença de estrutura genética para o gavião-real no Brasil utilizando SNPs e o mitogenoma;
- Gerar informações úteis para direcionar as estratégias de manejo para o gavião-real.

CAPÍTULO I

Desenvolvimento de novos marcadores microssatélites para manejo *ex situ* do gavião-real

Desenvolvimento de novos marcadores microssatélites para manejo *ex situ* do gavião-real

Mylena Kaizer^{1,2,3}, Pedro S. Bittencourt¹, Érico M. Polo¹, Tânia M. Sanaiotti^{2,4}, Izeni P. Farias^{1,2,3}, Tomas Hrbek^{1,2,3}, Aureo Banhos^{1,2,5}

¹ Laboratório de Evolução e Genética Animal - LEGAL, Departamento de Genética, Universidade Federal do Amazonas, Av. General Rodrigo Otávio Jordão Ramos, 3000, Coroado, 69077-000, Manaus, Amazonas, Brasil.

² Projeto Harpia, Instituto Nacional de Pesquisas da Amazônia - INPA, Av. André Araújo, 2936, Aleixo, 69067-375, Manaus, Amazonas, Brasil.

³ Programa de Pós-Graduação em Zoologia - PPGZOO, Universidade Federal do Amazonas, Av. General Rodrigo Otávio Jordão Ramos, 3000, Coroado, 69077-000, Manaus, Amazonas, Brasil.

⁴ Coordenação de Biodiversidade, Instituto Nacional de Pesquisas da Amazônia - INPA, Avenida André Araújo, 2936, Aleixo, 69067-375, Manaus, Amazonas, Brasil.

⁵ Departamento de Biologia, Centro de Ciências Exatas, Naturais e da Saúde - CCENS, Universidade Federal do Espírito Santo, Alto Universitário, Guararema, 29500-000, Alegre, Espírito Santo, Brasil.

Resumo

O gavião-real (*Harpia harpyja*), uma das maiores aves de rapina do mundo, está ameaçado de extinção, sendo classificado como Vulnerável em toda sua distribuição. Para garantir a sua conservação, além dos esforços *in situ*, é fundamental realizar avaliações genéticas da população *ex situ*, a fim de desenvolver estratégias de manejo que evitem o aumento da endogamia e a perda de diversidade genética ao longo do tempo. Neste estudo, foi utilizada a plataforma de sequenciamento de nova geração Ion Torrent PGM para desenvolver novos *primers* para regiões microssatélites específicas para o gavião-real e realizar a genotipagem dos indivíduos de forma simultânea. Estes marcadores foram validados em indivíduos de gavião-real mantidos em cativeiro por zoológicos e instituições conservacionistas brasileiras. Foram selecionados um conjunto de dez *loci* microssatélites com elevado polimorfismo e alto poder de exclusão de parentesco e de identificação individual para auxiliar no manejo da população *ex situ* da espécie. Os *loci* microssatélites desenvolvidos não apresentaram alelos nulos, não houve desequilíbrio de ligação e apenas quatro *loci* desviaram da expectativa do equilíbrio de Hardy-Weinberg. A diversidade genética encontrada para a população *ex situ* do gavião-real amostrada neste estudo foi baixa. Os microssatélites demonstraram eficiência em excluir os prováveis pais de um indivíduo, realizar a identificação individual e detectar os níveis de parentesco entre os indivíduos. Diante disso, este estudo fornece um novo conjunto de marcadores microssatélites eficientes para auxiliar no direcionamento de estratégias de manejo *ex situ*, visando melhores pareamentos reprodutivos, evitando a endogamia e mantendo a diversidade genética.

Palavras-Chave: Accipitridae. *Harpia Harpyja*. Conservação *ex situ*. Reprodução em cativeiro.

1. Introdução

Harpia harpyja (Linnaeus, 1758), popularmente chamada de gavião-real ou harpia, pertence à família Accipitridae e é considerada uma das maiores águias viventes do mundo. Sua distribuição geográfica compreende florestas primárias densas e de galeria das Américas Central e Sul, ocorrendo originalmente desde o sul do México ao leste da Bolívia, leste do Brasil e extremo nordeste da Argentina (Vargas et al., 2006). Ao longo da história, a espécie vem sofrendo com a redução populacional em grande parte da sua distribuição, devido à perda

de habitat e à caça (Trinca, 2008; Banhos et al., 2018). Estimativas apontam que a taxa de declínio populacional do gavião-real pode chegar a cerca de 50% nas próximas três gerações (BirdLife International, 2023). Por esses motivos, a espécie é considerada ameaçada de extinção em toda a sua distribuição, situada na categoria Vulnerável (BirdLife International, 2023).

Ações para a conservação *in situ* e *ex situ* têm sido desempenhadas para reduzir as ameaças de extinção do gavião-real. Enquanto as ações *in situ* focam na proteção dos ninhos da espécie e de seus territórios (Aguilar-Silva et al., 2015; Sanaiotti et al., 2015), as ações *ex situ* focam principalmente no resgate, reabilitação e destinação dos indivíduos feridos (Oliveira et al., 2022). Além disso, os zoológicos e criadouros dedicados à conservação do gavião-real possuem um acervo de indivíduos com grande potencial para servir como uma população reprodutiva de segurança. Isso possibilita futuras intervenções de manejo, incluindo translocações entre as instituições e a reprodução em cativeiro para reforçar as populações naturais já existentes ou, caso necessário, fundar novas populações que possam ser introduzidas na natureza (Soares et al., 2008; Oliveira et al., 2022).

A reprodução do gavião-real em cativeiro ocorre desde 1988, porém a seleção dos pares reprodutivos é realizada baseada na compatibilidade comportamental e em informações contida no histórico dos indivíduos, como idade, origem e sexo dos indivíduos (Watson et al., 2016; Oliveira et al., 2022). Entretanto, informações errôneas ou ausentes sobre o histórico dos indivíduos são frequentes e influenciam negativamente o manejo em cativeiro, como a subestimação do nível de parentesco entre os indivíduos reprodutores (Bowling et al., 2003; Oliveira et al., 2022). Desse modo, os dados moleculares podem auxiliar no manejo, avaliando a representatividade da população em cativeiro, detectando erros nas informações do histórico individual e, mais importante, desvendando o nível de parentesco entre os indivíduos cativos.

Os microssatélites (SSRs) podem auxiliar na obtenção de informações genéticas, visto que suas características permitem identificar parentes, a determinar da identidade de indivíduos e o estabelecer relações de parentesco. Devido a essa utilidade, muitos programas de reprodução em cativeiro fazem uso dos SSRs para realizar análises detalhadas da variação genética presente na população, permitindo que as informações sejam aplicadas para evitar o aumento da endogamia e a manter a viabilidade genética e demográfica da população (Henkel et al., 2012; Witzemberger e Hochkirch, 2013; Garg et al., 2016; Kleinman-Ruiz et al., 2019; Campos et al., 2021; Coleman et al., 2023). No caso do gavião-real, existem SSRs que foram transferidos de espécies filogeneticamente distantes (Banhos et al., 2008), contudo, essa abordagem nem sempre é o ideal, pois esses marcadores podem perder informações e conter alelos nulos, subestimando a verdadeira diversidade genética da espécie e afetando as análises de parentesco (Barbará et al., 2007).

Com a implementação das plataformas de sequenciamento de alto rendimento e a criação dos novos *pipelines* de bioinformática, surgiram técnicas avançadas e econômicas para isolar microssatélites espécie-específico e genotipar as amostras com maior automatização e padronização, demandando menos tempo e esforço (Gardner et al., 2011). Essa tecnologia

aumenta a resolução e acelera a obtenção de resultados, tornando as análises genéticas mais eficientes e contribuindo para a tomada de decisões mais rápidas quanto ao manejo genético da espécie. Em vista disso, Coser et al. (2014) desenvolveram microssatélites específicos para gavião-real. No entanto, a maioria dos *loci* apresentou baixo polimorfismo, não permitindo a distinção confiável entre os indivíduos de uma população. Como resultado, a capacidade de realizar análises precisas de parentesco e identificação individual para o gavião-real foi prejudicada.

Diante disso, o objetivo deste trabalho foi isolar e desenvolver novos *loci* microssatélites específicos para o gavião-real, contendo elevado polimorfismo, alto poder de exclusão de parentesco e de identificação individual para auxiliar principalmente no manejo da população *ex situ* da espécie e para garantir que a população *ex situ* se mantenha saudável e viável ao longo do tempo.

2. Material e Métodos

2.1. Desenvolvimento de novos microssatélites

Com intuito de isolar e caracterizar os novos microssatélites para o gavião-real, os *primers* foram projetados a partir de uma biblioteca genômica gerada no sequenciador Ion Torrent Personal Genome Machine (PGM) utilizando uma amostra de sangue de um indivíduo resgatado em Parintins, Amazonas, cedida pelo Projeto Harpia (www.projeto-harpia.org).

A obtenção da representação reduzida do genoma foi através da metodologia de ddRADseq com algumas adaptações (Hrbek et al., 2020), seguindo as etapas: (1) extração do DNA usando CTAB a 2% com adição de 15 mg/mL de proteinase K; (2) digestão do DNA genômico (200 ng) usando duas enzimas de restrição PstI (corte raro) e Csp6I (corte frequente) e ligação simultânea dos adaptadores P1 (corte raro) e AY (corte frequente); (3) enriquecimento do produto da digestão/ligação através da reação em cadeia da polimerase (PCR); (4) purificação do produto enriquecido com esferas magnéticas (*beads*) preparadas através do protocolo BOMB #4.2 (Oberacker et al., 2019); (5) seleção de tamanho dos fragmentos da amostra purificada entre 374 e 456 pares de base (pb) no Pippin Prep (Saga Science); (6) purificação com *beads* e quantificação no fluorômetro Qubit 2.0 (Thermo Fisher Scientific); (7) PCR clonal de emulsão e o carregamento do chip 318 Ion PGM (Life Technologies) no Ion Chef e (8) sequenciamento do chip no Ion Torrent PGM usando o kit Ion PGM Sequencing 400 (Life Technologies).

Após obter as sequências, o arquivo de saída gerado pelo sequenciador (.bam file) foi convertido para *fastq* através da função *bam2fq* do software *Samtools* 1.13 (Danecek et al., 2021) e foi filtrado para remover as sequências muito curtas (≤ 100 pb). Em seguida, os microssatélites foram filtrados e selecionados através do *pipeline* QDD3 (Meglécz et al., 2014), utilizando os parâmetros: (1) microssatélites contendo pelo menos 10 repetições (motivo tipo di ou trinucleotídeo) em comprimento; (2) ser preferencialmente simples e perfeitos; (3) seleção do tamanho do produto de PCR entre 120 e 300 pb; (4) exclusão de pares de *primers* localizados muito próximos (< 20 pb) da região de microssatélites alvo; (5) implementação do

"melhor desenho de *primer*" quando múltiplos pares de *primers* são sugeridos para amplificar a mesma região repetitiva; (6) tamanho do *primer* entre 18 e 27 pb e temperatura de pareamento entre 57 e 63°C e (7) sequências de consenso com polimorfismo detectado em suas leituras.

Os pares de *primers* foram selecionados a partir dos parâmetros estabelecidos e em seguida, foi adicionado às extremidades dos *primers forward* e *reverse* os *linkers*, 5'-CGCTCTTCCGATCT-3' e 5'-AAACGACGGCCAGT-3', complementares aos adaptadores P1 e AY da plataforma Ion Torrent. Para cada adaptador foram desenhadas variações de *barcodes* de seis nucleotídeos cada, permitindo diferentes combinações para indexação dos indivíduos. Os novos *primers* selecionados foram sintetizados pelo grupo GenOne Biotech e, posteriormente, testados referente a sua eficácia na amplificação via PCR, principalmente em relação a temperatura ideal para o pareamento correto. Por fim, os *loci* microssatélites que amplificaram corretamente foram sequenciados para as amostras de gavião-real.

2.2. Amostragem, extração de DNA e identificação do sexo

Ao todo, foram doadas amostras de sangue de 25 indivíduos de gavião-real cativos, 23 nascidos na natureza e dois nascidos em cativeiro, no Criadouro de Animais Silvestres da Itaipu Binacional (CASIB; Apêndice 1). As amostras foram disponibilizadas por quatro zoológicos e três instituições conservacionistas (Apêndice 1). Dentre os 23 indivíduos nascidos em vida livre, 13 são provenientes da floresta Amazônica, três da floresta Atlântica, um da região central do Brasil e seis de origem desconhecida. Destes indivíduos de origem desconhecida, três indivíduos são provavelmente oriundos do estado do Pará (Apêndice 1). O DNA foi extraído usando o kit Qiagen (Valencia, CA, EUA) e sua qualidade verificada no espectrofotômetro NanoDrop 2000 (Thermo Scientific). A identificação do sexo para todas as amostras foi realizada através do DNA seguindo o protocolo de Banhos et al. (2008).

2.3. Amplificação e preparação da biblioteca de sequenciamento

Para padronizar, em uma alíquota à parte, todas as amostras de DNA foram ajustadas para uma concentração de 20 ng/μL. Em seguida, todas as amostras foram amplificadas via PCR separadamente para cada par de *primers* desenvolvidos neste estudo, para evitar a formação de quimeras. A reação continha um volume final de 10 μL, consistindo em: 1 μL de DNA, 3,2 μL de Nuclease-Free ddH₂O, 1 μL de dNTPs (1 mM), 1 μL de MgCl₂ (2,5 mM), 1 μL de tampão 10X de Taq Buffer com (NH₄)₂SO₄ (75 mM Tris-HCl pH 8.8 a 25 °C, 20 mM de (NH₄)₂SO₄), 0,5 μL de BSA (2 mg), 1 μL (0,2 pM) de *primer forward* e *reverse* e 0,3 μL de Taq DNA polimerase (0,1 U). Todas as reações ocorreram no termociclador Veriti Thermal Cycler (Applied Biosystems) e as condições de ciclagem foram: (1) desnaturação inicial a 94°C por 60 s; (2) 35 ciclos de desnaturação a 94 °C por 30 s, pareamento dos *primers* a uma temperatura específica do *primer* (Tabela 1) por 30 s e extensão a 68 °C por 40 s; e (3) uma extensão final de 72 °C por 300 s.

Para formar o *pool* individual, foram transferidos 3 μL de cada produto de PCR por indivíduo/microssatélite para uma placa de 96 poços. Dessa forma, cada *pool* individual continha todas as regiões microssatélites amplificadas. Os *pools* foram purificados com as

beads para remover os produtos de PCR abaixo de 100 pb de cada *pool* individual. Para a purificação foi utilizado uma proporção de 0,8 µL de *beads* para 1 µL do *pool* individual, que foram ressuspensos no mesmo volume inicial de cada *pool*. O resultado das purificações foi visualizado em gel de agarose 1% (p/v).

Posteriormente, os produtos de PCR do *pool* individual purificados foram transferidos para uma nova placa de 96 poços, onde a indexação dos indivíduos foi realizada através da ligação dos adaptadores AY e P1 contendo combinações únicas de *barcodes* para cada indivíduo. Os reagentes para a PCR de ligação dos adaptadores foram preparados separadamente para o adaptador A (4,5 µL) e P1 (4,5 µL), contendo: 1,8 µL Nuclease-Free ddH₂O, 0,5 µL de dNTPs (2,22 mM), 0,5 µL de MgCl₂ (5,55 mM), 0,5 µL de tampão 10X de tampão (166,66 mM Tris-HCl pH 8.8 a 25 °C, 44,44 mM de (NH₄)₂SO₄), 0,25 BSA (4.44 mg), 0,4µL do primer A ou P1 (0,44 pM) e 0,15 µL de Taq DNA polimerase (0,1 U). Após os adaptadores A e P1 foram misturados em 1 µL do *pool* individual, totalizando em um volume final de 10 µL. As ligações dos adaptadores ocorreram no termociclador Veriti Thermal Cycler (Applied Biosystems) e as condições do ciclo da PCR de ligação foram: (1) desnaturação inicial a 94°C por 60 s; (2) 15 ciclos de desnaturação a 94 °C por 30 s, pareamento dos adaptadores a 48 °C por 30 s e extensão a 68 °C por 40 s; (3) extensão final de 72°C por 420 s.

Em seguida, foi transferido 3 µL de cada *pool* individual ligado aos adaptadores para um tubo de 1,5 mL e purificados novamente com as *beads*, formando a biblioteca de sequenciamento. Por fim, a preparação e o carregamento da biblioteca de sequenciamento foram realizados na plataforma Ion Chef (Thermo Fisher) e o sequenciamento no sequenciador automático Ion Torrent PGM (Thermo Fisher). Todo o processo de amplificação e preparação da biblioteca de sequenciamento está exemplificado na Figura 1. Para a demultiplexação das leituras de cada amostra sequenciada foi utilizada a ferramenta *fqgrep* (github.com/indraniel/fqgrep), buscando as combinações dos *barcodes* de seis pares de bases contidos nos adaptadores AY e P1 e armazenando dos *reads* em um arquivo *fastq* para cada amostra.

2.4. Genotipagem e análises estatísticas

Os arquivos individuais em formato *fastq* foram utilizados para identificar as sequências correspondentes aos microssatélites a partir do *pipeline* SSRseq (Lepais et al., 2020) e do programa FDSTools (Hoogenboom et al., 2017). Logo, para cada indivíduo, foi recuperado a sequência completa, contendo a região de microssatélite e as regiões extremas das repetições (região flanqueadora). Em seguida, um arquivo *fasta* foi criado para cada *locus*, contendo as variações alélicas de cada indivíduo identificadas pela ferramenta *allelefinder* do FDSTools. As sequências foram alinhadas e editadas manualmente através do AliView (Larsson, 2014). Por fim, foram gerados quatro bancos de genotipagem, sendo eles: (1) *FullLength* - o tamanho completo da sequência, incluindo as regiões flanqueadoras e a região microssatélite; (2) *RepeatFocused* - apenas o tamanho da região microssatélite; (3) SSRs - o tamanho da região microssatélite alvo e quaisquer outros tipos de repetição observados na sequência completa e

(4) CAT - dados concatenados, ou seja, todo tipo de polimorfismo encontrado na sequência (repetições + SNPs + *indels*) de maneira categórica.

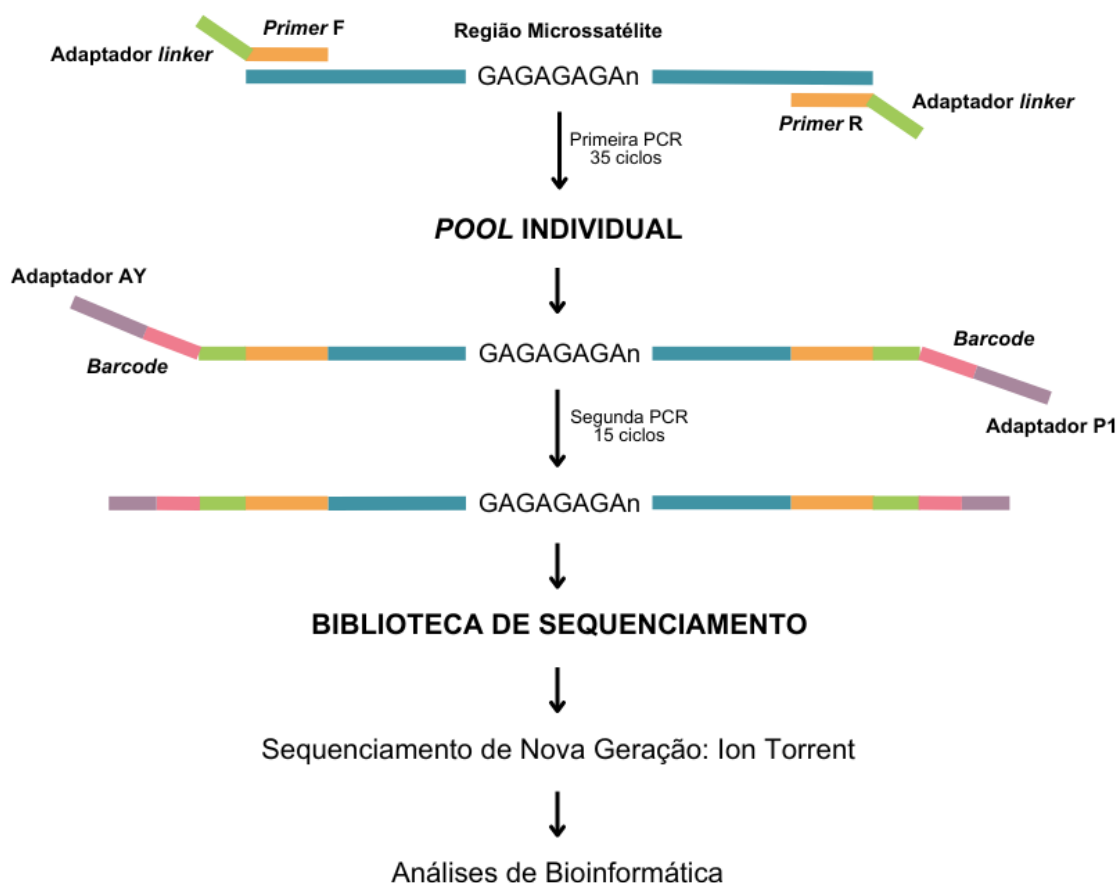


Figura 1. Representação esquemática da estratégia de amplificação e preparação da biblioteca para o sequenciamento de nova geração pela plataforma Ion Torrent.

Todos os bancos de dados foram filtrados através de um *script* personalizado no R (R Core Team, 2022), selecionando as amostras e os *loci* que continham somente 10% de dados faltantes em sua genotipagem. Os artefatos como, *stutters* da PCR, presença de alelos nulos e outros ruídos sistêmicos foram filtrados pelo FDSTools (Hoogenboom et al., 2017) e verificados pelo Micro-Checker v2.2.3 (Van Oosterhout et al., 2004). Os parâmetros de diversidade genética, incluindo o número de alelos (N_A), heterozigosidade observadas (H_O) e esperadas (H_E) e equilíbrio Hardy-Weinberg (HWE) foram obtidos no programa Arlequin 3.5 (Excoffier e Lischer, 2010). A significância estatística para o desequilíbrio de ligação foi estimada utilizando a correção de Bonferroni através do programa FSTAT 2.9.4 (Goudet, 2003). Como não houve diferenças significativas entre os quatro bancos de dados de genotipagem em relação aos parâmetros de diversidade, o método *FullLength* (1) foi escolhido para a caracterização dos microsatélites, por ser um método semelhante ao método usual de genotipagem, facilitando entendimento e o uso dessa ferramenta nos criadouros para monitoramento e no manejo reprodutivo.

A probabilidade de exclusão de paternidade (Q), a probabilidade conjunta de exclusão de paternidade em todos os *loci* (QC), a probabilidade de identidade genética (I) e a

probabilidade conjunta de identidade genética em todos os *loci* (IC) foram calculadas através do software GenAlEx 6.5 (Peakall e Smouse, 2012). Os resultados dessas probabilidades podem variar de 0 a 1. Portanto, a probabilidade de exclusão de paternidade indica a eficiência dos *loci* em excluir um indivíduo aleatório da população como pai e/ou mãe em potencial, então, quanto mais próximo de 1, mais informativo são os *loci*. Quanto à identificação individual, verifica-se a probabilidade de dois indivíduos, sorteados aleatoriamente, de uma população com o genótipo idêntico em múltiplos *loci*. Com isso, quanto menor o valor de I, maior a probabilidade de identificar o indivíduo.

Para investigar a existência de parentes próximos entre os indivíduos de gavião-real amostrados nesse estudo (N = 25), foi utilizado o software ML-Related (Kalinowski et al., 2006). Este programa calcula os coeficientes de parentesco par a par (r) com base em estimativas de máxima verossimilhança fornecendo as relações com maior probabilidade para cada par de indivíduos e os classifica em quatro tipos de relações de parentesco mais comuns: não relacionados (U), meio-irmãos (HS), irmãos completos (FS) e pais-filhos (PO). Adicionalmente o software gera um conjunto de confiança para cada relação entre os pares e compara as relações potenciais com as relações alternativas.

Com base nisso, adotou-se uma abordagem conservadora para a classificação de parentesco, sendo considerados como relacionados apenas os pares nos quais a categoria U não estava presente no conjunto de confiança de 95%. Aqueles com apenas a categoria U inclusa no conjunto foram considerados não relacionados, enquanto os pares que apresentaram qualquer combinação de categorias relacionadas e não relacionadas foram considerados como tendo uma relação ambígua. Os pares identificados como relacionados pela análise foram submetidos a um teste de hipótese utilizando um teste de razão de verossimilhança entre a relação com maior probabilidade e as relações alternativas estimadas pelo conjunto de confiança. Para cada teste foi utilizado 100.000 simulações e um limiar de significância de $p < 0,05$. Se o valor de p fosse menor que o nível de significância, era possível concluir que a relação com maior probabilidade se ajusta aos dados significativamente melhor do que as relações alternativas.

3. Resultados

3.1. Obtenção dos novos microssatélites

Após o sequenciamento e pré-processamento dos *reads* da biblioteca genômica do gavião-real, foi obtido um total de 141.281 *reads*, com fragmentos variando entre 101 a 605 pb, com tamanho médio de 311 pb. Destes, um total de 2.860 sequências apresentaram regiões microssatélites após executar o PIPE1 do QDD3. Após o PIPE2, um total de 1.046 *singletons* e 412 sequências consenso únicas (total de 1.458 sequências) foram obtidas, que após a execução do PIPE3, resultaram em 1.020 sequências contendo microssatélites dentro das nossas especificações e um total de 12.664 desenhos alternativos de *primers* para estas regiões microssatélites. Dentre as 1.020 sequências de microssatélites, as mais frequentes foram na forma de dinucleotídeos (78,52%), seguido por tri (16,67%), tetra (3,13%), penta

(1,18%) e hexa (0,50%). Desses microssatélites disponíveis, dezenove (15 dinucleotídeos e 4 tri) foram escolhidos para genotipar os indivíduos de gavião-real. Quanto ao teste de eficácia dos 19 pares *primers* recém-desenvolvidos, dois pares não foram eficientes na amplificação via PCR, restando 17 regiões microssatélites para a genotipagem.

3.2. Caracterização dos novos microssatélites

A genotipagem dos indivíduos obteve um total de 306.403 *reads* brutos, com o número de *reads* por amostra variando entre 8.859 a 15.793 e um valor médio de 12.256 *reads*. Dos 17 microssatélites sequenciados, apenas o *locus* Hha16 não foi sequenciado corretamente, sendo descartado do conjunto de dados, restando 16 *loci* no total. Destes, a filtragem de *locus* e de amostras com mais de 10% de dados faltantes removeu 4 *loci* (Hha07, Hha11, Hha13 e Hha15) e um indivíduo (HZ14). Além disso, dois *loci* (Hha10 e Hha17) apresentaram monomorfismo e foram removidos das análises subsequentes. Ao final, 10 *loci* e 24 indivíduos foram selecionados para a caracterização dos microssatélites (Tabela 1).

Quanto à integridade dos dados, não foram observados artefatos de *stutters* e nem a presença de alelos nulos para os *loci* desenvolvidos. Além disso, nenhuma evidência significativa foi detectada para o desequilíbrio de ligação após a correção de Bonferroni, indicando que os *loci* não estão ligados fisicamente e podem ser considerados marcadores independentes. Quatro *loci* (Hha02, Hha03, Hha06 e Hha09) apresentaram desvios significativos ($P < 0.05$) para o equilíbrio de Hardy-Weinberg (HWE). No que se refere às medidas de diversidade genética para a caracterização dos 10 *loci* microssatélites polimórficos, o número de alelos (N_A) variaram de 2 a 8, com o número médio de 4.4. A heterozigosidade observada (H_O) e esperada (H_E) variaram de 0.0416 a 0.8750 e de 0.0416 a 0.8102, respectivamente. As médias da heterozigosidade observada e esperada foram de 0.3541 e 0.4432, respectivamente (Tabela 1).

Em relação à probabilidade de exclusão de parentesco, os valores de Q variaram entre 0.0387 e 0.7328, somando os esforços, o valor de QC foi de 0.9938 (Tabela 1; Apêndice 2). Quanto à identificação individual, os valores de I variaram entre 0.0928 e 0.9208, e o IC foi de 2.6684×10^{-5} (Tabela 1; Apêndice 2). Contudo, selecionando os *loci* Hha02, Hha03, Hha05 e Hha06 e os combinando é possível ter um resultado similar e eficiente para exclusão de parentesco e identificação individual usando um menor número de marcadores ($QC = 0.9661$ e $IC = 1.2505 \times 10^{-3}$; Apêndice 3).

Analisando os pares de coeficiente de parentesco com maior probabilidade e suas respectivas categorias, foram 237 pares categorizados como U, 26 como HS, 20 como FS e 17 como PO (Apêndice 4). Porém, averiguando o conjunto de confiança (Apêndice 5) foi possível assumir apenas 18 pares relacionados, os quais apresentaram relacionamento HS, FS ou PO com conjunto de confiança em que a categoria U não estava presente; 44 não relacionados, com apenas a categoria U inclusa no conjunto e o restante apresentou relacionamentos ambíguos, impedindo a identificação do relacionamento entre os pares com maior exatidão.

Os pares com maior probabilidade de estarem relacionados, mas que incluem em seu conjunto de confiança a relação U (ou seja, um dado ambíguo) foram submetidos a testes de

hipótese. Como resultado desses testes, observou-se que 8 pares (HZ03-HZ19, HZ05-HZ13, HZ06-HZ23, HZ07-HZ13, HZ12-HZ13, HZ14-HZ18, HZ14-HZ22 e HZ14-HZ25) com alta probabilidade de estarem relacionados (HS, FS ou PO; Apêndice 6) apresentaram um valor menor que o nível de significância quando foram testados quanto à possibilidade de não serem relacionados. Isso indica que o relacionamento sugerido pela análise com maior probabilidade se ajusta melhor do que a alternativa de não haver relação entre os pares. Quanto aos 18 pares relacionados sem apresentar a relação U em seu conjunto de confiança, 13 foram considerados FS e cinco PO com maior probabilidade. Dentre esses, está presente uma família com relacionamento conhecido a priori. Essa família é composta por quatro indivíduos, onde HZ15 e HZ16 são irmãos completos nascidos no Criadouro de Animais Silvestres da Itaipu Binacional (CASIB) a partir de um cruzamento entre os indivíduos HZ12, uma fêmea de origem desconhecida, e HZ13, um macho oriundo do Paraná, divisa com Argentina.

Diante dessa informação a priori, os relacionamentos esperados entre HZ12-HZ15, HZ12-HZ16, HZ13-HZ15 e HZ13-HZ16 eram de pais-filhos (PO) e entre HZ15-HZ16 irmãos-completos (FS). O valor estimado para as relações foram corretos (Apêndice 4 e 5), exceto para o relacionamento entre HZ13-HZ16, que foi estimado como FS ao invés de PO e HZ15-HZ16 como PO ao invés de FS. Cabe ressaltar que os pais (HZ12-HZ13) também apresentaram alto valor de parentesco, com maior probabilidade de serem PO, porém o conjunto de confiança resgatou relacionamentos alternativos, incluindo U e HS (Apêndice 5). Por esse motivo e por optar pela conservação dos dados assumimos um relacionamento ambíguo, mas com grandes chances desse cruzamento ter sido realizado entre parentes pelo fato do teste de hipótese apresentar um valor abaixo do nível de significância (p -valor = 0.029) para o relacionamento U.

Excluindo a família com relacionamento conhecido, o restante dos pares relacionados (13 pares) a maioria dos casos envolveram indivíduos provenientes do mesmo estado ou de localidades próximas. Entre esses pares, seis casos corresponderam a indivíduos oriundos do estado do Pará (HZ03-HZ09, HZ06-HZ08, HZ06-HZ24, HZ08-HZ23, HZ08-HZ24 e HZ23-HZ24); outros quatro casos, entre indivíduos oriundos de estados vizinhos, sendo dois casos entre Pará e Bahia (HZ05-HZ19; HZ06-HZ19), um entre Pará e Amapá (HZ17-HZ25) e um entre Mato Grosso e Rondônia (HZ14-HZ20); e por último, três casos entre indivíduos de origem conhecida e outro de origem desconhecida, sendo um indivíduo oriundo do estado de Rondônia (HZ01-HZ21) e dois do estado do Pará (HZ03-HZ11 e HZ05-HZ12) (Apêndice 4).

3.3. Amostragem e identificação do sexo

Inicialmente, de acordo com o banco de dados fornecido pelas instituições, havia 12 machos e 13 fêmeas. No entanto, ao analisar o DNA dos 25 indivíduos, constatou-se que cinco indivíduos anteriormente identificados como machos eram, na verdade, fêmeas (HZ01; HZ05; HZ06; HZ09; HZ10; Apêndice 1), correspondendo a 20% de identificação incorreta da amostragem realizada neste estudo ($N = 25$). Além disso, não houve nenhuma identificação incorreta de fêmeas. Portanto, após essa correção, foram contabilizados sete machos e 18 fêmeas.

Tabela 1. Caracterização de 10 microssatélites em 24 indivíduos de gavião-real (*Harpia harpyja*) mantidos em cativeiro no Brasil.

<i>Locus</i>	Repetição Motif	Sequência do <i>primer</i> (5' - 3')	Tamanho (bp)	T _a (°C)	N	N _A	H _O	H _E	P _{HWE}	Q	I
Hha01	(AC) ₁₀	F- CTCCTGTGTTCCCATCTGCT R- GGGTGCGAGACAGTTCCTC	158	59	24	2	0.2500	0.2234	1.000	0.1667	0.6342
Hha02	(AC) ₁₀	F- TCTGTTGGAGTTTCCGGAGC R- CCGGGATGCATCTCCTTTGT	192	59	24	8	0.4583	0.8102	0.000*	0.7328	0.0928
Hha03	(AG) ₁₄	F- ATGGCAAGTCGCTAGAACGG R- GGCTGCTATTCCACTTGCCA	204	57	24	5	0.5416	0.6693	0.004*	0.5172	0.2154
Hha04	(AC) ₁₂	F- GTGGAAGGGAGTTAGCGTGG R- ATGGCTCTGTCTAACCTGCG	125	59	24	2	0.0416	0.0416	1.000	0.0387	0.9208
Hha05	(AGC) ₁₁	F- ACAAGGTCACATCTTCCCGC R- AAAGTGTCTGCCTCGCTGA	179	59	24	8	0.4583	0.5186	0.235	0.4269	0.3504
Hha06	(AG) ₁₂	F- ACCAGGCTGTAAGGGTTGAAG R- TGGACAGATGTGGAGTGTCTG	121	59	24	4	0.8750	0.6817	0.000*	0.5421	0.1783
Hha08	(AC) ₁₀	F- CATCCGTAGCCATGCACAGA R- CTTACCTGGGTGTCAGCACA	155	57	24	4	0.3333	0.4140	0.207	0.3507	0.3899
Hha09	(AC) ₁₂	F- ACCCAAGCTCTTCCTGGATG R- TCAAAGCGTATGAAGCCAGGA	198	57	24	4	0.0833	0.4609	0.000*	0.3786	0.3483
Hha12	(AG) ₁₁	F- TCCTCCTTGAAGCAGCAAG R- AGCTCAGTTTCTGGAGTCAGG	120	57	24	3	0.2916	0.3448	0.112	0.2527	0.4833
Hha18	(AC) ₁₁	F- GCCCACAGAAACGATACAGC R- ATTAGCTGCTCGCAGACAAA	150	57	24	4	0.2083	0.2677	0.252	0.2496	0.5564
							média = 0.3541	média = 0.4432		QC = 0.9938	IC = 2.6684x10 ⁻⁵

F, *forward*; R, *reverse*; bp, pares de base; T_a, temperatura de pareamento dos pares de *primers* (°C); N, número de indivíduos genotipados; N_A, número de alelos; H_O, heterozigosidade observada; H_E, heterozigosidade esperada; P_{HWE}, probabilidade de desvio para o equilíbrio de Hardy-Weinberg (*p*-valor); Q, probabilidade de exclusão de paternidade em um *locus*; QC, probabilidade de exclusão de paternidade para todos os *loci*; I, probabilidade de identidade genética em um *locus*; IC, probabilidade de identidade genética para todos os *loci*. Desvio significativo do equilíbrio de Hardy-Weinberg em **p*<0,05.

4. Discussão

No Brasil, esforços para reproduzir o gavião-real em cativeiro tiveram início em 1995 (Azeredo, 2005). Desde então, com a contribuição da Sociedade de Pesquisa de Fauna Silvestre (CRAX) e do RBV da Itaipu Binacional, o número de indivíduos nascidos em cativeiro aumentou com o passar dos anos e atualmente, quando comparado com outros países, o Brasil contém a maior população *ex situ* do gavião-real, contendo 86 indivíduos selvagens e 53 nascidos em cativeiro (N = 139) até o ano de 2020 (Oliveira et al., 2022). De fato, a reprodução em cativeiro do gavião-real se tornou um sucesso, porém os dados genéticos básicos da população *ex situ* não foram levantados anteriormente e a escolha dos pares reprodutivos ainda é realizada através da compatibilidade comportamental entre os indivíduos e dos registros históricos individuais como, origem, idade e sexo (Watson et al., 2016; Oliveira et al., 2022).

Entretanto, muitas vezes esses registros históricos possuem imprecisões ou lacunas de informações que podem comprometer significativamente o sucesso do manejo *ex situ* e a conservação da espécie (Bowling et al., 2003; Oliveira et al., 2022). Como exemplo, neste estudo, seis indivíduos de gavião-real não continham informação concreta sobre a sua origem e de acordo com Oliveira et al. (2022) cerca de 26% da população *ex situ* total possui origem desconhecida. A falta dessa informação dificulta direcionar boas estratégias para os programas que visam à reprodução em cativeiro e futuras reintroduções, pois populações de diferentes localidades podem apresentar variações genéticas que resultaram de adaptações específicas a ambientes locais. Então, se essa informação não for considerada, pode resultar em uma perda potencial de variação genética para a população em cativeiro (Allendorf et al., 2013; Frankham et al., 2017).

Outra imprecisão dos registros individuais notada neste estudo foi a identificação equivocada do sexo dos indivíduos. Por meio da sexagem molecular, foi constatado que cinco indivíduos tiveram seu sexo erroneamente determinado (Apêndice 1). Dado que o gavião-real não apresenta dimorfismo sexual e a distinção entre os sexos ocorre somente pelo tamanho corporal, com os machos sendo menores que as fêmeas, a identificação do sexo de forma visual pode ser imprecisa, principalmente quando os indivíduos são jovens. Esse erro tem um impacto direto em programas de reprodução principalmente na seleção de novos pares reprodutivos e na compatibilidade comportamental entre os indivíduos reprodutores. Portanto, o método mais confiável para determinar o sexo dos gaviões-reais é através da análise molecular do DNA (Banhos et al., 2008).

Ademais, vale enfatizar, que o pareamento dos indivíduos mantidos em cativeiro baseando-se apenas nas informações dos registros históricos pode subestimar o nível de parentesco entre os pares, podendo acarretar sérios problemas genéticos na população *ex situ*, como aumento da endogamia e perda da diversidade genética (Witzenberger e Hochkirch, 2011; Frankham et al., 2017). Portanto, para garantir a saúde e a viabilidade da população em cativeiro a longo prazo, realizar análises moleculares fornece dados mais realistas sobre as relações de parentesco dentro de populações cativas (Gonçalves-Silva et al., 2010;

Witzenberger e Hochkirch, 2011; 2013; Garg et al., 2016; Kleinman-Ruiz et al., 2019; Campos et al., 2021; Coleman et al., 2023). Nesse contexto, o conjunto de microssatélites desenvolvidos neste estudo desempenha um papel crucial ao possibilitar a obtenção de informações genéticas dos indivíduos facilitando um manejo mais eficiente e preciso da população *ex situ* do gavião-real.

Os novos loci que apresentaram desequilíbrio de HWE provavelmente são resultado da mistura de indivíduos pertencentes a diferentes populações (Apêndice 1), conferindo o efeito Wahlund, o qual é mais pronunciado à medida que aumenta a diferenciação entre as populações (Waples, 2015). Estudos anteriores evidenciaram diferenças genéticas para o gavião-real no Brasil (Banhos, 2009; Banhos et al., 2016). Tais informações também precisam ser consideradas no manejo *ex situ*, pois embora misturar indivíduos de origem diferentes possa reduzir a probabilidade de depressão endogâmica e aumentar a diversidade genética, também pode aumentar o risco de depressão exogâmica (Frankham et al., 2017). Assim, é importante avaliar e evitar a formação de pares reprodutivos no caso de existir diferentes unidades evolutivas e ao realizar translocações de uma área para outra ou reintroduzi-los em seu habitat natural é necessário ter cautela para não introduzir linhagens ou grupos genéticos inadequados para o local específico.

O alto valor para a exclusão de parentesco ($QC = 0.9938$) e o baixo índice de identidade genética ($IC = 2.6684 \times 10^{-5}$) indicam que o conjunto de loci desenvolvidos possuem grande potencial para excluir os prováveis pais de um indivíduo e realizar a identificação individual. Além disso, também é possível realizar a identificação de parentes e de indivíduos com alta confiabilidade utilizando um painel com somente quatro marcadores (Hha02, Hha03, Hha05 e Hha06; Apêndice 3). Essas informações individuais obtidas através dos marcadores moleculares são de extrema importância para o conhecimento da população *ex situ* do gavião-real, visto que muitos indivíduos chegam ao recinto sem informações sobre sua origem (Oliveira et al., 2022) e automaticamente são assumidos como tendo alto parentesco na população ou não relacionado (Nielsen et al., 2007; Gonçalves-Silva et al., 2010). Contudo, sem informações genéticas, essas suposições imprecisas podem excluir indivíduos importantes geneticamente para a reprodução ou subestimar o nível de parentesco entre os pares.

Espécies ameaçadas de extinção, como o gavião-real, geralmente possuem um número limitado de indivíduos reprodutores em cativeiro, com isso, aumenta a probabilidade de ocorrer cruzamentos entre parentes (Frankham et al., 2010). Assim, saber o coeficiente de parentesco entre os indivíduos possui um significado biológico extremamente importante para programas de reprodução em cativeiro e devem ser utilizados para acrescentar informações aos registros históricos, direcionar as translocações dos indivíduos entre as instituições e na formação dos pares reprodutivos menos aparentados, com intuito de gerar filhotes que irão amenizar a tendência endogâmica da população *ex situ* e, ao mesmo tempo, preservar a diversidade genética existente da espécie (Witzenberger e Hochkirch, 2011; Frankham et al., 2017).

A análise de parentesco evidenciou a eficiência dos novos marcadores desenvolvidos neste estudo em detectar vínculo genético entre os indivíduos. Tal eficiência pode ser observada claramente nos relacionamentos de primeiro grau esperados para a família de gaviões-reais pertencente à RVB da Itaipu Binacional (Apêndice 4). Também foi evidenciado que a união dos pais HZ12 e HZ13, baseada na compatibilidade comportamental e nos registros históricos dos indivíduos, subestimou a relação de parentesco com maior potencial entre esses indivíduos. Mesmo que a presença do relacionamento U no conjunto de confiança possa criar alguma ambiguidade, os testes de hipótese indicaram a existência de algum grau de parentesco entre HZ12 e HZ13 (Apêndice 5 e 6). Com base nessa informação, além dos valores de maior probabilidade de relacionamento, é importante avaliar com cautela o conjunto de confiança e realizar os testes de hipótese para obter maior exatidão ao excluir ou considerar possíveis pares relacionados.

Além disso, o novo conjunto de microssatélites revelou a existência de outros indivíduos que quando pareados apresentam um determinado nível de parentesco. A maioria desses indivíduos provém de áreas geograficamente próximas, incluindo estados vizinhos como Pará e Amapá, Mato Grosso e Rondônia, assim como municípios vizinhos dentro do mesmo estado, como Santarém e Tucuruí. Essas associações sugerem a ocorrência da reprodução do gavião-real nessas regiões e sua dispersão ao redor. Mais uma vez, torna-se evidente a importância dos dados moleculares para conhecer e orientar a formação de pares reprodutivos, destacando que não é suficiente considerar apenas a origem conhecida ao formar casais, uma vez que o gavião-real possui alta capacidade de explorar vastas áreas durante seus voos podendo se dispersar por longas distâncias (Urios et al., 2017; Naveda-Rodríguez et al. 2022).

Quanto à diversidade genética, os microssatélites desenvolvidos para o gavião-real neste estudo apresentaram níveis mais baixos quando comparados com o estudo realizado a partir de microssatélites transferidos por Banhos et al. (2016). Entretanto, os autores detectaram perda de diversidade genética ao longo do tempo para a espécie a partir de uma ampla amostragem histórica e atual na Amazônia e na Mata Atlântica (Banhos et al. 2016). Ademais, foram observados níveis de diversidade genética maiores para outras águias também classificadas como vulneráveis, como a *Aquila adalberti* (HO = 0.516 e HE = 0.549) e *Aquila heliaca* (HO = 0.563 e HE = 0.627) (Martínez-Cruz et al., 2004). Tais descobertas destacam a importância de monitorar a diversidade genética do gavião-real e implementar estratégias de manejo adequadas para garantir a sua manutenção ou até mesmo maximizar a diversidade a longo prazo.

Coser et al. (2014) também desenvolveram microssatélites específicos para o gavião-real e encontraram baixos níveis de diversidade genética para a espécie (HO = 0.364 e HE = 0.365). Curiosamente, seus resultados quanto à diversidade genética foram semelhantes aos encontrados neste estudo, sugerindo ser uma característica específica dos marcadores microssatélites para essa espécie. No entanto, os mesmos autores identificaram loci menos polimórficos, inviabilizando análises de parentesco e identificação individual. Por outro lado, no

estudo atual, a seleção criteriosa do melhor conjunto de marcadores e o uso da genotipagem por sequenciamento de nova geração foi crucial para resgatar loci com maior variação sendo úteis para identificar o parentesco entre os indivíduos. Além disso, a genotipagem através do sequenciamento de nova geração acelera a obtenção dos resultados nas tomadas de decisões para o manejo da população *ex situ* (Gardner et al., 2011; Putman and Carbone, 2014; Rosazlina et al., 2015).

A baixa diversidade genética encontrada nos microssatélites é preocupante para a viabilidade das populações do gavião-real, considerando o importante papel da diversidade genética na adaptação da espécie às mudanças ambientais (Frankham et al., 2010). No entanto, é relevante notar que o número limitado de indivíduos cativos amostrados (24 de 86 nascidos em vida livre) pode ter influenciado nesse resultado. Além disso, enquanto a conservação *ex situ* busca proteger a diversidade genética total da espécie, nossos dados indicam que a população em cativeiro estudada não abrange toda a variação genética na natureza, ao contrário de estudos anteriores com amostragens mais amplas encontrando maior diversidade (Banhos 2009; Banhos et al., 2016). Dessa forma, uma abordagem que incluía mais indivíduos e de diferentes origens pode recuperar maior diversidade genética, com a presença de diferentes alelos.

Por fim, os marcadores moleculares microssatélites juntamente com os marcadores sexuais aprimorados por Banhos et al. (2008) demonstraram ser ferramentas moleculares eficazes para preencher lacunas presentes nos registros históricos dos indivíduos e gerir a população *ex situ* de acordo com resultados adquiridos pela diversidade genética e dos níveis de parentesco encontrados. Adicionalmente sugerimos que esses marcadores sejam aplicados para toda a população *ex situ* do gavião-real, para obter mais informações sobre cada indivíduo não amostrado nesse estudo e poder direcionar estratégias mais assertivas para o manejo genético, atuando principalmente no pareamento de casais menos aparentados para manter populações autossustentáveis e demograficamente estáveis, capazes de reter diversidade genética e evitar a depressão endogâmica ao longo do tempo, cumprindo com o objetivo da conservação *ex situ*.

5. Agradecimentos

Às Instituições que forneceram amostras de sangue do gavião-real, Fundação Parque Zoológico de São Paulo, Hayabusa Falcoaria e Consultoria Ambiental, Parque das Aves, Refúgio Biológico Bela Vista (RBV) da Itaipu Binacional, Zoológico de Belo Horizonte, Zoológico de Brasília, Zoológico Municipal de Curitiba. À Universidade Federal do Amazonas e ao Laboratório de Evolução e Genética Animal onde as etapas laboratoriais foram desenvolvidas.

6. Financiamento

O presente trabalho teve apoio financeiro da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001, da fundação

Tiergarten-Nürnberg, da Associação de Zoológicos e Aquários do Brasil (AZAB) e do Projeto Harpia.

7. Referências

Aguiar-Silva FH, Junqueira TG, Sanaïotti TM, Guimarães VY, Mathias PVC, Mendonça CV (2015) Resource availability and diet in Harpy Eagle breeding territories on the Xingu River, Brazilian Amazon. *Braz J of Biol* 75:181-189. 10.1590/1519-6984.00914BM

Allendorf FW, Luikart GH, Aitken SN (2013) Conservation and the genetics of populations. John Wiley & Sons, New York

Azeredo R (2005) Reprodução de *Harpia harpyja* em cativeiro. *Rapina* 1:6-13

Banhos A, Hrbek T, Gravena W, Sanaïotti TM, Farias IP (2008) Genomic resources for the conservation and management of the harpy eagle (*Harpia harpyja*, Falconiformes, Accipitridae). *Genet and Mol Biol* 31:146-154. 10.1590/S1415-47572008000100025

Banhos A, Hrbek T, Sanaïotti TM, Farias IP (2016) Reduction of genetic diversity of the Harpy Eagle in Brazilian tropical forests. *PloS One* 11:1-12. <https://doi.org/10.1371/journal.pone.0148902>

Banhos A, Sanaïotti TM, Aguiar-Silva FH, Martins FD, Luz BB, Carvalho AS, Ruiz CM (2018) *Harpia harpyja* (Linnaeus, 1758). Livro Vermelho da Fauna Ameaçada de Extinção 3:124-128

Barbará T, Palma-Silva C, Paggi GM, Bered F, Fay MF, Lexer C (2007) Cross-species transfer of nuclear microsatellite markers: potential and limitations. *Mol Ecol* 16:3759-3767. <https://doi.org/10.1111/j.1365-294X.2007.03439.x>

BirdLife International (2023) Species factsheet: *Harpia harpyja*. (<http://datazone.birdlife.org/species/factsheet/harpy-eagle-harpia-harpyja>). Acesso em: 10 fev 2023

Bowling AT, Zimmermann W, Ryder O, Penado C, Peto S, Chemnick L, Yasinetskaya N, Zharkikh T (2003) Genetic variation in Przewalski's horses, with special focus on the last wild caught mare, 231 Orlitza III. *Cytogenet and Gen Res* 102:226-234. <https://doi.org/10.1159/000075754>

Campos CI, Martinez MA, Acosta D, Diaz-Luque JA, Berkunsky I, Lamberski NL, Cruz-Nieto J, Russello MA, Wright TF (2021) Genetic diversity and population structure of two endangered neotropical parrots inform *in situ* and *ex situ* conservation strategies. *Divers* 13:386. <https://doi.org/10.3390/d13080386>

Coleman JL, Wyffels JT, Penfold LM, Richardson D, Maddox JD (2023) Development of genetic markers for reproductive management of toucans. *Zoo Biol* 1-9. <https://doi.org/10.1002/zoo.21792>

Coser R, Martínez JG, Nunes M, Assunção EN, Astolfi-Filho S, Banhos A, Sanaiotti TM, Hrbek T, Farias IP (2014) Development of microsatellite markers for the near threatened eagles *Harpia harpyja* and *Morphnus guianensis* using next-generation sequencing. *Conserv Genet Resour* 6:999-1001. <https://doi.org/10.1007/s12686-014-0268-7>

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H (2021) Twelve years of SAMtools and BCFtools. *GigaSci* 10:giab008. [10.1093/gigascience/giab008](https://doi.org/10.1093/gigascience/giab008)

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 10:564-567. <https://doi.org/10.1111/j.1755-0998.2010.02847.x>

Frankham R, Ballou JD, Briscoe DA (2010) *Introduction to Conservation Genetics*. Cambridge University Press, Cambridge

Frankham R, Ballou JD, Ralls K, Eldridge M, Dudash MR, Fenster CB, Lacy RC, Sunnucks, P (2017) *Genetic management of fragmented animal and plant populations*. Oxford University Press, Oxford

Gardner MG, Fitch AJ, Bertozzi T, Lowe AJ (2011) Rise of the machines—recommendations for ecologists when using next generation sequencing for microsatellite development. *Mol Ecol Resour* 11:1093-1101. <https://doi.org/10.1111/j.1755-0998.2011.03037.x>

Garg S, Javed R, Sood S, Singh H (2016) Genetic evaluation of *ex situ* conservation breeding projects of cheer pheasant (*Catreus wallichii*) and western tragopan (*Tragopan melanocephalus*) in India. *Zoo Biol* 35:269-273. <https://doi.org/10.1002/zoo.21263>

Gonçalves-Silva A, Lalonde DR, Quse V, Shoemaker A, Russello MA (2010) Genetic approaches refine *ex situ* lowland tapir (*Tapirus terrestris*) conservation. *J of Hered* 101:581-590. <https://doi.org/10.1093/jhered/esq055>

Goudet J (1995) FSTAT (Version 1.2): A Computer Program to Calculate F-Statistics. *J of Heredity* 86:485-486

Henkel JR, Jones KL, Hereford SG, Savoie ML, Leibo SP, Howard JJ (2012) Integrating microsatellite and pedigree analyses to facilitate the captive management of the endangered Mississippi sandhill crane (*Grus canadensis pulla*). *Zoo Biol* 31:322-335. <https://doi.org/10.1002/zoo.20399>

Hoogenboom J, Van-der-Gaag KJ, Leeuw RH, Sijen T, Knijff P, Laros JF (2017) FDSTools: A software package for analysis of massively parallel sequencing data with the ability to recognise and correct STR stutter and other PCR or sequencing noise. *Forensic Sci Int Genet* 27:27-40. [10.1016/j.fsigen.2016.11.007](https://doi.org/10.1016/j.fsigen.2016.11.007)

Hrbek T, Martínez JG, Hernandez-Rangel SM, Assunção EN, Bertuol F, Canton R, Astolfi-Filho S, Farias IP (2020) Optimizing Next Generation Sequencing for biodiversity studies on the IonTorrent PGM: an in silico and in vitro analysis. PeerJ (accepted)

Kalinowski ST, AP Wagner, ML Taper (2006) ML-Relate: a computer program for maximum likelihood estimation of relatedness and relationship. Mol Ecol Notes 6:576-579. <https://doi.org/10.1111/j.1471-8286.2006.01256.x>

Kleinman-Ruiz D, Soriano L, Casas-Marce M, Szychta C, Sánchez I, Fernández J, Godoy JA (2019) Genetic evaluation of the *Iberian lynx ex situ* conservation programme. Hered 123:647-661. <https://doi.org/10.1038/s41437-019-0217-z>

Larsson A (2014) AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinform 30:3276-3278. <https://doi.org/10.1093/bioinformatics/btu531>

Lepais O, Chancerel E, Boury C, Salin F, Manicki A, Taillebois L, Dutech C, Aissi A, Bacles CFE, Daverat F, Launey S, Guichoux E (2020) Fast sequence-based microsatellite genotyping development workflow. PeerJ 8:e9085. 10.7717/peerj.9085

Lynch M, Ritland K (1999) Estimation of pairwise relatedness with molecular markers. Genet 152:1753-1766

Martínez-Cruz B, Godoy JA, Negro JJ (2004) Population genetics after fragmentation: the case of the endangered Spanish imperial eagle (*Aquila adalberti*). Mol Ecol 13:2243-2255. <https://doi.org/10.1111/j.1365-294X.2004.02220.x>

Megléczy E, Pech N, Gilles A, Dubut V, Hingamp P, Trilles A, Grenier R, Martin JF (2014) QDD version 3.1: A user friendly computer program for microsatellite selection and primer design revisited: Experimental validation of variables determining genotyping success rate. Mol Ecol Resour 14:1302-1313. <https://doi.org/10.1111/1755-0998.12271>

Naveda-Rodríguez A, Campbell-Thompson E, Watson RT, McCabe J, Vargas FH (2022) Dispersal and space use of captive-reared and wild-rehabilitated Harpy Eagles released in central American landscapes: implications for reintroduction and reinforcement management. Divers 14:886. <https://doi.org/10.3390/d14100886>

Nielsen RK, Pertoldi C, Loeschcke V (2007) Genetic evaluation of the captive breeding program of the Persian wild ass. J of Zool 272:349-357. Nielsen, R. K., Pertoldi, C., & Loeschcke, V. (2007). Genetic evaluation of the captive breeding program of the Persian wild ass. Journal of zoology, 272(4), 349-357.

Oberacker P, Stepper P, Bond DM, Höhn S, Focken J, Meyer V, Schelle L, Sugrue VJ, Jeunen GJ, Moser T, Hore SR, Meyenn FV, Hipp K, Hore TA, Jukowski TP (2019) Bio-On-Magnetic-Beads (BOMB): Open platform for high-throughput nucleic acid extraction and manipulation. PLoS Biol 17: e3000107. <https://doi.org/10.1371/journal.pbio.3000107>

- Oliveira MJ, Aguiar-Silva FH, Moraes W, Sanaiotti TM, Banhos A, Moreira N (2022) *Ex situ* population of the Harpy Eagle and its potential for integrated conservation. ZooKeys 1083:109-128. 10.3897/zookeys.1083.69047
- Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinform 28:2537-2539. 10.1093/bioinformatics/bts460
- Putman AI, Carbone I (2014) Challenges in analysis and interpretation of microsatellite data for population genetic studies. Ecol and Evol 4:4399-4428. <https://doi.org/10.1002/ece3.1305>
- R Core Team (2022) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Rosazlina R, Jacobsen N, Ørgaard M, Othman AS (2015) Utilizing next generation sequencing to characterize microsatellite loci in a tropical aquatic plant species *Cryptocoryne cordata* var. *cordata* (Araceae). Biochem Syst and Ecol 61:385-389. <https://doi.org/10.1016/j.bse.2015.06.033>
- Sanaiotti TM, Junqueira TG, Palhares V, Aguiar-Silva FH, Henriques LMP, Oliveira G, Guimarães VY, Castro V, Mota D, Trombin DF, Villar DNA, Lara KM, Fernandes D, Castilho L, Yosheno E, Alencar RM, Cesca L, Dantas SM, Laranjeiras TO, Mathias PC, Mendonça CV (2015) Abundance of Harpy and Crested Eagles from a reservoir-impact area in the Low-and Mid-Xingu River. Braz J of Biol 75:190-204. <https://doi.org/10.1590/1519-6984.00614BM>
- Soares ES, Amaral FSR, Carvalho-Filho EPM, Granzinolli MA, Albuquerque MJLB, Lisboa JS, Azevedo MAG, Moraes W, Sanaiotti TM, Guimarães IG (2008) Plano de Ação Nacional para Conservação de Aves de Rapina. Instituto Chico Mendes de Conservação da Biodiversidade, Coordenação-Geral de Espécies Ameaçadas. Brasília
- Trinca CT, Ferrari SF, Lees AC (2008) Curiosity killed the bird: arbitrary hunting of Harpy Eagles *Harpia harpyja* on an agricultural frontier in southern Brazilian Amazonia. Cotinga 30:12-15
- Turchetto-Zolet AC, Turchetto C, Zanella CM, Passaia G (2017) Marcadores moleculares na era genômica: Metodologias e aplicações. Sociedade Brasileira de Genética, Ribeirão Preto
- Urios V, Muñiz-López R, Vidal-Mateo J (2017) Juvenile dispersal of Harpy Eagles (*Harpia harpyja*) in Ecuador. J of Raptor Res 51:439-445. <https://doi.org/10.3356/JRR-16-54.1>
- Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes 4:535-538. 10.1111/j.1471-8286.2004.00684.x

Vargas GJJ, Whitacre D, Mosquera R, Albuquerque J, Piana R, Thiollay JM, Márquez C, Sánchez JE, López ML, Matola S, Aguilar S, Rettig N, Sanaïotti TM (2006) Estado y distribución actual del Águila Arpía (*Harpia harpyja*) en Centro y Sur América. Ornitol Neotrop 17:39-55. <https://repositorio.inpa.gov.br/handle/1/22603>

Waples RS (2015) Testing for Hardy-Weinberg proportions: have we lost the plot?. J of Heredity 106:1-19. <https://doi.org/10.1093/jhered/esu062>

Watson RT, McClure CJ, Vargas FH, Jenny JP (2016) Trial restoration of the Harpy eagle, a large, long-lived, Tropical Forest Raptor, in Panama and Belize. J of Raptor Res 50:3-22. <https://doi.org/10.3356/rapt-50-01-3-22.1>

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

Witzenberger KA, Hochkirch A (2013) Evaluating *ex situ* conservation projects: Genetic structure of the captive population of the Arabian sand cat. Mamm Biol 78:379-382 <https://doi.org/10.1016/j.mambio.2013.03.001>

8. Apêndices

Apêndice 1. Instituições doadoras das amostras de sangue dos gaviões-reais e as respectivas informações de cada indivíduo amostrado como a identificação individual, origem, nascimento e o sexo pelo histórico e pela sexagem molecular.

Amostras	Código	Microchip	Instituição doadora	Local de Origem - Estado	Nascimento	Sexo (Histórico)	Sexo (DNA)
HZ01	26439	977.200.001.259.029	Fundação Parque Zoológico de São Paulo	Rondônia	Natureza	Macho	Fêmea
HZ02	16166	977.200.000.676.341	Fundação Parque Zoológico de São Paulo	Britânia - Goiás	Natureza	Fêmea	Fêmea
HZ03	20189	AVID 069.862.570	Fundação Parque Zoológico de São Paulo	Desconhecida - Pará*	Natureza	Fêmea	Fêmea
HZ04	1725IB	963.008.000.141.725	RVB da Itaipu Binacional	Bahia	Natureza	Fêmea	Fêmea
HZ05	339610	963.000.000.339.610	RVB da Itaipu Binacional	Pará	Natureza	Macho	Fêmea
HZ06	211482	977.200.001.211.482	RVB da Itaipu Binacional	Pará	Natureza	Macho	Fêmea
HZ07	602	963.008.000.141.608	RVB da Itaipu Binacional	Tucuruí - Pará	Natureza	Fêmea	Fêmea
HZ08	594	963.008.000.141.620	RVB da Itaipu Binacional	Tucuruí - Pará	Natureza	Fêmea	Fêmea
HZ09	1929	963.000.000.001.929	Parque das Aves	Pará	Natureza	Macho	Fêmea
HZ10	4567IB	953.000.000.094.567	RVB da Itaipu Binacional	Pará	Natureza	Macho	Fêmea
HZ11	2828	977.200.001.170.386	Parque das Aves	Desconhecida	Natureza	Fêmea	Fêmea
HZ12	8831	900.115.000.378.831	RVB da Itaipu Binacional	Desconhecida	Natureza	Fêmea	Fêmea
HZ13	141710IB	963.008.000.141.710	RVB da Itaipu Binacional	Paraná	Natureza	Macho	Macho
HZ14	08932	900.032.001.008.932	Zoológico Municipal de Curitiba	Matupá - Mato Grosso	Natureza	Fêmea	Fêmea
HZ15	12919	985.121.005.412.919	Zoológico Municipal de Curitiba	CASIB	Cativeiro	Fêmea	Fêmea
HZ16	41678	963.008.000.141.678	Zoológico Municipal de Curitiba	CASIB	Cativeiro	Macho	Macho
HZ17	17108	977.200.000.701.102	Fundação Parque Zoológico de São Paulo	Desconhecida - Pará*	Natureza	Macho	Macho

HZ18	6956	977.200.004.892.688	Zoológico de Brasília	Desconhecida - Pará*	Natureza	Macho	Macho
HZ19	6916	977.200.004.928.177	Zoológico de Brasília	Salvador - Bahia	Natureza	Fêmea	Fêmea
HZ20	AV0728	963.000.000.195.035	Zoológico de Belo Horizonte	Guajará Mirim - Rondônia	Natureza	Macho	Macho
HZ21	AV0005	-	Zoológico de Belo Horizonte	Desconhecida	Natureza	Macho	Macho
HZ22	AV0007	963.000.000.167.101	Zoológico de Belo Horizonte	Rondônia	Natureza	Fêmea	Fêmea
HZ23	-	963.008.000.038.298	Hayabusa Falcoaria e Consultoria Ambiental	Santarém - Pará	Natureza	Macho	Macho
HZ24	-	963.008.000.033.766	Hayabusa Falcoaria e Consultoria Ambiental	Pará	Natureza	Fêmea	Fêmea
HZ25	HHF AP	-	Hayabusa Falcoaria e Consultoria Ambiental	Laranjal do Jari - Amapá	Natureza	Fêmea	Fêmea

*, provável estado de origem; RVB, Refúgio Biológico Bela Vista; CASIB, Criadouro de Animais Silvestres da Itaipu Binacional.

Apêndice 2. Combinação gradativa dos 10 *loci* em relação a probabilidade de identidade genética (I) e a probabilidade de exclusão de paternidade (Q).

Índices	1	1+2	1+2+3	1+2+3+4	1+2+3+4+5	1+2+3+4+5+6	1+2+3+4+5+6+8	1+2+3+4+5+6+8+9	1+2+3+4+5+6+8+9+12	1+2+3+4+5+6+8+9+12+18
Q	1.7E-01	7.8E-01	8.9E-01	9.0E-01	9.4E-01	9.7E-01	9.8E-01	9.9E-01	9.9E-01	9.9E-01
I	6.3E-01	5.9E-02	1.3E-02	1.2E-02	4.1E-03	7.3E-04	2.8E-04	9.9E-05	4.8E-05	2.7E-05

Apêndice 3. Combinação gradativa dos quatro *loci* em relação a probabilidade de identidade genética (I) e a probabilidade de exclusão de paternidade (Q).

Índices	2	2+3	2+3+5	2+3+5+6
Q	7.3E-01	8.7E-01	9.3E-01	9.7E-01
I	9.3E-02	2.0E-02	7.0E-03	1.3E-03

Apêndice 4. Matriz de coeficiente de parentesco (r) e suas respectivas relações com maior probabilidade entre pares dos 25 indivíduos de gavião-real mantidos em cativeiro amostrados neste estudo.

	HZ01	HZ02	HZ03	HZ04	HZ05	HZ06	HZ07	HZ08	HZ09	HZ10	HZ11	HZ12	HZ13	HZ14	HZ15	HZ16	HZ17	HZ18	HZ19	HZ20	HZ21	HZ22	HZ23	HZ24	HZ25
HZ01	-	U	HS	U	U	U	U	U	U	U	HS	U	U	U	U	U	U	U	U	U	FS	U	U	U	HS
HZ02	0.16	-	U	U	U	U	U	U	U	U	U	U	U	HS	U	U	U	U	U	U	U	U	U	HS	U
HZ03	0.31	0.01	-	U	HS	U	U	U	FS	U	PO	U	U	U	U	U	U	U	PO	U	U	U	U	U	U
HZ04	0.17	0.00	0.00	-	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U
HZ05	0.33	0.00	0.19	0.00	-	U	HS	U	PO	U	U	FS	PO	U	HS	U	U	U	FS	U	U	U	U	U	U
HZ06	0.00	0.03	0.00	0.00	0.00	-	U	FS	U	U	U	FS	U	PO	PO	U	U	PO	FS	FS	U	U	FS	FS	U
HZ07	0.04	0.02	0.00	0.00	0.19	0.00	-	U	U	HS	U	HS	PO	U	FS	HS	U	U	U	U	U	U	U	U	U
HZ08	0.04	0.09	0.33	0.00	0.00	0.57	0.00	-	U	HS	HS	U	U	U	U	U	U	U	U	U	U	U	FS	FS	HS
HZ09	0.12	0.00	0.38	0.08	0.39	0.00	0.00	0.05	-	U	U	U	U	U	U	U	U	U	HS	U	U	U	U	U	U
HZ10	0.00	0.00	0.00	0.00	0.00	0.23	0.19	0.27	0.00	-	HS	U	U	U	U	U	U	U	U	U	U	U	HS	U	U
HZ11	0.21	0.00	0.54	0.13	0.00	0.12	0.02	0.3	0.12	0.17	-	U	U	U	U	U	U	U	U	U	HS	U	U	U	U
HZ12	0.00	0.00	0.00	0.00	0.55	0.44	0.28	0.11	0.00	0.00	0.00	-	PO	U	PO	PO	U	U	U	U	U	U	PO	U	U
HZ13	0.00	0.00	0.00	0.00	0.50	0.24	0.5	0.00	0.07	0.00	0.00	0.50	-	U	PO	FS	U	U	U	U	U	HS	U	U	U
HZ14	0.06	0.20	0.00	0.00	0.00	0.57	0.00	0.11	0.00	0.13	0.00	0.00	0.00	-	HS	U	U	FS	U	FS	U	PO	U	PO	FS
HZ15	0.00	0.00	0.00	0.03	0.21	0.5	0.43	0.00	0.00	0.00	0.00	0.61	0.61	0.22	-	PO	U	HS	U	HS	U	U	PO	U	U
HZ16	0.00	0.00	0.00	0.00	0.10	0.08	0.25	0.00	0.00	0.00	0.00	0.50	0.76	0.22	0.70	-	U	U	U	U	U	HS	U	HS	U
HZ17	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.10	0.00	0.00	0.00	0.00	0.00	0.45	0.00	0.00	-	U	U	U	U	U	U	U	FS
HZ18	0.09	0.00	0.04	0.00	0.09	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.21	0.48	0.19	0.19	0.18	-	U	U	U	U	U	U	U
HZ19	0.09	0.00	0.38	0.00	0.43	0.71	0.00	0.15	0.23	0.11	0.04	0.00	0.44	0.36	0.00	0.00	0.00	0.36	-	U	U	U	U	U	U
HZ20	0.00	0.17	0.00	0.00	0.00	0.43	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.72	0.37	0.18	0.00	0.25	0.26	-	U	FS	U	U	U
HZ21	0.32	0.12	0.13	0.08	0.20	0.00	0.00	0.00	0.06	0.00	0.20	0.13	0.00	0.00	0.00	0.00	0.12	0.18	0.00	0.00	-	U	U	U	U
HZ22	0.21	0.20	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.14	0.55	0.15	0.24	0.00	0.23	0.05	0.36	0.15	-	U	U	HS
HZ23	0.00	0.10	0.00	0.00	0.00	0.45	0.32	0.6	0.00	0.27	0.00	0.50	0.18	0.00	0.47	0.08	0.11	0.00	0.00	0.00	0.00	0.00	-	FS	HS
HZ24	0.00	0.23	0.12	0.00	0.00	0.76	0.09	0.62	0.00	0.11	0.05	0.17	0.00	0.50	0.11	0.25	0.00	0.00	0.29	0.09	0.00	0.00	0.54	-	U
HZ25	0.18	0.00	0.16	0.00	0.07	0.15	0.00	0.13	0.00	0.06	0.00	0.00	0.00	0.67	0.00	0.00	0.71	0.00	0.05	0.05	0.00	0.16	0.16	0.00	-

U, não relacionados; HS, meio-irmãos; FS, irmãos completos; PO, pais-filhos.

Apêndice 5. Lista com as estimativas de máxima verossimilhança do coeficiente de parentesco e relacionamento com maior probabilidade entre pares dos 25 indivíduos de gavião-real mantidos em cativeiro, obtidas por meio do ML-Relate.

Indivíduo 1	Indivíduo 2	r	R	LnL(R)	Delta Ln(L)				95% IC
					U	HS	FS	PO	
HZ01	HZ02	0.1596	U	-30.52	-	0.5	1.49	9999	U, HS
HZ01	HZ03	0.3079	HS	-27.25	0.73	-	0.04	9999	U, HS, FS
HZ01	HZ04	0.1705	U	-35.44	-	1.68	3.6	9999	U, HS
HZ01	HZ05	0.3255	U	-27.48	-	0.26	1.35	9999	U, HS
HZ01	HZ06	0	U	-19.29	-	2.2	5.04	9999	U, HS
HZ01	HZ07	0.0441	U	-27.43	-	1.07	3.66	9999	U, HS
HZ01	HZ08	0.0417	U	-24.45	-	0.8	1.6	9999	U, HS
HZ01	HZ09	0.1166	U	-29.79	-	2.79	6.22	9999	U, HS
HZ01	HZ10	0	U	-31.24	0.47	-	1.07	9999	U, HS
HZ01	HZ11	0.2066	HS	-30.27	-	1.99	5.29	9999	U, HS, FS
HZ01	HZ12	0	U	-21.39	-	1.6	3.4	9999	U, HS
HZ01	HZ13	0	U	-20.67	-	0.31	1.16	9999	U, HS, FS
HZ01	HZ14	0.0577	U	-16.99	-	1.02	2.37	9999	U, HS, FS
HZ01	HZ15	0	U	-22.59	-	1.98	4.54	9999	U, HS
HZ01	HZ16	0	U	-26.96	-	1.29	2.93	9999	U, HS
HZ01	HZ17	0	U	-34.65	-	0.42	3.14	9999	U, HS
HZ01	HZ18	0.0928	U	-26.45	-	1.71	3.67	9999	U, HS
HZ01	HZ19	0.0928	U	-23.88	-	1.16	3.37	9999	U, HS
HZ01	HZ20	0	U	-27.3	-	1.28	3.43	9999	U, HS
HZ01	HZ21	0.3164	FS	-29.11	0.31	-	1.7	9999	HS, FS
HZ01	HZ22	0.2053	U	-26.32	-	1.32	2.94	9999	U, HS
HZ01	HZ23	0	U	-20.73	-	1.49	3.38	9999	U
HZ01	HZ24	0	U	-24.4	-	0.17	1.04	9999	U
HZ01	HZ25	0.1826	HS	-23.13	-	0.07	0.28	9999	U, HS, FS
HZ02	HZ03	0.0102	U	-30.11	-	1.6	4.13	9999	U, HS
HZ02	HZ04	0	U	-37.57	-	2.25	5.25	9999	U
HZ02	HZ05	0	U	-29.61	1.93	1.09	-	0.67	U
HZ02	HZ06	0.0296	U	-21.64	-	1.18	2.82	9999	U, HS, FS
HZ02	HZ07	0.0166	U	-29.56	-	0.64	1.94	9999	U, HS
HZ02	HZ08	0.0916	U	-26.58	-	2.76	6.14	9999	U, HS, FS
HZ02	HZ09	0	U	-31.92	0.58	0.35	-	9999	U
HZ02	HZ10	0	U	-33.37	-	0.22	1.79	9999	U, HS
HZ02	HZ11	0	U	-32.49	1.1	0.03	0.69	-	U, HS
HZ02	HZ12	0	U	-22.13	-	1.26	2.84	9999	U, HS
HZ02	HZ13	0	U	-20.47	-	1.15	3.56	9999	U, HS
HZ02	HZ14	0.2023	HS	-17.87	-	1.15	2.82	9999	U, HS, FS
HZ02	HZ15	0	U	-24.72	-	1.89	4.54	9999	U, HS
HZ02	HZ16	0	U	-29.09	-	1.18	2.8	9999	U
HZ02	HZ17	0	U	-36.77	-	0.51	1.87	1.73	U, HS
HZ02	HZ18	0	U	-28.58	-	2	5.29	9999	U, HS
HZ02	HZ19	0	U	-26.01	-	2.65	5.64	9999	U, HS
HZ02	HZ20	0.17	U	-29.43	-	0.88	1.65	9999	U, HS
HZ02	HZ21	0.1207	U	-31.93	0.71	-	0.29	9999	U, HS
HZ02	HZ22	0.1969	U	-28.67	0.09	-	0.2	9999	U, HS, FS
HZ02	HZ23	0.0959	U	-21.47	-	1.96	4.49	9999	U, HS, FS
HZ02	HZ24	0.2262	HS	-26.4	0.09	-	0.84	9999	U, HS, FS
HZ02	HZ25	0	U	-23.09	-	1.75	4.46	9999	U, HS, FS
HZ03	HZ04	0	U	-35.03	5.42	2.05	1.05	-	U, HS
HZ03	HZ05	0.1876	HS	-26.6	-	1.45	3.58	9999	U, HS, FS
HZ03	HZ06	0	U	-18.88	-	2.25	5.78	9999	U, HS
HZ03	HZ07	0	U	-27.02	-	1.17	2.33	9999	U, HS
HZ03	HZ08	0.3281	U	-24.04	-	0.72	3.22	9999	U, HS, FS
HZ03	HZ09	0.3815	FS	-28.8	0.24	-	0.49	9999	HS, FS
HZ03	HZ10	0	U	-30.83	-	0.08	1.77	9999	U, HS, FS

HZ03	HZ11	0.5373	PO	-24.53	0.2	-	1.32	9999	HS, FS, PO
HZ03	HZ12	0	U	-20.97	-	1.81	3.95	9999	U, HS
HZ03	HZ13	0	U	-20.26	-	1.03	2.87	9999	U, HS
HZ03	HZ14	0	U	-16.58	-	2.12	5.1	9999	U, HS
HZ03	HZ15	0	U	-22.17	-	1.2	3.4	9999	U, HS
HZ03	HZ16	0	U	-26.55	1.74	0.78	-	9999	U
HZ03	HZ17	0	U	-34.23	0.29	0.2	-	9999	U
HZ03	HZ18	0.041	U	-26.04	0.87	-	0.71	9999	U, HS
HZ03	HZ19	0.375	PO	-22.67	-	0.6	1.65	9999	U, HS, FS, PO
HZ03	HZ20	0	U	-26.89	-	0.41	1.94	9999	U, HS
HZ03	HZ21	0.1294	U	-29.39	-	1.98	4.38	9999	U, HS
HZ03	HZ22	0	U	-25.91	-	3.23	6.93	9999	U
HZ03	HZ23	0	U	-20.31	-	0.65	2.02	9999	U, HS
HZ03	HZ24	0.1233	U	-23.99	-	1.91	4.55	9999	U, HS, FS
HZ03	HZ25	0.1554	U	-22.89	-	1.34	3.25	9999	U, HS
HZ04	HZ05	0	U	-34.53	-	1.71	3.82	9999	U, HS
HZ04	HZ06	0	U	-26.56	2.01	0.85	0.7	-	U, HS
HZ04	HZ07	0	U	-34.48	-	0.3	0.75	9999	U, HS
HZ04	HZ08	0	U	-31.5	1.93	0.72	0.5	-	U, HS
HZ04	HZ09	0.0791	U	-36.84	-	1.26	3.07	9999	U, HS, FS
HZ04	HZ10	0	U	-38.29	-	0.5	1.38	9999	U, HS
HZ04	HZ11	0.1315	U	-37.41	-	1.77	3.97	9999	U, HS
HZ04	HZ12	0	U	-27.05	-	1.26	3.6	9999	U, HS
HZ04	HZ13	0	U	-27.72	1.85	0.82	1.07	-	U, HS
HZ04	HZ14	0	U	-22.88	-	0.23	1.15	9999	U, HS
HZ04	HZ15	0.0263	U	-29.64	0.09	-	0.45	9999	U, HS
HZ04	HZ16	0	U	-34.01	-	1.18	2.72	9999	U
HZ04	HZ17	0	U	-41.7	-	1.79	4	9999	U
HZ04	HZ18	0	U	-33.5	-	1.46	3.27	9999	U
HZ04	HZ19	0	U	-30.93	1.12	0.5	0.19	-	U, HS
HZ04	HZ20	0	U	-34.35	-	1.82	4.14	9999	U, HS
HZ04	HZ21	0.083	U	-36.85	-	0.01	0.64	0.27	U, HS
HZ04	HZ22	0	U	-33.6	-	1.43	3.19	9999	U
HZ04	HZ23	0	U	-26.39	-	1.05	2	9999	U, HS
HZ04	HZ24	0	U	-31.45	-	1.67	3.53	9999	U
HZ04	HZ25	0	U	-30.35	-	0.6	1.96	9999	U
HZ05	HZ06	0	U	-18.38	-	0.8	1.94	9999	U, HS
HZ05	HZ07	0.1873	HS	-26.21	-	1.53	3.7	9999	U, HS, FS
HZ05	HZ08	0	U	-23.54	-	1.71	4.08	9999	U
HZ05	HZ09	0.3883	PO	-27.78	-	1.83	4.85	9999	U, HS, FS, PO
HZ05	HZ10	0	U	-30.33	-	1.52	3.36	9999	U
HZ05	HZ11	0	U	-29.45	0.28	-	0.65	9999	U
HZ05	HZ12	0.5492	FS	-18.73	0.53	0.16	0.52	-	HS, FS
HZ05	HZ13	0.5	PO	-16.37	1.06	0.17	-	9999	U, HS, FS, PO
HZ05	HZ14	0	U	-16.08	-	0.33	1.93	0.98	U, HS
HZ05	HZ15	0.2061	HS	-21.4	-	0.99	2.98	9999	U, HS, FS
HZ05	HZ16	0.0986	U	-26.05	-	1.7	4.13	9999	U, HS, FS
HZ05	HZ17	0	U	-33.73	-	2.61	5.83	9999	U
HZ05	HZ18	0.0934	U	-25.54	3.04	1.34	0.55	-	U, HS
HZ05	HZ19	0.4321	FS	-20.89	1.93	0.88	0.25	-	HS, FS
HZ05	HZ20	0	U	-26.39	0.32	-	0.19	0.08	U
HZ05	HZ21	0.2027	U	-28.89	-	1.91	4.85	9999	U, HS
HZ05	HZ22	0	U	-25.41	-	2.6	5.94	9999	U
HZ05	HZ23	0	U	-19.81	-	2.6	6.08	9999	U, HS
HZ05	HZ24	0	U	-23.49	-	2.41	5.21	9999	U, HS
HZ05	HZ25	0.0654	U	-21	-	0.11	1.53	9999	U, HS
HZ06	HZ07	0	U	-18.55	-	0.14	0.82	9999	U, HS
HZ06	HZ08	0.5653	FS	-13.06	0.51	-	0.47	9999	HS, FS, PO
HZ06	HZ09	0	U	-20.68	-	1.1	3.15	9999	U, HS
HZ06	HZ10	0.2271	U	-21.78	-	1.76	4.21	9999	U, HS, FS
HZ06	HZ11	0.1195	U	-19.64	-	2.47	5.36	9999	U, HS

HZ06	HZ12	0.4363	FS	-13.63	-	3.12	6.67	9999	U, HS, FS
HZ06	HZ13	0.238	U	-12.07	2.15	0.96	1.45	-	U, HS, FS
HZ06	HZ14	0.5655	PO	-10.89	4.04	2.29	-	1.51	U, HS, FS, PO
HZ06	HZ15	0.5	PO	-13.17	-	0.18	0.71	9999	U, HS, FS, PO
HZ06	HZ16	0.0755	U	-16.24	2.92	1.25	0.01	-	U, HS, FS
HZ06	HZ17	0.0569	U	-23.22	-	2.1	4.95	9999	U, HS, FS
HZ06	HZ18	0.5	PO	-16.63	-	0.81	2.71	9999	U, HS, FS, PO
HZ06	HZ19	0.7137	FS	-13.96	-	2.74	5.75	9999	HS, FS
HZ06	HZ20	0.4312	FS	-18.07	-	3.66	7.8	9999	U, HS, FS
HZ06	HZ21	0	U	-20.7	-	3.33	7.08	9999	U, HS
HZ06	HZ22	0.0498	U	-19.92	-	0.22	1	9999	U, HS, FS
HZ06	HZ23	0.4461	FS	-11.58	-	3.47	7.61	9999	U, HS, FS, PO
HZ06	HZ24	0.7584	FS	-11	-	0.12	1.97	9999	FS
HZ06	HZ25	0.1524	U	-12.37	-	3.3	7	9999	U, HS, FS, PO
HZ07	HZ08	0	U	-23.49	-	1.49	3.82	9999	U, HS
HZ07	HZ09	0	U	-28.83	-	3	6.23	9999	U, HS
HZ07	HZ10	0.1911	HS	-29.57	-	1.95	4.75	9999	U, HS, FS
HZ07	HZ11	0.0165	U	-29.4	-	2.45	5.41	9999	U, HS
HZ07	HZ12	0.2804	HS	-20.97	-	0.63	0.83	9999	U, HS, FS
HZ07	HZ13	0.5016	PO	-16.68	-	2.64	5.96	9999	U, HS, FS, PO
HZ07	HZ14	0	U	-17.67	-	3.14	6.8	9999	U, HS
HZ07	HZ15	0.4281	FS	-20.57	-	1.8	3.39	9999	U, HS, FS
HZ07	HZ16	0.2523	HS	-25.49	-	1.9	4.16	9999	U, HS, FS
HZ07	HZ17	0	U	-33.68	-	1.33	2.86	9999	U
HZ07	HZ18	0	U	-25.49	-	2.6	6.09	9999	U, HS
HZ07	HZ19	0	U	-22.92	-	1.5	2.91	9999	U, HS
HZ07	HZ20	0	U	-26.34	0.71	0.26	0.27	-	U, HS
HZ07	HZ21	0	U	-28.84	-	1.72	4.16	9999	U, HS
HZ07	HZ22	0	U	-25.58	-	0.51	1.88	1.49	U
HZ07	HZ23	0.3234	U	-21.18	-	1.58	3.33	9999	U, HS, FS
HZ07	HZ24	0.0874	U	-23.44	-	1.77	3.58	9999	U, HS
HZ07	HZ25	0	U	-21.24	-	1.37	3.72	9999	U
HZ08	HZ09	0.0479	U	-25.84	-	0.53	1.78	9999	U, HS
HZ08	HZ10	0.2694	HS	-27.21	-	0.4	0.91	9999	U, HS, FS
HZ08	HZ11	0.299	HS	-26.18	1.4	0.66	-	0.28	U, HS, FS
HZ08	HZ12	0.1069	U	-18.86	0.2	-	0.43	0.49	U, HS, FS
HZ08	HZ13	0	U	-17.23	-	0.35	1.12	9999	U, HS
HZ08	HZ14	0.1107	U	-14.11	-	1.12	2.6	9999	U, HS, FS, PO
HZ08	HZ15	0	U	-18.64	-	1.36	2.71	9999	U, HS, FS, PO
HZ08	HZ16	0	U	-23.01	-	1.34	2.99	9999	U, HS
HZ08	HZ17	0.0993	U	-30.7	0.8	0	0.32	-	U, HS
HZ08	HZ18	0	U	-22.51	-	2.04	4.91	9999	U, HS, FS, PO
HZ08	HZ19	0.1482	U	-19.94	2.08	0.97	-	9999	U, HS, FS
HZ08	HZ20	0	U	-23.36	0.82	0.55	-	9999	U, HS, FS
HZ08	HZ21	0	U	-25.86	-	1.38	3.32	9999	U
HZ08	HZ22	0	U	-22.02	-	0.7	1.51	9999	U, HS
HZ08	HZ23	0.6039	FS	-15.63	0.55	-	0.55	9999	HS, FS, PO
HZ08	HZ24	0.6247	FS	-18.04	-	1.21	2.41	9999	FS
HZ08	HZ25	0.1348	HS	-19.8	-	0.7	2.03	9999	U, HS, FS, PO
HZ09	HZ10	0	U	-32.64	-	1.17	2.76	9999	U, HS
HZ09	HZ11	0.1166	U	-31.75	-	0.06	0.06	9999	U, HS, FS
HZ09	HZ12	0	U	-21.39	-	0.23	0.38	9999	U, HS, FS
HZ09	HZ13	0.0689	U	-22.06	-	0.55	1.67	9999	U, HS, FS
HZ09	HZ14	0	U	-17	-	1.32	2.89	9999	U, HS
HZ09	HZ15	0	U	-23.98	-	1.88	3.84	9999	U, HS
HZ09	HZ16	0	U	-28.35	-	0.16	0.18	9999	U, HS
HZ09	HZ17	0	U	-36.04	-	1.66	3.77	9999	U
HZ09	HZ18	0	U	-27.85	-	0.96	1.97	9999	U, HS
HZ09	HZ19	0.2256	HS	-24.73	-	1.97	4.52	9999	U, HS, FS
HZ09	HZ20	0	U	-28.7	-	1.65	3.89	9999	U
HZ09	HZ21	0.0568	U	-31.2	-	2.24	5.07	9999	U, HS, FS

HZ09	HZ22	0	U	-27.72	0.35	0.12	-	9999	U
HZ09	HZ23	0	U	-20.73	-	1.88	4.08	9999	U, HS
HZ09	HZ24	0	U	-25.8	-	0.46	1.59	9999	U, HS
HZ09	HZ25	0	U	-24.69	-	2.21	4.99	9999	U, HS
HZ10	HZ11	0.1675	HS	-33.01	-	1.83	3.8	9999	U, HS, FS
HZ10	HZ12	0	U	-22.85	-	2.75	5.49	9999	U, HS
HZ10	HZ13	0	U	-22.42	-	0.82	2.21	9999	U, HS
HZ10	HZ14	0.1322	U	-18.1	-	1.25	2.89	9999	U, HS
HZ10	HZ15	0	U	-25.44	1.36	0.82	-	9999	U, HS
HZ10	HZ16	0	U	-29.81	0.26	-	0.06	9999	U
HZ10	HZ17	0	U	-37.5	-	0.63	1.66	9999	U, HS
HZ10	HZ18	0	U	-29.3	-	2.42	4.73	9999	U, HS
HZ10	HZ19	0.1113	U	-26.73	-	0.04	0.39	9999	U, HS
HZ10	HZ20	0	U	-30.15	-	0.68	1.33	9999	U, HS
HZ10	HZ21	0	U	-32.65	0.7	0.19	-	9999	U
HZ10	HZ22	0	U	-28.82	-	1.05	2.72	9999	U
HZ10	HZ23	0.2685	HS	-22.03	-	0.47	2.07	9999	U, HS, FS
HZ10	HZ24	0.1115	U	-27.25	-	0.39	2.07	9999	U, HS, FS
HZ10	HZ25	0.0574	U	-25.05	-	0.9	1.96	9999	U, HS, FS
HZ11	HZ12	0	U	-24.77	-	2.16	4.63	9999	U
HZ11	HZ13	0	U	-22.64	-	1.24	4.34	9999	U, HS
HZ11	HZ14	0	U	-18.76	-	3.12	6.96	9999	U, HS
HZ11	HZ15	0	U	-24.55	-	0.31	1.89	9999	U
HZ11	HZ16	0	U	-28.92	-	2.45	5.77	9999	U
HZ11	HZ17	0	U	-36.61	0.01	-	1.37	9999	U
HZ11	HZ18	0	U	-28.42	-	0.79	2.15	9999	U, HS
HZ11	HZ19	0.0366	U	-25.85	-	0.99	2.66	9999	U, HS, FS
HZ11	HZ20	0	U	-29.27	-	0.79	2.38	9999	U
HZ11	HZ21	0.2005	HS	-31.76	-	1.48	4.26	9999	U, HS
HZ11	HZ22	0	U	-26.68	-	2.25	5.49	9999	U
HZ11	HZ23	0	U	-24.11	-	1.02	2.99	9999	U, HS
HZ11	HZ24	0.0534	U	-26.37	-	1.15	2.7	9999	U, HS, FS
HZ11	HZ25	0	U	-25.26	-	1.92	3.94	9999	U, HS
HZ12	HZ13	0.5	PO	-13.73	-	2.22	5	9999	U, HS, FS, PO
HZ12	HZ14	0	U	-14.19	-	0.42	1.08	9999	U, HS, FS
HZ12	HZ15	0.6057	PO	-13.96	-	0.22	0.8	9999	HS, FS, PO
HZ12	HZ16	0.5	PO	-19.21	-	2.44	5.3	9999	HS, FS, PO
HZ12	HZ17	0	U	-27.35	-	2.62	6.15	9999	U, HS
HZ12	HZ18	0	U	-19.15	-	2.71	5.81	9999	U, HS, FS
HZ12	HZ19	0	U	-18.29	-	0.41	1.26	9999	U, HS
HZ12	HZ20	0	U	-21.71	-	2.52	5.49	9999	U, HS
HZ12	HZ21	0.1312	U	-22.51	-	1.05	2.83	9999	U, HS
HZ12	HZ22	0	U	-20.95	-	2.8	6.23	9999	U, HS
HZ12	HZ23	0.5	PO	-16.57	-	2.41	5.04	9999	U, HS, FS, PO
HZ12	HZ24	0.1747	U	-18.81	-	2.59	5.26	9999	U, HS, FS
HZ12	HZ25	0	U	-16.79	-	1.67	3.96	9999	U, HS
HZ13	HZ14	0	U	-11.38	0.14	-	1.2	9999	U, HS, FS
HZ13	HZ15	0.6141	PO	-13.63	1.58	0.51	0.13	-	HS, FS, PO
HZ13	HZ16	0.7647	FS	-15.89	-	0.6	1.55	9999	FS
HZ13	HZ17	0	U	-24.59	0.76	-	0.59	9999	U
HZ13	HZ18	0.2093	U	-19.42	-	0.88	2.47	9999	U, HS, FS
HZ13	HZ19	0.4448	U	-14.77	-	0.51	1.13	9999	U, HS, FS
HZ13	HZ20	0	U	-20.28	-	1.27	2.57	9999	U, HS
HZ13	HZ21	0	U	-22.08	0.05	0.01	-	9999	U, HS
HZ13	HZ22	0.1447	HS	-19.15	-	0.98	2.36	9999	U, HS, FS
HZ13	HZ23	0.1843	U	-14.92	-	2.49	5.7	9999	U, HS, FS
HZ13	HZ24	0	U	-17.18	-	0.1	1.01	9999	U, HS
HZ13	HZ25	0	U	-20.95	-	1.07	2.61	9999	U, HS
HZ14	HZ15	0.2151	HS	-12.5	-	1.71	3.8	9999	U, HS, FS, PO
HZ14	HZ16	0.2152	U	-15.36	-	0.69	2.37	9999	U, HS, FS
HZ14	HZ17	0.4549	U	-20.63	1.1	0.53	-	0.32	U, HS, FS

HZ14	HZ18	0.4787	FS	-13.36	-	0.55	0.99	9999	U, HS, FS, PO
HZ14	HZ19	0.3566	U	-13.9	2.57	1.39	-	0.78	U, HS, FS
HZ14	HZ20	0.717	FS	-16.18	-	1.06	2.8	9999	FS
HZ14	HZ21	0	U	-18.11	0.16	-	0.25	9999	U, HS
HZ14	HZ22	0.552	PO	-17.46	-	1.1	3.42	9999	U, HS, FS, PO
HZ14	HZ23	0	U	-12.95	1.13	0.36	0.45	-	U, HS, FS, PO
HZ14	HZ24	0.5	PO	-12.38	-	0.64	1.24	9999	U, HS, FS, PO
HZ14	HZ25	0.6737	FS	-8.94	-	0.12	1.15	0.45	U, HS, FS, PO
HZ15	HZ16	0.6977	PO	-18.23	0.38	0.08	0.48	-	HS, FS, PO
HZ15	HZ17	0	U	-28.84	-	0.47	1.32	9999	U
HZ15	HZ18	0.1872	HS	-20.44	-	0.05	1.95	9999	U, HS, FS, PO
HZ15	HZ19	0	U	-18.08	-	0.44	1.86	1.27	U, HS, FS
HZ15	HZ20	0.3698	HS	-21.24	-	0.96	2.33	9999	U, HS, FS
HZ15	HZ21	0	U	-24	-	0.73	2.41	9999	U, HS
HZ15	HZ22	0.1544	U	-20.74	-	1.56	4.08	9999	U, HS, FS
HZ15	HZ23	0.4697	PO	-15.95	-	1.31	3.65	9999	U, HS, FS, PO
HZ15	HZ24	0.1066	U	-18.6	-	2.25	4.97	9999	U, HS, FS, PO
HZ15	HZ25	0	U	-18.19	0.14	-	0.41	9999	U, HS
HZ16	HZ17	0	U	-33.21	-	0.84	1.87	9999	U
HZ16	HZ18	0.1873	U	-25.01	-	2.36	5.72	9999	U, HS, FS
HZ16	HZ19	0	U	-22.45	-	1.94	4.49	9999	U, HS
HZ16	HZ20	0.1749	U	-25.87	2.68	1.54	-	0.74	U, HS, FS
HZ16	HZ21	0	U	-28.37	-	0.87	2.06	9999	U
HZ16	HZ22	0.2438	HS	-22.52	2.42	1.41	-	0.99	U, HS, FS
HZ16	HZ23	0.0847	U	-20.71	-	1.92	4.41	9999	U, HS, FS
HZ16	HZ24	0.2533	HS	-22.19	-	0.34	1.08	9999	U, HS, FS
HZ16	HZ25	0	U	-22.56	-	0.53	1.84	9999	U, HS
HZ17	HZ18	0.183	U	-32.7	-	0.3	0.89	9999	U, HS
HZ17	HZ19	0	U	-30.14	-	0.95	2.31	9999	U, HS
HZ17	HZ20	0	U	-33.55	0.43	0.13	0.3	-	U
HZ17	HZ21	0.1245	U	-36.06	-	0.02	1.17	0.29	U, HS
HZ17	HZ22	0	U	-30.25	0.78	-	0.33	9999	U, HS
HZ17	HZ23	0.1129	U	-26.69	-	1.23	3.2	9999	U, HS
HZ17	HZ24	0	U	-30.66	-	0.2	1.12	0.8	U, HS
HZ17	HZ25	0.7148	FS	-22.08	-	0.39	0.75	9999	FS, PO
HZ18	HZ19	0.3601	U	-21.94	-	0.15	0.83	9999	U, HS, FS
HZ18	HZ20	0.2469	U	-25.36	-	2.81	6.2	9999	U, HS, FS
HZ18	HZ21	0.1811	U	-27.86	-	0.74	2.07	9999	U, HS
HZ18	HZ22	0.2275	U	-24.38	1.6	0.85	-	0.64	U, HS, FS
HZ18	HZ23	0	U	-18.5	0.17	-	0.8	9999	U, HS, FS, PO
HZ18	HZ24	0	U	-22.46	-	0.63	2.01	9999	U, HS, FS, PO
HZ18	HZ25	0	U	-22.05	-	0.67	1.74	9999	U, HS
HZ19	HZ20	0.2608	U	-22.79	-	2.51	5.65	9999	U, HS, FS
HZ19	HZ21	0	U	-25.29	-	1.06	2.76	9999	U, HS
HZ19	HZ22	0.0498	U	-21.81	-	0.04	0.3	0.28	U, HS
HZ19	HZ23	0	U	-17.63	-	2.32	5.45	9999	U, HS
HZ19	HZ24	0.2937	U	-19.89	0.06	-	1.33	0.58	U, HS, FS
HZ19	HZ25	0.0463	U	-17.4	-	2.15	4.84	9999	U, HS, FS
HZ20	HZ21	0	U	-28.71	-	0.34	1.66	9999	U
HZ20	HZ22	0.3641	FS	-25.4	-	2.05	4.38	9999	U, HS, FS
HZ20	HZ23	0	U	-21.05	-	1.77	4.05	9999	U, HS
HZ20	HZ24	0.0922	U	-23.31	-	1.99	4.64	9999	U, HS, FS
HZ20	HZ25	0.0462	U	-22.9	1.33	0.81	-	0.47	U, HS
HZ21	HZ22	0.1481	U	-27.73	-	1.48	3.8	9999	U, HS
HZ21	HZ23	0	U	-21.85	-	1.99	4.64	9999	U, HS
HZ21	HZ24	0	U	-25.81	5.14	1.96	-	0.88	U
HZ21	HZ25	0	U	-24.71	-	1.29	2.86	9999	U, HS
HZ22	HZ23	0	U	-19.72	-	0.73	1.69	9999	U, HS
HZ22	HZ24	0	U	-20.72	-	1.27	2.57	9999	U, HS
HZ22	HZ25	0.1574	HS	-19.53	-	1.4	3.57	9999	U, HS, FS, PO
HZ23	HZ24	0.5392	FS	-16.55	0.07	-	0.7	0.58	HS, FS, PO

HZ23	HZ25	0.164	HS	-16	0.13	-	1.37	0.43	U, HS, FS, PO
HZ24	HZ25	0	U	-19.81	-	1.06	2.5	9999	U, HS

r, coeficiente de parentesco; R, relacionamento com maior probabilidade; LnL, verossimilhança logarítmica de R; Delta Ln(L), probabilidades de log para os outros relacionamentos; 95% IC, relacionamentos alternativos no nível de significância de 0,05; U, não relacionados; HS, meio-irmãos; FS, irmãos completos; PO, pais-filhos; 9999, relação não possível.

Apêndice 6. Teste de razão de verossimilhança realizado no ML-Related para os pares relacionados e com ambiguidade em seu conjunto de confiança.

Ind 1	Ind 2	r	R	LnL(R)	Delta Ln(L)				95% IC	Teste p-valor			
					U	HS	FS	PO		U	HS	FS	PO
HZ01	HZ03	0.3079	HS	-27.25	0.73	-	0.04	9999	U, HS, FS	0.100	-	0.279	-
HZ01	HZ11	0.2066	HS	-30.27	-	1.99	5.29	9999	U, HS, FS	0.244	-	0.103	-
HZ01	HZ21	0.3164	FS	-29.11	0.31	-	1.7	9999	HS, FS	-	0.210	-	-
HZ01	HZ25	0.1826	HS	-23.13	-	0.07	0.28	9999	U, HS, FS	0.221	-	0.110	-
HZ02	HZ14	0.2023	HS	-17.87	-	1.15	2.82	9999	U, HS, FS	0.245	-	0.178	-
HZ02	HZ24	0.2262	HS	-26.4	0.09	-	0.84	9999	U, HS, FS	0.228	-	0.184	-
HZ03	HZ05	0.1876	HS	-26.6	-	1.45	3.58	9999	U, HS, FS	0.146	-	0.072	-
HZ03	HZ09	0.3815	FS	-28.8	0.24	-	0.49	9999	HS, FS	-	0.173	-	-
HZ03	HZ11	0.5373	PO	-24.53	0.2	-	1.32	9999	HS, FS, PO	-	0.005	0.068	-
HZ03	HZ19	0.375	PO	-22.67	-	0.6	1.65	9999	U, HS, FS, PO	0.076	0.366	0.259	-
HZ05	HZ07	0.1873	HS	-26.21	-	1.53	3.7	9999	U, HS, FS	0.185	-	0.021	-
HZ05	HZ09	0.3883	PO	-27.78	-	1.83	4.85	9999	U, HS, FS, PO	0.059	0.355	0.146	-
HZ05	HZ12	0.5492	FS	-18.73	0.53	0.16	0.52	-	HS, FS	-	0.209	-	-
HZ05	HZ13	0.5	PO	-16.37	1.06	0.17	-	9999	U, HS, FS, PO	0.025	0.135	0.141	-
HZ05	HZ15	0.2061	HS	-21.4	-	0.99	2.98	9999	U, HS, FS	0.193	-	0.136	-
HZ05	HZ19	0.4321	FS	-20.89	1.93	0.88	0.25	-	HS, FS	-	0.074	-	-
HZ06	HZ08	0.5653	FS	-13.06	0.51	-	0.47	9999	HS, FS, PO	-	0.059	-	0.049
HZ06	HZ12	0.4363	FS	-13.63	-	3.12	6.67	9999	U, HS, FS	0.104	0.210	-	-
HZ06	HZ14	0.5655	PO	-10.89	4.04	2.29	-	1.51	U, HS, FS, PO	0.059	0.231	0.305	-
HZ06	HZ15	0.5	PO	-13.17	-	0.18	0.71	9999	U, HS, FS, PO	0.092	0.320	0.193	-
HZ06	HZ18	0.5	PO	-16.63	-	0.81	2.71	9999	U, HS, FS, PO	0.081	0.294	0.279	-
HZ06	HZ19	0.7137	FS	-13.96	-	2.74	5.75	9999	HS, FS	-	0.134	-	-
HZ06	HZ20	0.4312	FS	-18.07	-	3.66	7.8	9999	U, HS, FS	0.098	0.229	-	-
HZ06	HZ23	0.4461	FS	-11.58	-	3.47	7.61	9999	U, HS, FS, PO	0.049	0.137	-	0.113
HZ06	HZ24	0.7584	FS	-11	-	0.12	1.97	9999	FS	-	-	-	-
HZ07	HZ10	0.1911	HS	-29.57	-	1.95	4.75	9999	U, HS, FS	0.244	-	0.214	-
HZ07	HZ12	0.2804	HS	-20.97	-	0.63	0.83	9999	U, HS, FS	0.082	-	0.124	-
HZ07	HZ13	0.5016	PO	-16.68	-	2.64	5.96	9999	U, HS, FS, PO	0.027	0.167	0.203	-
HZ07	HZ15	0.4281	FS	-20.57	-	1.8	3.39	9999	U, HS, FS	0.050	0.216	-	-
HZ07	HZ16	0.2523	HS	-25.49	-	1.9	4.16	9999	U, HS, FS	0.141	-	0.174	-
HZ08	HZ10	0.2694	HS	-27.21	-	0.4	0.91	9999	U, HS, FS	0.244	-	0.235	-
HZ08	HZ11	0.299	HS	-26.18	1.4	0.66	-	0.28	U, HS, FS	0.203	-	0.170	-
HZ08	HZ23	0.6039	FS	-15.63	0.55	-	0.55	9999	HS, FS, PO	-	0.037	-	0.038
HZ08	HZ24	0.6247	FS	-18.04	-	1.21	2.41	9999	FS	-	-	-	-
HZ08	HZ25	0.1348	HS	-19.8	-	0.7	2.03	9999	U, HS, FS, PO	0.254	-	0.045	0.025
HZ09	HZ19	0.2256	HS	-24.73	-	1.97	4.52	9999	U, HS, FS	0.131	-	0.156	-
HZ10	HZ11	0.1675	HS	-33.01	-	1.83	3.8	9999	U, HS, FS	0.211	-	0.046	-
HZ10	HZ23	0.2685	HS	-22.03	-	0.47	2.07	9999	U, HS, FS	0.221	-	0.225	-
HZ11	HZ21	0.2005	HS	-31.76	-	1.48	4.26	9999	U, HS	0.266	-	-	-
HZ12	HZ13	0.5	PO	-13.73	-	2.22	5	9999	U, HS, FS, PO	0.029	0.145	0.064	-
HZ12	HZ15	0.6057	PO	-13.96	-	0.22	0.8	9999	HS, FS, PO	-	0.050	0.188	-
HZ12	HZ16	0.5	PO	-19.21	-	2.44	5.3	9999	HS, FS, PO	-	0.113	0.022	-
HZ12	HZ23	0.5	PO	-16.57	-	2.41	5.04	9999	U, HS, FS, PO	0.058	0.268	0.218	-
HZ13	HZ15	0.6141	PO	-13.63	1.58	0.51	0.13	-	HS, FS, PO	-	0.130	0.333	-
HZ13	HZ16	0.7647	FS	-15.89	-	0.6	1.55	9999	FS	-	-	-	-
HZ13	HZ22	0.1447	HS	-19.15	-	0.98	2.36	9999	U, HS, FS	0.227	-	0.057	-
HZ14	HZ15	0.2151	HS	-12.5	-	1.71	3.8	9999	U, HS, FS, PO	0.181	-	0.236	0.095
HZ14	HZ18	0.4787	FS	-13.36	-	0.55	0.99	9999	U, HS, FS, PO	0.037	0.117	-	0.119
HZ14	HZ20	0.717	FS	-16.18	-	1.06	2.8	9999	FS	-	-	-	-
HZ14	HZ22	0.552	PO	-17.46	-	1.1	3.42	9999	U, HS, FS, PO	0.038	0.227	0.326	-
HZ14	HZ24	0.5	PO	-12.38	-	0.64	1.24	9999	U, HS, FS, PO	0.097	0.332	0.263	-
HZ14	HZ25	0.6737	FS	-8.94	-	0.12	1.15	0.45	U, HS, FS, PO	0.039	0.092	0.082	-
HZ15	HZ16	0.6977	PO	-18.23	0.38	0.08	0.48	-	HS, FS, PO	-	0.060	0.372	-
HZ15	HZ18	0.1872	HS	-20.44	-	0.05	1.95	9999	U, HS, FS, PO	0.214	-	0.181	0.032
HZ15	HZ20	0.3698	HS	-21.24	-	0.96	2.33	9999	U, HS, FS	0.197	-	0.271	-

HZ15	HZ23	0.4697	PO	-15.95	-	1.31	3.65	9999	U, HS, FS, PO	0.102	0.339	0.262	-
HZ16	HZ22	0.2438	HS	-22.52	2.42	1.41	-	0.99	U, HS, FS	0.095	-	0.149	-
HZ16	HZ24	0.2533	HS	-22.19	-	0.34	1.08	9999	U, HS, FS	0.092	-	0.204	-
HZ17	HZ25	0.7148	FS	-22.08	-	0.39	0.75	9999	FS, PO	-	-	-	0.030
HZ20	HZ22	0.3641	FS	-25.4	-	2.05	4.38	9999	U, HS, FS	-	-	-	0.000
HZ22	HZ25	0.1574	HS	-19.53	-	1.4	3.57	9999	U, HS, FS, PO	0.248	-	0.127	0.025
HZ23	HZ24	0.5392	FS	-16.55	0.07	-	0.7	0.58	HS, FS, PO	-	0.086	-	0.056
HZ23	HZ25	0.164	HS	-16	0.13	-	1.37	0.43	U, HS, FS, PO	0.230	-	0.042	0.039

r, coeficiente de parentesco; R, relacionamento com maior probabilidade; LnL, verossimilhança logarítmica de R; Delta Ln(L), probabilidades de log para os outros relacionamentos; 95% IC, relacionamentos alternativos no nível de significância de 0,05; U, não relacionados; HS, meio-irmãos; FS, irmãos completos; PO, pais-filhos; 9999, relação não possível.

CAPÍTULO II

Investigação da estrutura genética do gavião-real (*Harpia harpyja*) no Brasil a partir de um sequenciamento de baixa cobertura

Investigação da estrutura genética do gavião-real (*Harpia harpyja*) no Brasil a partir de um sequenciamento de baixa cobertura

Mylena Kaizer^{1,2,3}, Jeronymo Dalapicolla⁴, Alexandre Aleixo⁴, Érico M. Polo¹, Tânia M. Sanaiotti^{2,5}, Izeni P. Farias^{1,2,3}, Aureo Banhos^{1,2,6}, Tomas Hrbek^{1,2,3}

¹ Laboratório de Evolução e Genética Animal - LEGAL, Departamento de Genética, Universidade Federal do Amazonas, Av. General Rodrigo Otávio Jordão Ramos, 3000, Coroado, 69077-000, Manaus, Amazonas, Brasil.

² Projeto Harpia, Instituto Nacional de Pesquisas da Amazônia - INPA, Av. André Araújo, 2936, Aleixo, 69067-375, Manaus, Amazonas, Brasil.

³ Programa de Pós-Graduação em Zoologia - PPGZOO, Universidade Federal do Amazonas, Av. General Rodrigo Otávio Jordão Ramos, 3000, Coroado, 69077-000, Manaus, Amazonas, Brasil.

⁴ Instituto Tecnológico da Vale - ITV, Rua Boaventura da Silva, 955, Nazaré, 66055-090, Belém, Pará, Brasil

⁵ Coordenação de Biodiversidade, Instituto Nacional de Pesquisas da Amazônia - INPA, Avenida André Araújo, 2936, Aleixo, 69067-375, Manaus, Amazonas, Brasil.

⁶ Departamento de Biologia, Centro de Ciências Exatas, Naturais e da Saúde, Universidade Federal do Espírito Santo, Alto Universitário, Guararema, 29500-000, Alegre, Espírito Santo, Brasil.

Resumo

A Amazônia e Mata Atlântica são florestas tropicais megadiversas que atualmente são separadas por um corredor seco de vegetação aberta. Diversos estudos investigam a diversificação de espécies ou linhagens irmãs distribuídas de forma disjunta nessas florestas. O gavião-real (*Harpia harpyja*), uma ave de rapina ameaçada de extinção, é um exemplo, ocorrendo principalmente na Amazônia e na Mata Atlântica com escassos registros nesse corredor seco. Estudos para compreender a estrutura populacional do gavião-real no Brasil mostraram diferenças genéticas entre populações da Amazônia e da Mata Atlântica por meio do DNA mitocondrial (DNAm), porém não houve sinal de estrutura através dos microssatélites. Neste estudo, foram utilizados marcadores de polimorfismo de nucleotídeo único gerados a partir de um sequenciamento de baixa cobertura para investigar e testar hipóteses sobre a existência de estrutura genética para o gavião-real no Brasil. Foram selecionadas 49 amostras de gavião-real com origem conhecida e foram sequenciadas por meio da plataforma Illumina. Para identificar e resgatar as variantes foram utilizadas duas metodologias diferentes. Como resultado, todos os *pipelines* utilizados resgataram um baixo sinal de estrutura genética para o gavião-real no Brasil. O *pipeline* do *DiscoSnpRad* demonstrou maior eficácia para lidar com o sequenciamento de baixa cobertura resgatando informações significativas e consistentes com os resultados encontrados pelo DNAm com a população localizada ao Norte do Rio Amazonas apresentando maior diferenciação genética e as populações localizadas ao Sul do Rio Amazonas e na Mata Atlântica apresentando semelhanças genéticas. A diversidade genética apresentou valores moderados a alto, com menor valor em regiões com alto índice de desmatamento. No entanto, ainda há tempo para preservar e maximizar a diversidade genética por meio de um manejo genético eficiente e com a proteção de seus territórios. Com isso, espera-se que os resultados obtidos neste estudo sejam utilizados para orientar as estratégias de conservação para o gavião-real.

Palavras-chave: Harpia. Genética da Conservação. Diversidade Genética. Estrutura genética.

1. Introdução

A Amazônia e a Mata Atlântica são compostos por florestas tropicais e estão entre as florestas mais diversas do mundo (Jenkins et al., 2013). Atualmente esses dois biomas brasileiros estão separados por um amplo corredor seco de vegetação aberta, compreendendo as ecorregiões Cerrado, Caatinga e Chaco Seco. Entretanto, estudos biogeográficos, geológicos e palinológicos evidenciaram que a Amazônia e a Mata Atlântica foram contínuas no passado e que existem conexões históricas entre elas (Oliveira-Filho e Ratter, 1995; Costa, 2003). Assim, diversos estudos utilizam dados moleculares para investigar a existência e a diversificação de espécies ou linhagens irmãs que estão distribuídas de forma disjunta nessas florestas megadiversas (Costa, 2003; Batalha-Filho et al., 2013; Machado et al., 2021).

O gavião-real ou harpia, cientificamente chamado de *Harpia harpyja*, é uma ave de rapina de grande porte e está ameaçada de extinção, classificada como Vulnerável em toda a sua distribuição (BirdLife International, 2023). A espécie ocorre em florestas tropicais da América Central e do Sul, com maior abundância no território brasileiro, principalmente na Amazônia e na Mata Atlântica. Todavia, existem raros registros de nidificação e de deslocamento da espécie no corredor de vegetação aberta do centro do Brasil (Oliveira e Silva, 2006; Silva et al., 2013). Esses registros geralmente estão associados às bordas dos rios com presença de matas ciliares ou de manchas florestais que funcionam como uma conexão entre as florestas tropicais brasileiras (Banhos, 2009).

Apesar do gavião-real possuir alta capacidade de explorar vastas áreas durante seus voos, a espécie depende do habitat florestal para sua locomoção e sobrevivência (Álvarez-Cordero, 1996). Assim, a separação histórica da Amazônia e da Mata Atlântica e o aumento do desmatamento das conexões de floresta presente no corredor seco (MapBiomas, 2022), pode ter impactado negativamente a dispersão da espécie entre essas duas regiões, podendo ao longo do tempo ocasionar o isolamento reprodutivo entre essas populações. Em vista disso, Banhos (2009) investigou a presença de distintas populações de gavião-real no Brasil, encontrando diferenças genéticas significativas entre as populações da Amazônia e da Mata Atlântica por meio da região controle do DNA mitocondrial (DNAMt), de modo que quanto maior a distância geográfica entre as populações, maior era diferenciação genética entre elas.

Como o DNAMt possui herança materna, a estrutura genética encontrada por Banhos (2009) precisa ser esclarecida por marcadores moleculares biparentais que permitem a detecção da estrutura populacional não confundida pelo sexo. Dessa forma, a mesma hipótese foi testada utilizando microssatélites transferidos para o gavião-real, os quais não encontraram presença de estrutura genética (Banhos, 2009; Banhos et al., 2016). Por mais que os microssatélites sejam considerados altamente polimórficos, a ausência de estrutura genética evidenciada por Banhos (2009) e Banhos et al. (2016) pode ser efeito da baixa informação genética contida em poucos loci distribuídos no genoma, necessitando ser investigada mais a fundo por meio de marcadores moleculares que fornecem uma cobertura genômica maior, como os polimorfismos de nucleotídeo único (SNPs).

O uso de marcadores SNPs se tornou mais acessível para a pesquisa científica devido ao avanço tecnológico do sequenciamento de nova geração (NGS), o qual permite sequenciar o genoma completo de espécies não modelo e detectar variações genéticas em escala genômica. Atualmente, uma das estratégias utilizadas é o sequenciamento completo do genoma de baixa cobertura que contém capacidade de identificar corretamente as variantes com menor custo (Li et al., 2021). Entretanto, é preciso ter mais cautela no processamento dos dados de baixa cobertura para reduzir a probabilidade de resgatar SNPs falsos positivos. Além disso, diversas técnicas para estudos com SNPs têm sido aprimoradas para identificar e filtrar as variantes e os diferentes programas e *pipelines* de NGS criados para identificar essas variantes podem gerar resultados quantitativos e qualitativos diferentes (Lee et al., 2021).

Diante disso, o presente trabalho tem como objetivo esclarecer a presença de estrutura genética do gavião-real no Brasil, por meio de marcadores biparentais em escala genômica (SNPs), gerados pelo sequenciamento de baixa cobertura utilizando diferentes metodologias para descobrir e testar a presença de estrutura genética para a espécie.

2. Material e Métodos

2.1. Amostragem biológica e extração do DNA

Para realizar o estudo foram utilizadas amostras de gavião-real cedidas pelo Projeto Harpia (www.projetoharpia.org), que possui um acervo de amostras da espécie provenientes de coletas anteriores realizadas durante as atividades de monitoramento (Apêndice 1). Para realizar a genômica, foram selecionadas do acervo 49 amostras de sangue ou tecido de gaviões-reais nascidos em vida livre e de localidade conhecida. Dentre as amostras selecionadas, contém exemplares depositados em coleções, animais de vida livre (indivíduos em ninhos, resgates, reabilitados e caçados) e animais de cativeiro nascidos em vida livre. Além disso, as amostras são de diferentes localidades do Brasil, situadas principalmente na Amazônia e na Mata Atlântica.

Todas as amostras de sangue e tecido foram enviadas para o Laboratório de Evolução e Genética Animal (LEGAL) da Universidade Federal do Amazonas (UFAM) juntamente com suas respectivas identificações como, a identificação individual, sexo, idade e origem de cada indivíduo. As amostras foram acondicionadas em freezer e depositados na Coleção de Tecidos de Genética Animal (CTGA) da UFAM, em Manaus. O DNA foi extraído usando o kit Qiagen (Valencia, CA, EUA) e sua qualidade, concentração e pureza do DNA foram verificadas no espectrofotômetro NanoDrop 2000 (Thermo Scientific).

As análises subsequentes (ver adiante) para realizar testes de hipótese da presença de estrutura genética e das estimativas da diversidade genética das populações de gavião-real foram definidas a partir de estudos anteriores (Banhos, 2009; Banhos et al., 2016). O número amostral e a distribuição dos indivíduos dentro dos grupos são apresentados de acordo com as amostras selecionadas para esse estudo e pelas filtragens do banco de dados. Desse modo, para as análises de teste de hipótese todos os bancos de dados foram divididos em três

populações: (1) Norte do Amazonas, indivíduos amostrados ao norte do Rio Amazonas; (2) Sul do Amazonas, indivíduos amostrados ao sul do Rio Amazonas e (3) Mata Atlântica (Figura 1).



Figura 1. Distribuição geográfica das amostras de gavião-real.

2.2.Preparação da biblioteca genômica e sequenciamento

A estratégia utilizada para obter as sequências genômicas foi uma alternativa que busca melhor custo e benefício na geração de dados genômicos em larga escala, o sequenciamento completo do genoma de baixa cobertura do tipo *Shotgun*, o qual as amostras são sequenciadas em *paired-end*, utilizando uma abordagem que produz duas leituras para cada fragmento de DNA sequenciado, uma referente à fita *forward* (*read 1* - R1) e outra à fita *reverse* (*read 2* - R2). O preparo das bibliotecas genômicas e seu o sequenciamento das mesmas foram realizados no Instituto Tecnológico da Vale, em Belém (PA) (www.itv.org).

O primeiro passo para obtenção das sequências foi realizar a preparação da biblioteca genômica seguindo o protocolo “*SureSelect^{QXT} Target Enrichment for the Illumina Platform*” (Agilent Technologies, 2021), o qual utiliza o kit *SureSelect^{QXT}* (Agilent Technologies) e possui

6 etapas, sendo elas: (1) fragmentação do DNA por meio de enzimas e adição dos adaptadores às extremidades dos fragmentos em uma única reação; (2) purificação das bibliotecas marcadas com adaptadores usando esferas magnéticas (*beads*) AMPure XP; (3) amplificação das bibliotecas marcadas com adaptadores por meio da reação em cadeia da polimerase (PCR); (4) purificação das bibliotecas amplificadas usando *beads* AMPure XP; (5) avaliação do tamanho médio dos fragmentos gerados nas bibliotecas através do 4200 *TapeStation System* (Agilent Technologies) e (6) quantificação da concentração final das bibliotecas no Qubit 3.0 (Thermo Fisher Scientific).

O segundo passo foi juntar volumes equimolar das bibliotecas em um único tubo para obter o *pool* das bibliotecas, quantificar a concentração e verificar a qualidade da biblioteca completa no Qubit 3.0 (Thermo Fisher Scientific). Em seguida, realizou-se a desnaturação do *pool* para que as fitas de DNA permanecessem abertas durante o processo de sequenciamento. Por fim, o sequenciamento completo do genoma de baixa cobertura foi realizado na plataforma Illumina NextSeq 500 (Illumina Inc, USA) utilizando o kit Illumina NextSeq High Output 300 ciclos (Illumina Inc, USA).

Após o sequenciamento, foi realizado o pré-processamento dos dados brutos através da bioinformática. De início, foi verificada a qualidade do sequenciamento através do *BaseSpace*, um programa da Illumina para acompanhamento do sequenciamento à distância. Posteriormente, foi realizada a demultiplexação das bibliotecas através do *Follow-UP*, um programa que atribui às leituras sequenciadas a sua amostra de origem com base na sequência dos seus adaptadores que foram adicionados durante a preparação das bibliotecas, gerando dois arquivos *fastq* para cada indivíduo (R1.fastq e R2.fastq). Então, a qualidade das sequências foram avaliadas usando FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc/) e as sequências com baixa qualidade foram removidas.

2.3. Genoma de referência do gavião-real

O genoma de referência do gavião-real depositado no Genbank (GCF_026419915.1) foi baixado para ser utilizado na identificação e extração dos SNPs. No total, o gavião-real contém 30 cromossomos, 28 autossômicos e dois sexuais (Z e W). Além desses cromossomos, o genoma de referência também inclui *scaffolds* adicionais. Dentre eles, alguns pertencem a cromossomos específicos, mas não puderam ser montados com segurança, enquanto outros *scaffolds* não foram montados. Esses *scaffolds* são principalmente compostos por regiões repetitivas e um deles contém sequências do genoma mitocondrial. Portanto, para utilizar o genoma de referência na extração dos SNPs, o genoma foi filtrado mantendo apenas as sequências referentes aos cromossomos e aos *scaffolds* atribuídos aos cromossomos.

2.4. Extração e pré-filtragem dos SNPs

Para extração dos SNPs, foram utilizadas duas metodologias diferentes: (1) o *pipeline* baseado principalmente no *bwa-mem2* (Vasimuddin et al., 2019), *samtools* (Danecek et al., 2021) e *Genome Analysis Toolkit* (GATK) (gatk.broadinstitute.org) e (2) o *pipeline* *Discovering*

Single Nucleotide Polymorphism (DiscoSnp++) (<https://github.com/GATB/DiscoSnp>). O *pipeline* do GATK requer um genoma de referência para realizar o alinhamento contra as sequências individuais e identificar as variantes (McKenna et al., 2010). Essa abordagem foi projetada principalmente para pesquisas envolvendo o genoma humano e é muito utilizada para diagnósticos médicos (Hsu et al., 2017). Em contrapartida, o *DiscoSnp++* não exige nenhum genoma de referência para detectar as variantes e seu algoritmo é baseado na construção do grafo *Bruijn* (Uricaru et al., 2014; Peterlongo et al., 2017).

O *DiscoSnp++* foi inicialmente desenvolvido para identificar variantes a partir de dados de sequenciamento genômico completo (Uricaru et al., 2015). No entanto, uma vez que a maioria dos dados utilizados em análises populacionais são baseados em sequenciamento reduzido representativo do genoma, uma adaptação do *DiscoSnp++* foi feita para lidar especificamente com esse tipo de dado, resultando no desenvolvimento do *pipeline DiscoSnpRad* (Gauthier et al., 2020). Tanto o *pipeline* do GATK quanto o *DiscoSnp++* foram projetados para dados com cobertura relativamente alta. Em oposição a isso, o *DiscoSnpRad* foi desenvolvido para lidar com dados de cobertura heterogênea, em que algumas regiões podem ter cobertura igual a zero e outras com cobertura variável, desde baixa até alta.

O fluxo de trabalho do *pipeline* do GATK geralmente compreende quatro etapas de análise: (1) pré-processamento dos dados de sequência bruta; (2) mapeamento das sequências contra o genoma de referência; (3) descoberta das variantes e (4) filtragem das variantes (*hard* e *soft filters*). Neste estudo, o fluxo de trabalho começou com a junção das leituras pareadas R1 e R2 utilizando *seqtk*. Em seguida, o restante dos adaptadores foram removidos através do *cutadapt* e as leituras foram mapeadas contra o genoma de referência do gavião-real usando *bwa-mem2*. Os arquivos mapeados foram ordenados e anotados utilizando *samtools*. Posteriormente, as variantes foram identificadas e extraídas pelo *pipeline* do GATK e convertido em um arquivo VCF (*variant call file*). Esse arquivo VCF contendo as variantes (SNPs e *indels*) foi submetido a uma filtragem usando *bcftools*, implementando os “*hard filters*” recomendados pelo GATK Team e Kuderna et al. (2023). Todas essas etapas descritas acima foram repetidas para cada indivíduo. Por fim, o *bcftools* foi utilizado para juntar todos os arquivos VCFs individuais em um único arquivo VCF, que foi posteriormente filtrado usando os “*soft filters*” implementados no pacote R (R Core Team, 2022) *vcf2others* (Apêndice 2).

Quanto ao *DiscoSnp++* e o *DiscoSnpRad*, ambos seguem um fluxo de trabalho semelhante, composto por duas etapas: (1) a construção o grafo de *Bruijn* e detecção das variantes (SNPs e *indels*) e o (2) mapeamento das variantes encontradas nas leituras originais (Uricaru et al., 2015). No caso do *DiscoSnpRad*, existe uma terceira etapa, onde as variantes previstas são agrupadas por locus (Gauthier et al., 2020). Sendo assim, os arquivos em fastq (R1 e R2) foram submetidos aos *pipelines*, onde o primeiro passo para ambos foi construir o grafo de *Bruijn* para detecção dos SNPs e *indels* em todas as sequências. Após, as variações encontradas foram mapeadas ao conjunto de sequências originais de todos os indivíduos, determinando e avaliando a cobertura de leitura por alelo e por conjunto de leitura de cada

variante. Além disso, nessa etapa cada variante foi associada a uma pontuação de confiança e genotipada em cada conjunto de leitura.

Para o *DiscoSnpRad*, sua etapa adicional agrupou as variantes encontradas por locus, recuperando informações associadas às proximidades físicas das variações. Tanto o *DiscoSnp++* quanto o *DiscoSnpRad*, assim como o GATK, possuem em seus *pipelines* a inclusão de um conjunto de “*hard filters*” para realizar a filtragem das variantes identificadas. Por fim, foi gerado um arquivo de saída em VCF para cada *pipeline* do *DiscoSnp* utilizado e foram posteriormente filtrados usando os “*soft filters*” implementados no pacote R (R Core Team, 2022) *vcf2others* (Apêndice 2). Os mesmos “*soft filters*” foram aplicados para os VCFs de todos os *pipelines*.

Em resumo, para testar a existência de estrutura genética para o gavião-real no Brasil, foram gerados três VCFs finais: (1) SNPs resgatados pelo GATK; (2) SNPs resgatados pelo *DiscoSnp++*; (3) SNPs resgatados pelo *DiscoSnpRad*. Como todos os bancos de dados utilizaram as mesmas sequências brutas geradas pelo sequenciamento de baixa cobertura e os mesmos “*soft filters*”, a diferença entre eles está na maneira como as variantes foram detectadas e extraídas. Esses testes foram conduzidos para investigar se a metodologia de extração das variantes afeta os resultados tanto em termos quantitativos quanto qualitativos para bancos de dados com baixa cobertura.

2.5. Filtragem dos SNPs no R (*soft filters*)

Com intuito de obter bancos de dados mais coesos e robustos, a filtragem dos SNPs foi realizada na plataforma de programação R (R Core Team, 2022) através de um script personalizado *vcf2others* (Apêndice 2). Os parâmetros utilizados nessa etapa foram chamados de “*soft filters*” e serviram para compensar as perdas de informações causadas pelo sequenciamento completo de baixa cobertura, garantir a ausência de alelos falso positivos e descartar possíveis erros de sequenciamento que em análises futuras acarretaria em resultados espúrios.

Todos os bancos de dados foram filtrados baseados em cinco parâmetros na seguinte ordem: (1) porcentagem de dados faltantes por indivíduo, retirando os indivíduos com mais de 90% de dados faltantes; (2) a frequência dos alelos menores (MAF), excluindo os loci cujo segundo alelo mais comum apresentasse uma frequência abaixo de 3%; (3) cobertura dos *reads*, removendo *loci* com uma cobertura menor que 5 vezes; (4) proximidade física das variantes, retraindo somente SNPs não ligados, ou seja, um SNP por locus; (5) porcentagem de dados faltantes por locus, retirando os loci presentes em menos de 50% das amostras. Por fim, novamente foi aplicado o filtro que avalia a porcentagem de dados faltantes por indivíduo, removendo aqueles que ainda apresentavam mais de 50% dos loci obtidos ausentes.

2.6. Análises de estrutura genética

Alguns métodos, como o Structure, são baseados em modelos e pressupostos, que quando são violados podem impactar na obtenção dos melhores agrupamentos genéticos (Pritchard et al., 2000). Essas violações podem ocorrer devido a diversos fatores, como

populações com padrões consistente ao isolamento por distância (Ruiz-Gonzalez et al., 2015), subdivisões populacionais com baixo valor de F_{ST} (Latch et al., 2006) e pequeno tamanho amostral (Evanno et al., 2005). Cabe ressaltar que a metodologia de sequenciamento adotada neste estudo também pode resultar em violações do modelo devido à baixa cobertura, resultando na possível perda de heterozigotos por falta de informação e violando o pressuposto do equilíbrio de Hardy-Weinberg (HWE).

Considerando análises anteriores da estrutura genética do gavião-real (Banhos, 2009; Banhos et al., 2016) e levando em conta os possíveis fatores que podem afetar os pressupostos dos métodos baseados em modelos, optamos por realizar análises que não dependem desses pressupostos, para explorar a diferença genética entre as populações de gavião-real no Brasil. Assim, foram realizadas duas abordagens diferentes para as análises da estrutura genética: uma sem informação *a priori* sobre a estrutura (i.e. ferramentas para descobrir se existe estrutura a partir dos dados genéticos), utilizando a análise de componentes principais (PCA; Lever et al., 2017); e outra com informações *a priori* da estrutura (i.e. testes de hipótese da estrutura e a quantificação da mesma), utilizando a análise discriminante de componentes principais (DAPC; Jombart et al., 2010) e a análise de variância molecular (AMOVA), para testar a divergência entre os grupos previamente designados.

A PCA e a DAPC foram realizadas utilizando o pacote *adeigenet* (Jombart, 2008) na plataforma R (R Core Team, 2022). Enquanto a PCA não recebeu nenhuma informação prévia sobre as amostras, a DAPC utilizou informações sobre a distribuição geográfica dos indivíduos categorizados como Norte da Amazônia, Sul da Amazônia e Mata Atlântica. Para calcular a análise de variância molecular (AMOVA) foi utilizado o programa Arlequin 3.5 (Excoffier e Lischer, 2010). Essa análise também considerou a origem das amostras, fornecendo os valores dos índices de fixação (F_{ST}) e de endogamia (F_{IS} e F_{IT}), juntamente com seus respectivos valores de p, com nível de significância de $< 0,05$. Os conhecimentos prévios utilizados na DAPC e AMOVA foram baseados em estudos anteriores (Banhos, 2009) e no presente estudo. Sendo assim, as amostras foram subdivididas em três grupos: (1) Norte do Amazonas, com indivíduos acima do Rio Amazonas; (2) Sul do Amazonas, com indivíduos abaixo do Rio Amazonas e (3) Mata Atlântica (Apêndice 1).

Considerando que o banco de dados recuperado pelo *DiscoSnpRad* obteve maior quantidade de SNPs (ver adiante), foi realizada uma reamostragem utilizando SNPs randomizados para investigar se a quantidade de variantes interferiu nos resultados finais das análises de estrutura genética. Para isso, a partir do conjunto de dados completo, foram criadas 500 subamostras aleatórias contendo 1000 SNPs. Para cada subamostra, foram calculados os valores de F_{ST} e seus respectivos valores de p, com nível de significância de $< 0,05$. Em seguida, esses valores foram sumarizados e comparados com o valor obtido a partir do banco de dados completo. As análises foram realizadas utilizando o pacote *poppr* (Kamvar et al., 2014) na plataforma R (R Core Team, 2022).

2.7. Análise de diversidade genética

Para avaliar a diversidade genética dentro da população de gavião-real foram utilizados diferentes parâmetros de diversidade genética, incluindo a proporção de SNPs polimórficos (P_n) e a heterozigosidade observada (H_o) e esperada (H_e). Todos os parâmetros foram obtidos através do programa Arlequin 3.5 (Excoffier e Lischer, 2010).

2.8. Extração do mitogenoma

A localização e extração do mitogenoma nas sequências fastq (R1 e R2), foi realizada através do *pipeline* GetOrganelle (Jin et al., 2020) com adaptações. Para facilitar a identificação das sequências mitocondriais das amostras de gavião-real e realizar o mapeamento das leituras de um modo mais rápido e preciso, o banco de dados de referência usado pelo GetOrganelle (animal_mt) foi substituído pelo mitogenoma completo da águia-cobra-de-peito-preto (*Circaetus pectoralis*), uma espécie filogeneticamente próxima do gavião-real, pertencente à mesma família Accipitridae. O mitogenoma completo da espécie foi extraído do Genbank (NC_052805.1).

Diante disso, o GetOrganelle utiliza o banco de dados de referência como “isca” para recrutar as leituras associadas ao alvo. Isso permite que a ferramenta faça a montagem do mitogenoma para cada indivíduo. Uma vez obtidas as sequências mitocondriais, elas foram editadas e alinhadas separadamente para cada região mitocondrial usando o software Aliview (Larsson, 2014). Com as sequências alinhadas, foi possível identificar que alguns genes eram extremamente conservados, ou seja, não havia variação entre os indivíduos. Portanto, apenas as regiões mitocondriais com variação foram selecionadas para análises subsequentes.

2.9. Análises filogenéticas do mitogenoma

As respostas sobre a história evolutiva do gavião-real e descoberta sobre diferentes linhagens dentro da espécie, foi através de uma análise com inferência bayesiana no Beast2 (Suchard et al., 2018) utilizando o mitogenoma dos indivíduos amostrados. Porém, antes de executar a análise filogenética, foi utilizado o PartitionFinder 2 (Lanfear et al., 2017) para selecionar automaticamente o esquema de partição com melhor ajuste e informar o modelo de substituição de nucleotídeos mais adequado para cada partição. Essa etapa foi utilizada para aumentar a eficiência das análises filogenéticas, demandando menor esforço computacional.

Para iniciar a análise no PartitionFinder 2 todas as sequências mitocondriais editadas e alinhadas foram combinadas em um conjunto de dados concatenados. No arquivo de entrada foi informado o nome do gene e o seu respectivo número inicial e final da sequência. Para os genes codificadores de proteínas foi adicionado o esquema de três partições, separando todas as primeiras, segundas e terceiras posições do códon. Além disso, foi utilizado o algoritmo greedy, com comprimentos de ramificação estimados como desvinculados entre si e selecionado o modelo de sob o Critério de Informação de Akaike Corrigido (AICC), para determinar qual era o melhor ajuste para os dados. Os esquemas de particionamento mais adequados determinados pelo PartitionFinder 2 foram implementados na análise filogenética.

Os arquivos de entrada foram preparados no BEAUti2 (Suchard et al., 2018) para realizar a análise filogenética no Beast, onde os genes codificadores de proteínas foram agrupados de acordo com o melhor ajuste encontrado no PartitionFinder 2. O modelo de evolução molecular para cada partição foi estimado durante a corrida com o bModelTest, considerando-se todos os modelos rotulados sob "namedExtended" (Bouckaert e Drummond, 2017). Assumiu-se um relógio molecular relaxado com *prior* log-normal para estimar os tamanhos relativos dos ramos e um modelo coalescente exponencial como *prior* da árvore. Por fim, foi executado o método Monte Carlo via Cadeias de Markov (MCMC) por 100 milhões de gerações, com dados amostrados a cada 10.000 etapas e exclusão dos primeiros 10% de cada execução, considerado como a fase de queima.

Para visualizar a concordância entre os resultados, foram realizadas duas análises utilizando a mesma estratégia. A partir disso, os arquivos de saída do Beast foram combinados através do LogCombiner (Suchard et al., 2018) e observados e avaliados no Tracer (Rambaut et al., 2018). Posteriormente, foi criada uma árvore de consenso, sendo realizada através da árvore de Credibilidade Máxima de Clado (MCC) com um *burnin* de 10%, usando o programa TreeAnnotator (Suchard et al., 2018). A etapa final foi o carregamento da árvore consenso no Figtree (Rambaut, 2018), permitindo visualizar a árvore e as informações resumidas produzidas pelo TreeAnnotator.

3. Resultados

3.1. Sequenciamento e extração dos SNPs

O sequenciamento do genoma completo de baixa cobertura para os 49 indivíduos de gavião-real resultou em um total de 269 milhões de leituras, com média de 5.48 milhões de leituras por indivíduo (variando de 2.13 milhões a 11.14 milhões). Dado que o genoma do gavião-real (<https://www.ncbi.nlm.nih.gov/genome/?term=txid202279>) tem 1.268,38 Mb (1.268 bilhões de nucleotídeos), o sequenciamento foi feito numa profundidade média de 1.29x (variando de 0.50x a 2.64x) para cada indivíduo. Ao mapear essas sequências ao genoma de referência do gavião-real utilizando o *pipeline* do GATK, foram identificados 7.416.452 SNPs. Após a aplicação dos "*hard filters*" recomendados pelo GATK Team e Kuderna et al. (2023), restaram 11.257 SNPs. Ao utilizar os "*soft filters*", o número de SNPs foi reduzido para 6.935 para um total de 49 indivíduos (Tabela 1).

Por outro lado, ao utilizar os programas *DiscoSnp++* e *DiscoSnpRad*, foram resgatados 502.166 e 465.854 SNPs, respectivamente, após a aplicação dos "*hard filters*" internos dos programas. Em seguida, ao aplicar os "*soft filters*", das variantes resgatadas pelo *DiscoSnp++*, apenas 859 SNPs foram mantidos e 6 indivíduos (H240, H294, H316, H321, H328, Xingu) foram removidos do conjunto de dados. No caso do banco de dados recuperado pelo *DiscoSnpRad*, os "*soft filters*" filtraram 30.130 SNPs e removeram 3 indivíduos (H328, H240, H316) das análises subsequentes (Tabela 1).

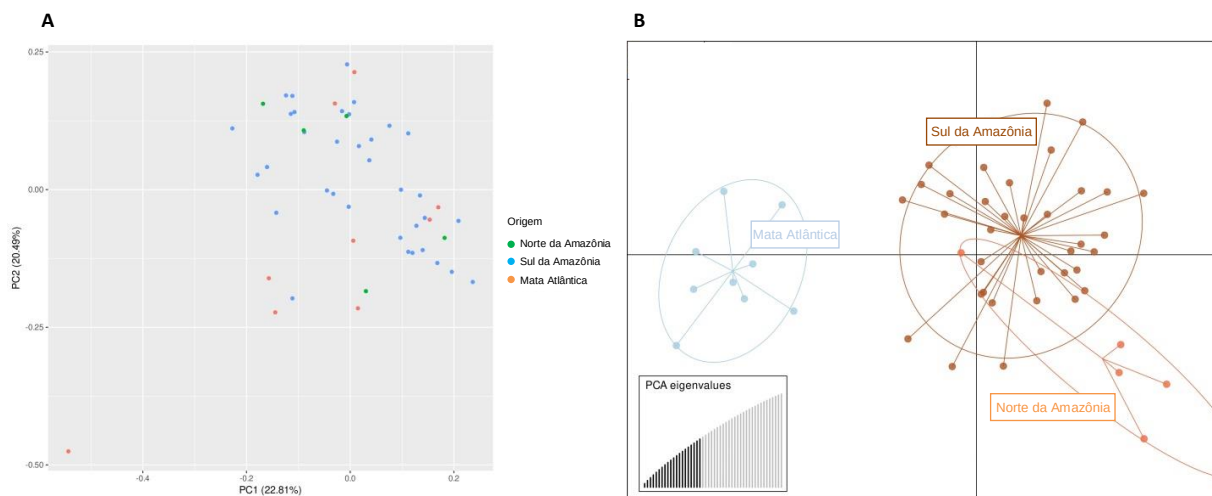
Tabela 1. Número de SNPs filtrados após a implementação do *hard* e *soft filters* e o número de indivíduos utilizados nas análises subsequentes por *pipeline*.

<i>Pipeline</i>	<i>Hard filters</i>	<i>Soft filters</i>	Número de indivíduos
GATK	11.257	6.935	49
<i>DiscoSnp++</i>	502.166	859	43
<i>DiscoSnpRad</i>	465.854	30.130	46

3.2. Estrutura populacional

Os resultados da PCA para os conjuntos de dados extraídos pelos *pipelines* GATK (Figura 2A) e *DiscoSnp++* (Figura 2C) não mostraram formação de *clusters* evidentes. Porém, o conjunto de dados extraído pelo *DiscoSnpRad* recuperou cinco *clusters* com distribuição desigual das amostras geográficas entre os *clusters* (Figura 2E). Para a DAPC, o GATK mostra sobreposição entre Norte e Sul da Amazônia e fraca separação da Mata Atlântica (Figura 2B). Em contrapartida, o *DiscoSnp++* (Figura 2D) e o *DiscoSnpRad* (Figura 2F) mostram uma leve sobreposição entre Mata Atlântica e Sul da Amazônia e uma maior separação do Norte da Amazônia.

Para todos os conjuntos de dados a AMOVA evidenciou um sinal muito fraco de estruturação populacional (Tabela 2), com valor significativo apenas para o conjunto de dados resgatado pelo *DiscoSnpRad* (Tabela 2). Os valores de F_{IS} e F_{IT} não foram significativos para todos os bancos de dados (Tabela 2), indicando ausência de endogamia para as populações. Em relação às reamostragens realizadas para o conjunto de dados do *DiscoSnpRad*, a média obtida para o F_{ST} foi de 0.0176, com valor de p significativo de 0.0048+0.00537. Esses valores indicam um baixo sinal de diferenciação genética para o gavião-real, assim como foi observado utilizando o painel completo de variantes para o *DiscoSnpRad* (Tabela 2).



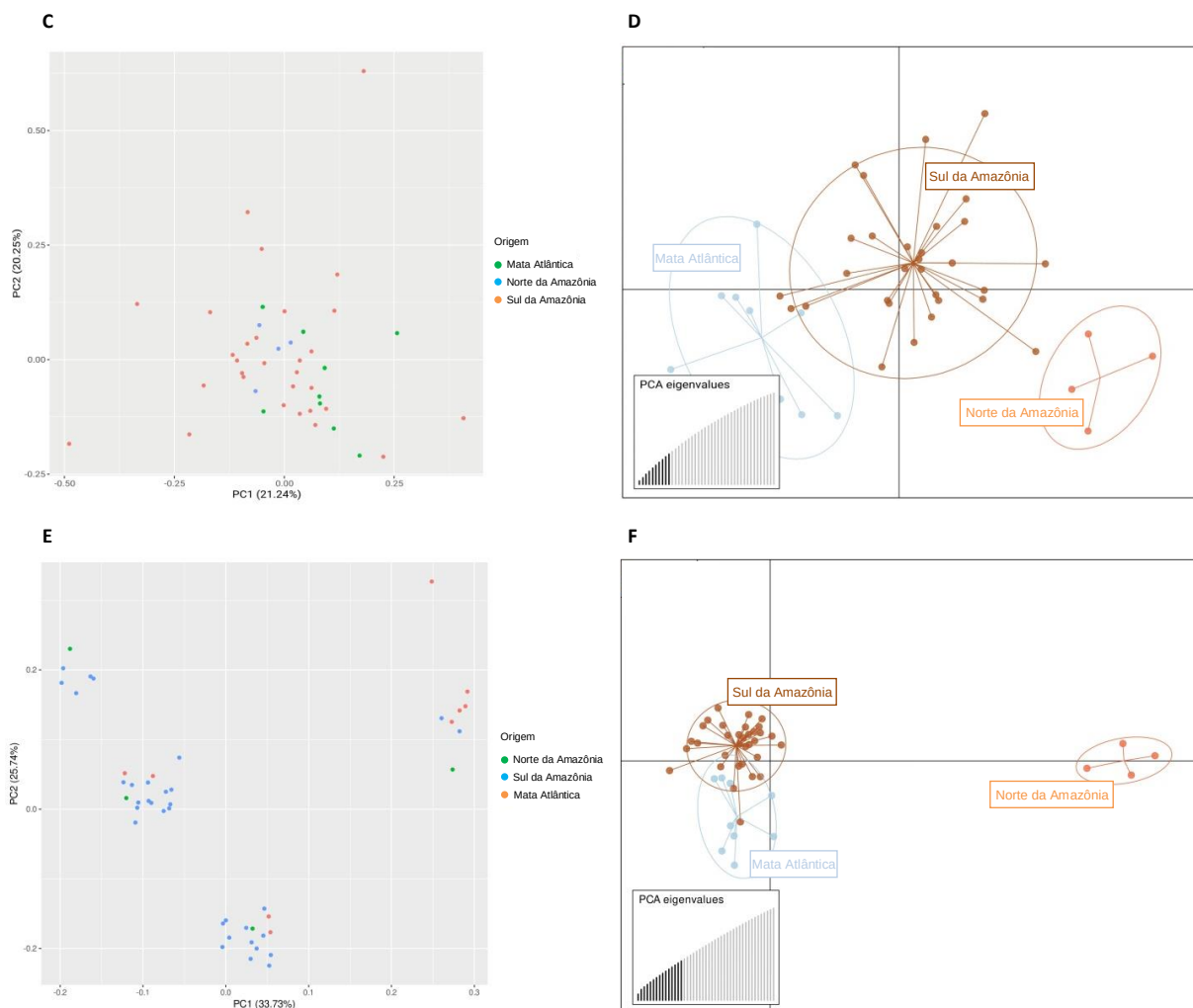


Figura 2. Resultados da análise de componentes principais (esquerda) e da análise discriminante de componentes principais (direita) para os diferentes conjuntos de dados. A e B para GATK, C e D para *DiscoSnp++* e E e F para *DiscoSnpRad*.

Tabela 2. Valores do F_{ST} , F_{IS} e F_{IT} e seus respectivos valores de p e desvio padrão para cada *pipeline* utilizado no estudo.

<i>Pipeline</i>	F_{ST}	P-valor	F_{IS}	P-valor	F_{IT}	P-valor
GATK	0.0006	0.3613 +0.0045	-1.0551	1.0000 +0.0000	-1.0537	1.0000 +0.0000
<i>DiscoSnp++</i>	0.0038	0.9815 +0.0014	-0.5345	1.0000 +0.0000	-0.5286	1.0000 +0.0000
<i>DiscoSnpRad</i>	0.0149	0.0154* +0.0010	-0.7614	1.0000 +0.0000	-0.7351	1.0000 +0.0000

* Nível de significância ($p < 0,05$).

3.3. Diversidade genética

Os resultados dos parâmetros de diversidade genética calculados para os diferentes grupos de gavião-real estão resumidos na Tabela 3. Os valores da heterozigosidade observada (H_o) foram maiores em comparação a heterozigosidade esperada (H_e) para todos os grupos, independentemente do *pipeline* utilizado para recuperar os SNPs. Para o *pipeline* do GATK, H_o

variou de 0.65897 a 0.73427 e H_e de 0.37331 a 0.5027. Quanto ao *DiscoSnp++*, H_o variou de 0.42995 a 0.61288 e H_e de 0.31083 a 0.47568. No caso do *DiscoSnpRad*, H_o variou de 0.47281 a 0.59198 e H_e de 0.32757 a 0.44238.

Considerando todos os conjuntos de dados, o grupo do Norte da Amazônia obteve maiores valores de H_o e H_e quando comparado com os grupos do Sul da Amazônia e da Mata Atlântica. O grupo Sul, por sua vez, obteve os menores valores. Cabe ressaltar que, apesar do *DiscoSnp++* ter recuperado um número bastante reduzido de SNPs, os valores de heteroziguidade foram semelhantes aos dos outros *pipelines*. Em relação ao Pn, a maior porcentagem foi observada para o grupo do Norte da Amazônia (41.10 %) pelo conjunto de dados extraído pelo *DiscoSnpRad* (Tabela 3). Assim como ocorreu para H_o e H_e , o grupo do Norte da Amazônia apresentou as maiores porcentagens de SNPs polimórficos, exceto no conjunto de dados do GATK, onde o grupo da Mata Atlântica com maior porcentagem (27.03 %) (Tabela 3).

Tabela 3. Estimativas da diversidade genética para os grupos de gavião-real, determinadas pelo número de SNPs resgatados pelos diferentes *pipelines*, proporção de SNPs polimórficos, número de indivíduos e heteroziguidade observada e esperada para cada grupo.

<i>Pipeline</i>	SNPs	Grupo	N	Pn (%)	H_o	H_e
GATK	6.395	NAM	5	25.44	0.73427	0.50275
		SAM	35	22.79	0.60720	0.37311
		ATF	9	27.08	0.65897	0.42604
<i>DiscoSnp++</i>	859	NAM	4	29.45	0.61288	0.47568
		SAM	30	7.45	0.42995	0.31083
		ATF	9	9.08	0.50655	0.37660
<i>DiscoSnpRad</i>	30.130	NAM	4	41.10	0.59198	0.47281
		SAM	33	12.26	0.44238	0.32757
		ATF	9	18.31	0.50605	0.38470

N, número de indivíduos; Pn, proporção de SNPs polimórficos; H_o , heteroziguidade observada; H_e , heteroziguidade esperada; NAM, Norte da Amazônia; SAM, Sul da Amazônia; ATF, Mata Atlântica.

3.4. Extração do mitogenoma

O GetOrganelle foi capaz de resgatar as sequências do mitogenoma de 48 indivíduos, com exceção da amostra H240. Essas sequências incluíram a região de controle (d-loop) e 28 genes, contando com 13 genes codificantes de proteínas (ATP6, ATP8, COX1, COX2, COX3, CYTB, ND1, ND2, ND3, ND4, ND4L, ND5 e ND6), 13 tipos de RNA transportador (trnE_uuc, trnG_ucc, trnH_gug, trnI_gau, trnL_uua, trnL_uag, trnM_cau, trnN_guu, trnP_ugg, trnR_ucg, trnS_uga, trnW_uca, trnY_gua) e dois tipos de RNA ribossômico (12S e 16S). Os genes ATP8, ND3, 12S, 16S e todos os RNAs transportadores não apresentaram variação entre os indivíduos, evidenciando serem genes conservados para a espécie, assim, foram excluídos das

análises subsequentes, restando somente o d-loop e 11 genes codificantes do mitogenoma (Tabela 4). No total, incluindo as regiões constantes e variáveis, foram resgatados 12.019 pb do mitogenoma, com média de aproximadamente 8.310 pb por amostra. Cada gene continha ao menos uma variação, resultando em 112 sítios segregantes, dos quais 46 foram parcimoniosamente informativos. Os esquemas de particionamento mais adequados determinados pelo PartitionFinder 2 agrupando o d-loop e os 11 genes com suas partições em 4 *subsets* diferentes (Tabela 4).

Tabela 4. Esquemas de particionamento determinados pelo PartitionFinder 2 utilizando o d-loop e 11 genes mitocondriais.

Subset	Região mitocondrial
1	ATP6_2, ND4_1, ND2_1, ND5_3, COX3_1, ND4L_1, CYTB_1, ND1_1, COX2_1, COX1_3
2	COX3_2, COX2_2, ND2_2, ATP6_3, COX1_1, CYTB_2, ND4L_2, ND5_1, ND1_2, ND4_2
3	COX3_3, ND4_3, COX2_3, ND5_2, CYTB_3, ND4L_3, ATP6_1, ND2_3, COX1_2, ND1_3
4	ND6_3, ND6_1, d-loop, ND6_2

A análise filogenética baseada no método bayesiano a partir das regiões do mitogenoma resultou na construção de uma árvore filogenética consistente, evidenciando a existência de duas linhagens evolutivas que se divergiram dentro da espécie. Os resultados da análise bayesiana apresentam um alto valor de bootstrap (BS = 1), indicando uma clara separação genética entre os indivíduos do Norte da Amazônia e os indivíduos do Sul da Amazônia e da Mata Atlântica. Além disso, os grupos do Sul da Amazônia e da Mata Atlântica compartilham haplótipos entre si.

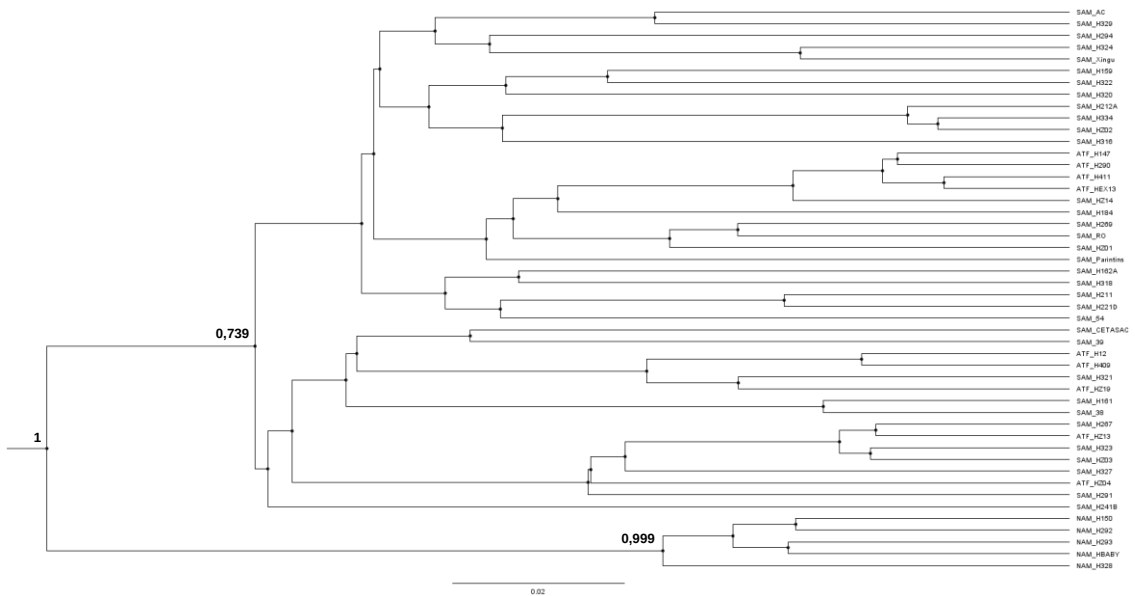


Figura 3. Árvore filogenética para o gavião-real, evidenciando diferentes linhagens para a espécie, com o grupo do Norte da Amazônia (NAM) separado geneticamente dos grupos do Sul da Amazônia (SAM) e Mata Atlântica (ATF).

4. Discussão

Compreender a estrutura genética de espécies ameaçadas de extinção e com ampla distribuição geográfica é de extrema importância para desenvolver e direcionar estratégias de manejo eficazes para sua conservação (Frankham et al., 2010). No caso do gavião-real, estudos anteriores foram conduzidos para investigar a diferenciação genética para a espécie no Brasil (Banhos, 2009; Banhos et al., 2016). No entanto, o uso de diferentes marcadores moleculares apresentaram resultados distintos, enquanto os microssatélites mostraram ausência de estrutura genética, o DNAm_t revelou diferenças genéticas significativas entre as populações da Amazônia e da Mata Atlântica, de forma que quanto maior a distância geográfica entre as populações, maior era diferença genética entre elas (Banhos, 2009).

Neste estudo, o mitogenoma também encontrou a presença de estrutura genética para o gavião-real no Brasil, de modo que as populações do Sul da Amazônia (ao sul do Rio Amazonas) e da Mata Atlântica são mais semelhantes geneticamente entre si e a população localizada ao norte do Rio Amazonas apresenta uma maior divergência genética entre as demais populações. Banhos et al. (2016) encontraram resultados similares para a espécie utilizando os microssatélites heterólogos. Embora os autores não tenham encontrado sinais de estrutura por meio de análises bayesianas, eles observaram baixos valores, porém significativos, de F_{ST} entre a população do Norte da Amazônia e as populações do Sul da Amazônia + Mata Atlântica. Além disso, observaram similaridade genética entre as populações do Sul da Amazônia e da Mata Atlântica, indicando a ocorrência de fluxo gênico entre essas populações (Banhos et al., 2016).

Como as análises com o DNAm_t mostraram sinais claros de estrutura genética (Banhos, 2009; este estudo) e os microssatélites não foram capazes de confirmar esse padrão (Banhos et al., 2016), o uso de marcadores nucleares mais abrangente no genoma, como os SNPs, tendem auxiliar a elucidar dinâmicas populacionais mais complexas (Kulski, 2016). No entanto, os resultados em relação à presença de estrutura genética utilizando os marcadores SNPs neste estudo variaram de acordo com a metodologia utilizada para identificar e extrair as variantes. Essa variação está diretamente ligada a abordagem de sequenciamento utilizada, onde o sequenciamento de baixa cobertura, apesar de ter capacidade de resgatar variantes importantes e concordantes com sequenciamentos de média e alta cobertura (Li et al., 2011; Rustagi et al., 2017; Gilly et al., 2019; Chaubey et al., 2020), possui um conjunto de dados com uma baixa quantidade de informações disponíveis por *loci*, dificultando a inferência estatística para identificar variantes verdadeiras (Nielsen et al., 2012).

Além disso, os *pipelines* NGS usados para identificar essas variantes não foram projetados para esse tipo de sequenciamento e podem produzir resultados qualitativos e quantitativos divergentes mesmo utilizando o mesmo dado bruto (McKenna et al., 2010; Uricaru et al., 2015; Lee et al., 2021), como observado neste estudo. Em relação ao *pipeline* do GATK, como é amplamente utilizado em diagnósticos médicos, seus critérios para determinar a presença de uma variante são mais rigorosos para minimizar o erro alfa. Essa rigidez, combinada com os dados gerados por um sequenciamento de baixa cobertura, resulta na

exclusão de muitas variantes verdadeiras que poderiam ser relevantes para detectar sinais de estrutura genética para o gavião-real. Apesar do GATK não ter sido o *pipeline* que manteve o menor número de SNPs, foi o único a revelar uma configuração de *clusters* diferente dos outros *pipelines* pelo método da DAPC, observando uma leve divergência entre as populações dos biomas Amazônia e Mata Atlântica, com os grupos do Norte e Sul da Amazônia se sobrepondo em alguns pontos.

Quanto ao *DiscoSnp++*, foi o *pipeline* que apresentou menor quantidade de SNPs após as filtrações. Assim como no GATK, o conjunto de dados obtidos pelo *DiscoSnp++* não revelou a descoberta de *clusters* através da PCA e nem apresentou diferenciação significativa para AMOVA. No entanto, mesmo com a recuperação de um número muito reduzido de SNPs, no método da DAPC o *pipeline* conseguiu resgatar sinais de estrutura genética semelhante àquela observada para o DNAm. No caso do *DiscoSnpRad*, o qual recuperou uma maior número de SNPs, apesar da PCA não ter identificado o número de *clusters* correlatos com a distribuição geográfica das populações, as variações extraídas por esse *pipeline* evidenciaram um F_{ST} significativo e mostrou por meio da DAPC que a população do Norte da Amazônia representa o grupo geneticamente mais distinto, com os agrupamentos do Sul da Amazônia e da Mata Atlântica apresentando maior similaridade genética. Além disso, cabe ressaltar que os achados utilizando os SNPs obtidos através do *pipeline* do *DiscoSnpRad*, são consistentes com os resultados encontrados através da análise filogenética com o mitogenoma.

Ademais, foi possível observar que os resultados divergentes entre os diferentes *pipelines* utilizados neste estudo não são inferidos pela quantidade de SNPs recuperados, mas sim pela metodologia empregada para identificar as variantes. Essa constatação foi confirmada pela a reamostragem realizada para o *DiscoSnpRad*, onde mesmo com um conjunto menor de SNPs, os resultados foram consistentes com aqueles obtidos utilizando o painel completo. Embora a metodologia utilizada para sequenciar as amostras neste estudo não tenha sido baseada na representação reduzida do genoma (RAD-Seq), os padrões genéticos do gavião-real identificados pelo *pipeline* *DiscoSnpRad* foram mais eficientes e semelhantes aos resultados observados para a espécie através do DNAm. Portanto, esses resultados sugerem que o *pipeline* do *DiscoSnpRad* é capaz de identificar e resgatar variantes mesmo com baixa cobertura de sequenciamento, mostrando sua eficácia ao lidar com esse tipo de dado e produzir resultados consistentes.

Mesmo utilizando uma amostragem do genoma mais ampla por meio dos SNPs, o gavião-real permanece apresentando baixo sinal de estrutura genética a partir de marcadores biparentais, indicando que ainda há mistura entre os grupos. A discrepância observada entre o grau de estrutura genética detectada pelos diferentes marcadores moleculares pode ser atribuída à dispersão desigual entre os sexos e as características específicas de cada marcador. Para o gavião-real, Banhos (2009) constatou maior dispersão para os machos. Como o DNAm possui herança materna e as fêmeas se dispersam menos, esse marcador consegue revelar uma estrutura genética histórica e, conseqüentemente, mais acentuada. Por outro lado, os marcadores biparentais apresentam um sinal mais fraco provavelmente devido a

esses machos migrantes, que quando se reproduzem na nova população, ocorre a mistura gênica e pode acabar sendo difundida com o passar de poucas das gerações.

Assim como o gavião-real, outras grandes aves de rapina com alto potencial de dispersão, como a águia-dourada (*Aquila chrysaetos*), também apresentam baixo sinal de estrutura genética, mostrando um padrão de variação genética consistente com isolamento por distância causado pela dispersão natal (Judkins et al., 2023). O padrão de variação genética observado no gavião-real aparentemente segue esse mesmo princípio, onde populações mais próximas têm maior chance de migrar e trocar genes entre si do que populações mais distantes (Banhos, 2009). Os eventos migratórios para a espécie foram observados com maior frequência originando da Mata Atlântica em direção a Amazônia, pelo leste do corredor seco utilizando as manchas florestais do Nordeste do Brasil (Banhos, 2009). Esse padrão de migração explica a maior similaridade genética encontrada entre a população do Sul da Amazônia e da Mata Atlântica e, além disso, indica que as conexões históricas entre a Mata Atlântica e a Amazônia tiveram grande importância na dispersão para a espécie.

A maior divergência genética observada no grupo do Norte da Amazônia pode ser explicada pela hipótese de que o Rio Amazonas atua como uma barreira geográfica. O Rio Amazonas teve sua formação há cerca de 7 milhões de anos e devido à sua grande extensão, divide a floresta Amazônica em dois blocos, um ao norte e outro ao sul. Além disso, ele é reconhecido como a barreira geográfica mais forte e eficaz entre todos os rios que compõem a Bacia Amazônica (Hoorn et al., 2010). Nesse sentido, para determinadas espécies, o rio pode reduzir ou impedir o fluxo gênico entre as populações localizadas em lados opostos do rio, inclusive para espécies que possuem alto poder de dispersão. Um exemplo disso é observado nas populações de onças-pintadas (*Panthera onca*), que também apresentaram diferenças genéticas significativas por causa da restrição do fluxo gênico entre a população do norte e sul do Rio Amazonas (Eizirik et al., 2001).

Analisando a diversidade genética encontrada neste estudo, mesmo com a disparidade do número de indivíduos entre os grupos, o grupo do Norte da Amazônia apresentou maior número de *loci* polimórficos e, conseqüentemente, maior heterozigosidade do que os demais grupos genéticos avaliados. Em contraste, o grupo localizado ao sul do Rio Amazonas obteve menor diversidade genética. O mesmo padrão de variação genética foi identificado em Banhos et al. (2016) através dos microssatélites transferidos e com base em amostras históricas e contemporâneas, os autores encontraram uma correlação entre a perda de diversidade genética com os altos índices de desmatamento na região do Sul da Amazônia e Mata Atlântica. Enquanto isso, as amostras situadas no Norte da Amazônia, região pouco alterada pela pressão do desmatamento, apresentou maior diversidade genética (Banhos et al., 2016; este estudo).

De fato, a manutenção da diversidade genética desempenha um papel essencial na sobrevivência e na capacidade de adaptação das espécies (Frankham et al., 2010). Embora a diversidade genética do gavião-real tenha sido encontrada em níveis mais baixos em áreas com altos índices de desmatamento, ela ainda é maior em comparação com outras espécies

de águias que não estão ameaçadas de extinção (Judkins et al., 2020; Judkins et al., 2023). Além disso, geralmente SNPs por serem marcadores bialélicos costumam captar menor índice de diversidade genética (Allendorf et al., 2013; Turchetto-Zolet et al., 2017). Neste estudo, mesmo utilizando os SNPs para quantificar a diversidade genética, as populações de gavião-real apresentaram uma diversidade moderada a alta para a espécie.

Outros estudos utilizando diferentes marcadores moleculares também encontraram uma diversidade genética variando de moderada a alta para a espécie (Banhos, 2009; Lerner et al., 2009). Isso significa que ainda há tempo para preservar e maximizar a diversidade genética por meio de um manejo genético eficiente e com a proteção de seus territórios. Vale ressaltar que a conservação da diversidade genética do gavião-real não deve se limitar apenas às áreas onde existem altos índices de desmatamento. As medidas de conservação devem ser implementadas em todas as áreas onde a espécie ocorre levando em consideração as diferentes unidades evolutivas da espécie, garantindo que as diferentes populações do gavião-real se mantenham geneticamente saudáveis e capazes de se adaptar às mudanças do ambiente.

Este estudo possibilitou melhor entendimento acerca da estrutura genética para o gavião-real no Brasil, indicando a existência de unidades evolutivas significativas (ESUs). Assim, é necessário levar em consideração as novas informações exploradas neste estudo para direcionar o manejo e a proteção das populações com objetivo de preservar a diversidade genética existente, garantir a evolução contínua da espécie e assegurar a capacidade de adaptação às mudanças ambientais a longo prazo. Além disso, é prudente que sejam evitadas translocações e reintroduções de indivíduos entre as diferentes ESUs, pois populações geneticamente diferenciadas não devem ser misturadas e requerem manejo separado (Frankham et al., 2010; Allendorf et al., 2013). Por fim, com este estudo também foi possível visualizar a importância de escolher a ferramenta correta para resgatar as variáveis (SNPs) de acordo com a abordagem de sequenciamento utilizada. A partir dos resultados obtidos aqui indicamos utilizar o *pipeline* do *DiscoSnpRad* para um conjunto de dados gerados a partir de um sequenciamento de baixa cobertura.

5. Agradecimentos

Ao Projeto Harpia por ceder amostras para realização deste estudo, ao Instituto Tecnológico da Vale por realizar o sequenciamento das amostras, à Universidade Federal do Amazonas e ao Laboratório de Evolução e Genética Animal onde as etapas de bioinformática foram desenvolvidas.

6. Financiamento

O presente trabalho teve apoio financeiro da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001, da Vale S.A e do Projeto Harpia.

7. Referência

Agilent Technologies (2021) SureSelect^{QXT} Target Enrichment for the Illumina Platform: Featuring Transposase-Based Library Prep Technology. Agilent Technologies, California

Allendorf FW, Luikart GH, Aitken SN (2013) Conservation and the genetics of populations. John Wiley & Sons, New York

Álvarez-Cordero E (1996) Biology and conservation of the Harpy Eagle in Venezuela and Panamá. Dissertation, University of Florida, Florida. 212 p.

Banhos A (2009) Genética, Distribuição e Conservação do Gavião-real (*Harpia harpyja*) no Brasil. Tese de Doutorado, Instituto Nacional de Pesquisas da Amazônia, Manaus, Amazonas. 153p.

Banhos A, Hrbek T, Sanaiotti TM, Farias IP (2016) Reduction of genetic diversity of the Harpy Eagle in Brazilian tropical forests. PLoS One 11:1-12. <https://doi.org/10.1371/journal.pone.0148902>

Batalha-Filho H, Fjeldså J, Fabre PH, Miyaki CY (2013) Connections between the Atlantic and the Amazonian forest avifaunas represent distinct historical events. J of Ornithol 154:41-50. <https://doi.org/10.1007/s10336-012-0866-7>

BirdLife International (2023). Species factsheet: *Harpia harpyja*. (www.birdlife.org). Acesso em: 10/02/2023.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina Sequence Data. Bioinform 30:2114-2120. <https://doi.org/10.1093/bioinformatics/btu170>

Bouckaert RR, Drummond AJ (2017). bModelTest: Bayesian phylogenetic site model averaging and model comparison. BMC Evolut Biol 17:1-11. <https://doi.org/10.1186/s12862-017-0890-6>

Chaubey A, Shenoy S, Mathur A, Ma Z, Valencia CA, Nallamilli BRR, Szekeres E, Stansberry L, Liu R, Hegde MR (2020) Low-pass genome sequencing: validation and diagnostic utility from 409 clinical cases of low-pass genome sequencing for the detection of copy number variants to replace constitutional microarray. The J of Mol Diagn 22:823-840. <https://doi.org/10.1016/j.jmoldx.2020.03.008>

Costa LP (2003) The historical bridge between the Amazon and the Atlantic forest of Brazil: a study of molecular phylogeography with small mammals. J Biogeogr 30:71-86 <https://doi.org/10.1046/j.1365-2699.2003.00792.x>

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H (2021) Twelve years of SAMtools and BCFtools. GigaSci 10:giab008. 10.1093/gigascience/giab008

Eizirik E, Kim JH, Menotti-Raymond M, Crawshaw PG, O'Brien SJ, Johnson WE (2001) Phylogeography, population history and conservation genetics of jaguars (*Panthera onca*, Mammalia, Felidae). *Mol Eco* 10:65-79. <https://doi.org/10.1046/j.1365-294X.2001.01144.x>

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14:2611-2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 10:564-567. <https://doi.org/10.1111/j.1755-0998.2010.02847.x>

Frankham R, Ballou JD, Briscoe DA (2010) *Introduction to Conservation Genetics*. Cambridge University Press, Cambridge

Gauthier J, Mouden C, Suchan T, Alvarez N, Arrigo N, Riou C, Lemaitre C, Peterlongo P (2020) DiscoSnp-RAD: de novo detection of small variants for RAD-Seq population genomics. *PeerJ* 8:e9291 <https://doi.org/10.7717/peerj.9291>

Gilly A, Southam L, Suveges D, Kuchenbaecker K, Moore R, Melloni GE, Hatzikotoulas K, Farmaki AE, Ritchie G, Schwartzentruber J, Danecek P, Kilian B, Pollard MO, Ge X, Tsafantakis E, Dedoussis G, Zeggini E (2019) Very low-depth whole-genome sequencing in complex trait association studies. *Bioinform* 35:2555-2561. [10.1093/bioinformatics/bty1032](https://doi.org/10.1093/bioinformatics/bty1032)

Hoorn C, Wesselingh FP, Ter-Steege H, Bermudez MA, Mora A, Sevink J, Sanmartín I, Sanchez-Meseguer A, Anderson CL, Figueiredo JP, Jaramillo C, Riff D, Negri FR, Hooghiemstra H, Lundberg J, Stadler T, Särkinen T, Antonelli A (2010) Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Sci* 330:927-931. [10.1126/science.1194585](https://doi.org/10.1126/science.1194585)

Hsu YC, Hsiao YT, Kao TY, Chang JG, Shieh GS (2017) Detection of somatic mutations in exome sequencing of tumor-only samples. *Sci Rep* 7:15959. <https://doi.org/10.1038/s41598-017-14896-7>

Jenkins CN, Pimm SL, Joppa LN (2013) Global patterns of terrestrial vertebrate diversity and conservation. *Proc of the Natl Acad of Scienc* 110:E2602-E2610. <https://doi.org/10.1073/pnas.1302251110>

Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, Yi TS, Li DZ (2020) GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genom Biol* 21:241. <https://doi.org/10.1186/s13059-020-02154-5>

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinform* 24:1403-1405. <https://doi.org/10.1093/bioinformatics/btn129>

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet* 11:94. <https://doi.org/10.1186/1471-2156-11-94>

Judkins ME, Couger BM, Warren WC, Van-Den-Bussche RA (2020) A 50K SNP array reveals genetic structure for bald eagles (*Haliaeetus leucocephalus*). *Conserv Genet* 21:65-76. <https://doi.org/10.1007/s10592-019-01216-x>

Judkins ME, Roemer GW, Millsap BA, Barnes JG, Bedrosian BE, Clarke SL, Domenech R, Herring G, Lamont M, Smith BW, Stahlecker DW, Stuber MJ, Warren WC, Van-Den-Bussche RA (2023) A 37 K SNP array for the management and conservation of Golden Eagles (*Aquila chrysaetos*). *Conserv Genet* 24:391-404. <https://doi.org/10.1007/s10592-023-01508-3>

Kamvar ZN, Tabima JF, Grünwald NJ (2014) *Poppr*: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2:e281 <https://doi.org/10.7717/peerj.281>

Kuderna LFK et al. (2023) A global catalog of whole-genome diversity from 233 primate species. *Sci* 380:906-913. [1126/science.abn7829](https://doi.org/10.1126/science.abn7829)

Kulski JK (2016) Next-generation sequencing—an overview of the history, tools, and “Omic” applications. In: Kulski JK (ed) *Next generation sequencing-advances, applications and challenges*, IntechOpen, London, 10.5772/61964

Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B (2017) PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. *Mol Biol Evol* 34:772-773. [10.1093/molbev/msw260](https://doi.org/10.1093/molbev/msw260)

Larsson A (2014) AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinform* 30:3276-3278. <https://doi.org/10.1093/bioinformatics/btu531>

Latch EK, Dharmarajan G, Glaubitz JS, Rhodes OE (2006) Relative performance of bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. *Conserv Genet* 7:295-302. <https://doi.org/10.1007/s10592-005-9098-1>

Lee JH, Kweon S, Park YR (2021) Sharing genetic variants with the NGS pipeline is essential for effective genomic data sharing and reproducibility in health information exchange. *Sci Rep* 11:2268 <https://doi.org/10.1038/s41598-021-82006-9>

Lever J, Krzywinski M, Altman N (2017) Principal Component Analysis. *Nat Methods* 14:641–642. <https://doi.org/10.1038/nmeth.4346>

Li JH, Mazur CA, Berisa T, Pickrell JK (2021) Low-pass sequencing increases the power of GWAS and decreases measurement error of polygenic risk scores compared to genotyping arrays. *Genom Res* 31:529-537. 10.1101/gr.266486.120

Li Y, Sidore C, Kang HM, Boehnke M, Abecasis GR (2011) Low-coverage sequencing: implications for design of complex trait association studies. *Genom Res* 21:940-951. 10.1101/gr.117259.110

Machado AF, Ritter CD, Miranda CL, Bredin YK, Pereira MJR, Duarte L (2021) Potential mammalian species for investigating the past connections between Amazonia and the Atlantic Forest. *PloS One*, 16:e0250016. <https://doi.org/10.1371/journal.pone.0250016>

MapBiomas (2022) Relatório Anual de Desmatamento 2021. (alerta.mapbiomas.org). Acesso em: 03/02/2023.

McKenna A, Hanna M, Bancos E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genom Res* 20:1297-1303. 10.1101/gr.107524.110

Nielsen R, Korneliussen T, Albrechtsen A, Li Y, Wang J (2012) SNP calling, genotype calling, and sample allele frequency estimation from new-generation sequencing data. *PloS One* 7:e37558. <https://doi.org/10.1371/journal.pone.0037558>

Oliveira AL, Silva RS (2006) Registro de Harpia (*Harpia harpyja*) no cerrado de Tapira, Minas Gerais, Brasil. *Revista Bras de Ornitol* 14:433-434.

Oliveira-Filho A, Ratter J (1995) A study of the origin of central Brazilian forests by the analysis of plant species distribution patterns. *Edinb J of Botany* 52:141-194. 10.1017/S0960428600000949

Peterlongo P, Riou C, Drezen E, Lemaitre C (2017) DiscoSnp++: de novo detection of small variants from raw unassembled read set(s). *BioRxiv* 209965. <https://doi.org/10.1101/209965>

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genet* 155:945-959. <https://doi.org/10.1093/genetics/155.2.945>

R Core Team (2022) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>

Rambaut A (2018) FigTree. (tree.bio.ed.ac.uk/software/figtree). Acesso em: 06/12/2022.

Rambaut A, Drummond AJ, Xie D, Baele G and Suchard MA (2018) Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. *Syst Biol* 67:901-904. 10.1093/sysbio/syy032

Ruiz-Gonzalez A, Cushman SA, Madeira MJ, Randi E, Gomez-Moliner BJ (2015) Isolation by distance, resistance and/or clusters? Lessons learned from a forest-dwelling carnivore inhabiting a heterogeneous landscape. *Mol Ecol* 24:5110-5129. <https://doi.org/10.1111/mec.13392>

Rustagi N, Zhou A, Watkins WS, Gedvilaite E, Wang S, Ramesh N, Muzny D, Gibbs RA, Jorde LB, Yu F, Xing, J (2017) Extremely low-coverage whole genome sequencing in South Asians captures population genomics information. *BMC Genom* 18:1-12. <https://doi.org/10.1186/s12864-017-3767-6>

Slatko BE, Gardner AF, Ausubel FM (2018) Overview of next-generation sequencing technologies. *Curr Protoc in Mol Biol* 122:e59. <https://doi.org/10.1002/cpmb.59>

Silva DA, Melo FRD, Guimarães-Júnior IG (2013) Historical and recent records of the Harpy Eagle (*Harpia harpyja*) in the Cerrado biome of the state of Goiás, Brazil. *Revista Bras de Ornitol* 21:260-263.

Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A (2018) Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol* 4:vey016. [10.1093/ve/vey016](https://doi.org/10.1093/ve/vey016)

Turchetto-Zolet AC, Turchetto C, Zanella CM, Passaia G (2017) Marcadores moleculares na era genômica: Metodologias e aplicações. Sociedade Brasileira de Genética, Ribeirão Preto

Uricaru R, Rizk G, Lacroix V, Quillery E, Plantard O, Chikhi R, Lemaitre C, Peterlongo P (2015) Reference-free detection of isolated SNPs. *Nucl Acids Res* 43:e11. <https://doi.org/10.1093/nar/gku1187>

Vasimuddin M, Misra S, Li H, Aluru S (2019) Efficient architecture-aware acceleration of BWA-MEM for multicore systems. *IEEE Parallel and Distrib Process Symp* 314-324. [10.1109/IPDPS.2019.00041](https://doi.org/10.1109/IPDPS.2019.00041)

8. Apêndices

Apêndice 1. Local de origem e estado das amostras de gavião-real, grupo genético, idade e identificação do sexo.

Amostra	Local de origem - Estado	Grupo	Idade	Sexo
38	Mato Grosso	SAM	Filhote	F
39	Mato Grosso	SAM	Jovem	F
54	Extrema - Rondônia	SAM	Jovem	F
AC	Boca do Acre - Acre	SAM	Adulto	M
Cetas AC	Porto Acre - Acre	SAM	Jovem	F
H12	Porto Seguro - Bahia	ATF	Adulto	F
H147	Fazenda Aliança, Itagimirim - Bahia	ATF	Jovem	F
H150	Faro - Pará	NAM	Filhote	F
H159	Vila Amazônia, Parintins - Pará	SAM	Adulto	M
H161	Comunidade Granja Andirá, Mamuru, Barreirinha - Amazonas	SAM	Jovem	F
H162A	Assentamento Vila Amazônia, Parintins - Amazonas	SAM	Filhote	F
H184	Novo Aripuanã - Amazonas	SAM	Adulto	F
H211	Novo Aripuanã - Amazonas	SAM	Adulto	F
H212A	Floresta Nacional de Carajás, Parauapebas - Pará	SAM	Filhote	M
H221D	Floresta Nacional de Carajás, Parauapebas - Pará	SAM	Filhote	F
H240	Fazenda Campina, Barão de Melgaço - Mato Grosso	SAM	Filhote	F
H241B	Fazenda Campina, Barão de Melgaço - Mato Grosso	SAM	Filhote	F
H267	Tucuruí - Pará	SAM	Jovem	F
H269	Represa Santo Antônio, Porto Velho - Rondônia	SAM	Filhote	M
H290	Eunápolis, Porto Seguro - Bahia	ATF	Jovem	F
H291	Dom Eliseu - Pará	SAM	Filhote	F
H292	ZF3, Km 41, Rio Preto da Eva - Amazonas	NAM	Filhote	F
H293	Paricatuba, Iranduba - Amazonas	NAM	Adulto	F
H294	BR-313, Km 186, Capixaba - Acre	SAM	Filhote	F
H316	Floresta Estadual de Antimary, Antimary - Acre	SAM	Adulto	F
H318	Cocalzinho - Goiânia	SAM	Adulto	F
H320	Comunidade Tabatinga, Juruti - Pará	SAM	Adulto	M
H321	Presídio, CEA-Nenegeria, Altamira - Pará	SAM	Adulto	M
H322	Volta Grandre do Xingu, Juruti - Pará	SAM	Adulto	F
H323	Mojui dos Campos - Pará	SAM	Adulto	F
H324	Santarém - Pará	SAM	Jovem	M
H327	Vicinal Linha 25, Nova Mamoré - Rondônia	SAM	Filhote	M
H328	BR-174, Barreira Policia, Manaus - Amazonas	NAM	Jovem	M
H329	Cacoal - Rondônia	SAM	Filhote	F
H334	Babaçulândia - Tocantins	SAM	Adulto	F
H409	Sítio Santa Rita, Córrego Farias, Linhares - Espírito Santo	ATF	Jovem	F
H411	Reserva Biológica de Sooretama, Linhares - Espírito Santo	ATF	Adulto	F
HBABY	Cacaia do Rumo Certo, Lago Balbina, Presidente Figueiredo - Amazonas	NAM	Filhote	F
HEX 13	Reserva Biológica de Sooretama, Sooretama - Espírito Santo	ATF	Jovem	F
HZ01	Rondônia	SAM	-	F
HZ02	Britania - Goiás	SAM	-	F
HZ03	Desconhecida - Pará*	SAM	-	F
HZ04	Bahia	ATF	-	F
HZ13	Paraná	ATF	-	M
HZ14	Matupá - Mato Grosso	SAM	-	F
HZ19	Salvador - Bahia	ATF	-	F
Parintins	Parintins - Amazonas	SAM	Adulto	F
RO	Ariquemes - Rondônia	SAM	Filhote	F
Xingu	Altamira - Pará	SAM	Filhote	F

SAM, Sul da Amazônia; NAM, Norte da Amazônia; ATF, Mata Atlântica; F, Fêmea; M, Macho; -, Indeterminado, *, provável estado de origem.

Apêndice 2. Comandos utilizados para mapear as sequências individuais contra o genoma de referência do gavião-real (*Genome mapping*), para resgatar dos SNPs utilizando o GATK (*SNP calling*) e os *hard* e *soft filters* para realizar a filtragem dos bancos de dados.

Genome mapping (implementado no Shell)

```
seqtk mergepe "$fastq_path"/"$sample"_R1_001.fastq.gz  
"$fastq_path"/"$sample"_R2_001.fastq.gz | cutadapt -j "$threads" --interleaved -m "$length" -a  
file:"$adapter" -A file:"$adapter" - | bwa-mem2.avx2 mem -pt "$threads" "$reference" - |  
samtools sort -@ "$threads" -n -T tmp/part.map.sorted.TMP -O bam - | samtools fixmate -@  
"$threads" -m - "$results_mapped"/"$sample"_mapped_tmp.bam
```

```
samtools sort -@ "$threads" -T tmp/part.map.sorted.TMP -O bam  
"$results_mapped"/"$sample"_mapped_tmp.bam | samtools markdup -@ "$threads" -r -  
"$results_mapped"/"$sample"_mapped_sorted.bam
```

SNP calling (implementado no Shell)

```
picard AddOrReplaceReadGroups -I "$results_mapped"/"$sample"_mapped_sorted.bam -O  
"$results_mapped"/"$sample"_mapped_sorted_addRG.bam -RGID "$sample" -RGLB lib1 -  
RGPL illumina -RGPU unit1 -RGSF "$sample"
```

```
samtools index -@ "$threads" "$results_mapped"/"$sample"_mapped_sorted_addRG.bam
```

```
gatk HaplotypeCaller -R "$reference" -I  
"$results_mapped"/"$sample"_mapped_sorted_addRG.bam -ERC BP_RESOLUTION -O  
"$results_vcf"/"$sample"_raw_snps_indels.gvcf.gz
```

```
gatk IndexFeatureFile -I "$results_vcf"/"$sample"_raw_snps_indels.gvcf.gz
```

```
gatk GenotypeGVCFs --include-non-variant-sites FALSE -R "$reference" -V  
"$results_vcf"/"$sample"_raw_snps_indels.gvcf.gz -O  
"$results_vcf"/"$sample"_genotyped.vcf.gz
```

Hard filter parameters

Flexible parameters for filtering SNPs following Hard filters by GATK Team and Kundera et al. 2023:

```
QD=2 #quality by depth  
MQRank=-12.5 #Mapping Quality Rank Sum Test  
FS=60 #Fisher Strand  
SOR=3 #Strand Odds Ratio  
ReadPos_snp=-8 #Read Position Rank Sum  
ReadPos_ind=-20 #Read Position Rank Sum  
snpgap=10 #distance from indels  
indelgap=5 #distance between indels  
pvalue=0.000001 #polymorphism detection  
MQ=20 #mapping quality  
QUAL=20 #base call quality  
MAF=0.05 #minimum allele frequency  
mindepth=0.8 #min coverage mean for site  
maxdepth=300 #max coverage mean for site  
missingdataSITE=0.9 #threshold for missing data amount by site
```

```
MIN_HET_AD=0.25 #minimum frequency of rare allele at het site  
# in Kundera et al. 2023 calculated modal coverage (excluding 0s) using mosdepth (v 0.2.9),  
and then MIN_COV was 1/3x mode and MAX_COV 2x mode  
MIN_COV=2 #minimum coverage per site  
MAX_COV=100 #maximum coverage per site
```

Hard filters (implementado no Shell)

```
bcftools filter -e "TYPE='snp' & (GT='het' & FMT/AD[:]<$MIN_HET_AD) | AC>2 | QD<$QD | FS>$FS | SOR>$SOR | MQRankSum<$MQRank | ReadPosRankSum<$ReadPos_snp | MQ<$MQ | FMT/DP<$MIN_COV | FMT/DP>$MAX_COV" --SnpGap $snpgap "$results_vcf"/"$sample"_genotyped.bcf.gz | bcftools filter -e "TYPE='indel' & (GT='het' & FMT/AD[:]<$MIN_HET_AD) | QD<$QD | FS>$FS | SOR>$SOR | ReadPosRankSum<$ReadPos_ind | MQ<$MQ | FMT/DP<$MIN_COV | FMT/DP>$MAX_COV" - -Ob -o "$results_vcf_filtered"/"$sample"_genotyped_hardfilters.bcf.gz
```

Soft filters (implementado no R)

```
source("https://raw.githubusercontent.com/legalLab/protocols-scripts/master/scripts/vcf2others.R")
```

```
vcf <- read.vcfR("harpia_all.vcf")
```

```
indivs <- read.table("indivs", header = TRUE)$id %>%  
  as.character()
```

```
vcf1 <- vcf_sub_indivs(vcf, indivs, whitelist = FALSE) %>%  
  vcf_filter_missing_indivs(.9) %>%  
  vcf_filter_maf(.03) %>%  
  vcf_filter_coverage(5) %>%  
  vcf_filter_oneSNP() %>%  
  vcf_filter_missingness(.5)
```

CONCLUSÃO GERAL

Os estudos da Genética da Conservação são essenciais para desenvolver estratégias de manejo eficazes, garantindo a preservação da diversidade genética e a sobrevivência de espécies a longo prazo. Neste estudo foram desenvolvidos novos marcadores microsatélites para o gavião-real, os quais podem ser aplicados em toda a população *ex situ* presente em zoológicos e criadouros no Brasil para fornecer mais informações sobre os indivíduos cativos ao *Studbook* da espécie. Assim, é possível auxiliar na definição de estratégias para o manejo populacional adequado, como realizar a translocação de indivíduos entre cativeiros e direcionar a formação de novos casais a partir de dados históricos e genéticos para aumentar a diversidade genética, manter a estrutura genética existente e evitar a consanguinidade.

Com base nas informações genéticas obtidas neste estudo foi possível elucidar a presença de estrutura genética para o gavião-real no Brasil a partir de marcadores moleculares biparentais (SNPs). Apesar desses marcadores resgatarem um baixo sinal de estrutura, eles encontraram padrões semelhantes ao que tem se observado para a espécie com o DNA mitocondrial, com a população localizada ao Norte do Rio Amazonas apresentando ser mais distante geneticamente das outras populações. Ademais, essa população possui uma alta diversidade genética e contém maior número de *loci* polimórficos, demandando atenção especial para sua conservação e manutenção da diversidade genética local. Por outro lado, as populações situadas ao Sul do Rio Amazonas e na Mata Atlântica apresentaram maior similaridade genética entre si e menor diversidade genética. Portanto, é crucial implementar medidas de manejo dessas populações visando manter e maximizar a diversidade genética, antes que ela diminua ainda mais. Tais ações pode incluir a proteção de seus territórios, uma vez que existem indícios de perda de diversidade genética devido aos altos índices de desmatamento na região.

Neste estudo, também foi possível sugerir o *pipeline* mais adequado para a detecção e recuperação de variantes a partir de um sequenciamento com baixa cobertura. Considerando que estudos anteriores sobre o gavião-real não apresentavam uma estrutura genética clara para marcadores biparentais, e que os programas e *pipelines* de bioinformática existentes foram desenvolvidos para lidar com dados de maior cobertura, a inferência estatística para identificar variantes verdadeiras torna-se mais desafiadora em um sequenciamento com baixa cobertura, dificultando a obtenção de informações sobre a estrutura genética. Portanto, foi necessário testar diferentes *pipelines* e encontrar aquele que melhor se adequasse ao nosso conjunto de dados, que no caso foi o *DiscoSnpRad*. Por fim, esse trabalho gerou novas informações e ferramentas genéticas para auxiliar no manejo e conservação do gavião-real *in situ* e *ex situ*.