

UNIVERSIDADE FEDERAL DE MINAS GERAIS  
INSTITUTO DE CIÊNCIAS BIOLÓGICAS  
DEPARTAMENTO DE GENÉTICA, ECOLOGIA E  
EVOLUÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA



TESE DE DOUTORADO

**Aplicações da genética na conservação e combate ao  
tráfico de animais silvestres**

ALUNA: ANA LUIZA LEMOS QUEIROZ

ORIENTADOR: PROF. DR. EVANGUEDES KALAPOTHAKIS

BELO HORIZONTE

Novembro – 2021

Ana Luiza Lemos Queiroz

## **Aplicações da genética na conservação e combate ao tráfico de animais silvestres**

Tese apresentada ao curso de Doutorado em Genética do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais como requisito parcial para obtenção de título de Doutora em Genética

Orientador: Prof. Dr. Evanguedes Kalapothakis

Belo Horizonte

2021

043

Queiroz, Ana Luiza Lemos.

Aplicações da genética na conservação e combate ao tráfico de animais silvestres [manuscrito] / Ana Luiza Lemos Queiroz. – 2021.

181 f. : il. ; 29,5 cm.

Orientador: Prof. Dr. Evanguedes Kalapothakis.

Tese (doutorado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Programa de Pós-Graduação em Genética.

1. Genética. 2. Animais Selvagens. 3. Tráfico. 4. Biologia Sintética. I. Kalapothakis, Evanguedes. II. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. III. Título.

CDU: 575



UNIVERSIDADE FEDERAL DE MINAS GERAIS  
Instituto de Ciências Biológicas  
Programa de Pós-Graduação em Genética

### FOLHA DE APROVAÇÃO

"Aplicações da genética na conservação e combate ao tráfico de animais silvestres"

Ana Luiza Lemos Queiroz

Tese aprovada pela banca examinadora constituída pelos Professores:

Evanguedes Kalapothakis  
UFMG

Daniel Ambrózio da Rocha Vilela  
IBAMA

Fábio José Viana Costa  
Polícia Federal

Alvaro Cantini Nunes  
UFMG

Fernanda Antunes Carvalho  
UFMG

Belo Horizonte, 16 de novembro de 2021.



Documento assinado eletronicamente por **Evanguedes Kalapothakis, Presidente**, em 16/11/2021, às 17:51, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Alvaro Cantini Nunes, Professor do Magistério Superior**, em 17/11/2021, às 09:47, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Daniel Ambrózio da Rocha Vilela, Usuário Externo**, em 17/11/2021, às 11:06, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Fábio Jose Viana Costa, Usuário Externo**, em 22/11/2021, às 10:27, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).





Documento assinado eletronicamente por **Fernanda Antunes Carvalho, Professora do Magistério Superior**, em 04/05/2022, às 21:14, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site [https://sei.ufmg.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **1080582** e o código CRC **9881EEEE**.

## AGRADECIMENTOS

Ao Professor Evanguedes Kalapothakis e ao programa de pós-graduação em genética da UFMG pela oportunidade.

Aos colegas do LBMM, pela convivência diária enquanto foi permitido e, em especial ao Bruno, Nazaré, Adriana e Grécia pelas discussões a respeito do projeto.

Ao professor George Church e todos os membros do Church lab, por me aceitarem em seu laboratório me permitindo um enorme crescimento profissional, em especial à Eriona Hysolli, que esteve comigo todos os dias durante meu período no exterior.

À Chao-ting Wu, por ter aberto as portas do seu laboratório para a realização dos nossos experimentos e ser sempre presente nas discussões, assim como a todos os seus alunos que de alguma forma participaram e colaboraram com o nosso projeto.

Ao Henrique, por ter sido o principal suporte e apoiador durante todo esse processo, e com quem eu aprendi tantos valores profissionais e pessoais.

À equipe do CETAS-BH, em especial à Erika, Cecília e Daniel, e ao perito criminal federal Rodrigo Mayrink, por cooperarem com este projeto de inúmeras formas, desde a obtenção de amostras até as discussões.

À CAPES pelo financiamento deste trabalho através do edital pró-forenses, à minha bolsa de estudos no país e no exterior. Ao CNPq e Fapemig pelo financiamento de equipamentos utilizados neste estudo. Ao Revive & Restore pelo financiamento da pesquisa desenvolvida no exterior.

À Aline Chagas, Isabella Pena e Prof. Francisco Lobo por terem sido da banca de qualificação, contribuindo com melhorias para a realização desta tese.

Aos orientadores que tanto me ajudaram nesses anos de pesquisa até aqui: Vitor, Prof. Vera Lucia, Prof. Mauro, Calliope, Marcia e Luisa, em especial ao Elisão, por ter sido o primeiro a me ensinar sobre ciência e me incentivar e guiar no começo deste caminho.

À Raíssa e à Dani, por serem sempre tão prestativas e dispostas a nos ajudar.

À Izinara Rosse, pela ajuda fundamental no processamento dos dados.

À família e amigos, por estarem sempre comigo.

Aos responsáveis pelo CEPAD (centro de processamento de alto desempenho) do ICB/UFMG, pela manutenção das ferramentas necessárias para as análises.

À Sociedade Brasileira de Ciências Forenses (SBCF) e Rede Mineira de Ciências Forenses, que durante este período proporcionaram inúmeras oportunidades de aprendizado e divulgação deste trabalho.

## RESUMO

Diferentes abordagens na área da genética são capazes de auxiliar a conservação de espécies. Uma delas é através da identificação de animais silvestres ilegais em cativeiro. Além disto, recentemente, a resolução do CONAMA nº 487/2018 trouxe a obrigatoriedade de genotipagem de matrizes em criatórios para verificação de filiação para algumas espécies, sendo *Saltator similis* a espécie de maior prioridade devido à demanda do tráfico. Assim, este trabalho teve como principal foco o desenvolvimento de um teste de filiação para a espécie baseado em microssatélites e com uso da tecnologia de sequenciamento massivo de DNA (NGS) tanto para o desenvolvimento como para a análise dos marcadores. O teste de paternidade foi capaz de identificar os progenitores mesmo quando utilizado menos de dez *loci*, indicando que os *loci* desenvolvidos neste trabalho são bastante informativos e estão aptos a compor o painel de genotipagem, além de que também poderão ser usados em projetos de manejo e conservação. Adicionalmente, o mtDNA completo da espécie também foi sequenciado e anotado, e poderá auxiliar investigações de crimes contra a espécie quando a utilização apenas do COI não for informativa o suficiente e também em estudos evolutivos.

Outra abordagem é a biologia sintética, que pode auxiliar na manutenção de espécies reduzindo o risco de extinção e até mesmo criar espécies próximas às extintas, mas que mantenham a função ecológica no ecossistema. Um destes projetos pioneiros de *de-extinction*, no qual parte deste trabalho foi realizado, é a criação de um animal com genoma de elefante-asiático (espécie modelo) e mamute-lanoso, que mantenha as funções da espécie extinta. Para isso, é necessário conhecimento dos genomas das duas espécies e realização de edições gênicas. Assim, foram desenhados sgRNA para *knockout* de genes de elefante-asiático e *primers* para confirmação destas edições, além de padronização de protocolo de FISH para melhor entendimento da estrutura nuclear de mamute-lanoso. Concomitantemente, este tipo de projeto também costuma agir na preservação da espécie modelo, que, no caso do elefante-asiático, tem o vírus EEHV como um fator complicador na sua conservação. Ainda pouco se sabe sobre os mecanismos de patogenicidade deste vírus e possíveis tratamentos pois seu cultivo em laboratório ainda não foi possível. Visando possibilitar estes estudos, neste trabalho foram amplificados e extraídos fragmentos de EEHV para posterior construção do material genético do vírus que poderá ser utilizado em experimentos.

**Palavras-chave:** tráfico de animais silvestres, *Saltator similis*, microsatélites, *de-extinction*, biologia sintética.

## ABSTRACT

Different approaches in the area of genetics are able to help species conservation. One is through the identification of wild animals maintained illegally in captivity. Recently the CONAMA resolution No. 487/2018 made the genotyping mandatory in breeding facilities to verify the affiliation of some species, being *Saltator similis* the species with the highest priority due to the demand for trafficking. Thus, this work had as main focus the development of a paternity test for the species based on microsatellites (STR) using the technology of massively parallel DNA sequencing both for the development and for the analysis of the markers sequences. The paternity test was able to identify the parents even when less than ten *loci* were used, indicating that the *loci* developed in this work are very informative and are able to compose the genotyping panel, in addition to the fact that they can also be used in management projects and conservation. Additionally, the complete mtDNA of the species has also been sequenced and annotated, and may assist investigations of crimes against the species when the use of COI alone is not informative enough and also in evolutionary studies.

Another approach is synthetic biology, which can help maintain species by reducing the risk of extinction and even creating proxies of extinct species that maintain the ecological function in the ecosystem. One of these pioneering *de-extinction* projects, in which part of this work was carried out, is the creation of a hybrid of Asian elephant (template species) and woolly mammoth that maintains the ecological functions of the extinct one. For this, it is necessary to know the genomes of both species and perform gene editing. sgRNAs were designed to *knockout* Asian elephant genes and *primers* to confirm these edits, as well as FISH protocol standardization for a better understanding of the nuclear structure of woolly mammoth. At the same time, this type of project usually helps in the preservation of template species. In the case of the Asian elephant, EEHV infections have been a complicating factor in its conservation. Little is known about the pathogenicity mechanisms of this virus and possible treatments because its cultivation in the laboratory has not yet been possible. In order to make these studies possible, in this work EEHV fragments were amplified and extracted for later construction of the virus genetic material that could be used in experiments.

**Keywords:** wildlife trafficking, *Saltator similis*, STR, *de-extinction*, synthetic biology.

Considerações iniciais

## Lista de ilustrações

Figura 1- Indivíduo da espécie <i>Saltator similis</i> . .....	21
Figura 2- Área de ocorrência de <i>S. similis</i> . .....	21
Figura 3- Processo de anilhamento de passeriforme.....	25
Figura 4- Unidades da federação que possuíam Centros de Triagem de Animais Silvestres (Cetas) sob gestão do Ibama em 2016. ....	28
Figura 5- Exemplos da classificação dos microssatélites quanto ao motivo da repetição. ....	33
Figura 6- Exemplos de microssatélites quanto à forma em que aparecem no DNA.....	34
Figura 7- Alterações do tamanho do microssatélite.....	36
Figura 8- Representação simplificada de sequenciamento na plataforma MiSeq.....	41
Figura 9- Organização do mtDNA em mamíferos e em aves. ....	43
Figura 10- Representação esquemática da metodologia utilizada no sequenciamento do DNA mitocondrial. ....	47
Figura 11- Representação esquemática da metodologia utilizada para desenvolvimento do teste de filiação. ....	48
Figura 12- Amostra de sangue de indivíduo de criatório em papel FTA.....	50
Figura 13- Representação esquemática do protocolo de amplificação das sequências alvo para sequenciamento. ....	54
Figura 14- Representação esquemática da estratégia de genotipagem utilizada.....	56
Figura 15- Qualidade das bases das reads R1 do sequenciamento do DNA total.....	60
Figura 16- Resultado da PCR com os pares de primers sintetizados sem adaptadores em gel de agarose.....	62
Figura 17- Resultado da PCR com os conjuntos de primers com adaptadores.....	63
Figura 18- Qualidade das bases das reads R1 do sequenciamento do teste 1 no MiSeq.....	67
Figura 19- Qualidade das bases das reads R2 do sequenciamento do teste 1 no MiSeq.....	68
Figura 20- Exemplo de mutação na região microssatélite.....	73
Figura 21- Qualidade das bases das reads do sequenciamento do teste 2 no MiSeq.....	74
Figura 22- Heredogramas com as amostras de famílias de <i>S. similis</i> oriundas de criatório. ....	81
Figura 23- Qualidade das bases das reads do sequenciamento do teste com amostras de criatórios.....	83
Figura 24- Alelos encontrados para o locus L11. ....	89
Figura 25- Alelos encontrados para o locus L14. ....	90
Figura 26- Alelos encontrados para o locus L15. ....	90
Figura 27- Alelos encontrados para o locus L16. ....	90
Figura 28- Alelos encontrados para o locus L19.....	91
Figura 29- Alelos encontrados para o locus L20.....	91
Figura 30- Alelos encontrados para o locus L23.....	91
Figura 31- Alelos encontrados para o locus L24.....	92
Figura 32- Alelos encontrados para o locus L26.....	92
Figura 33- Alelos encontrados para o locus L29.....	92
Figura 34- Alelos encontrados para o locus L30.....	93
Figura 35- Alelos encontrados para o locus L32.....	93
Figura 36- Alelos encontrados para o locus L35.....	93
Figura 37- Alelos encontrados para o locus L36.....	94
Figura 38- Alelos encontrados para o locus L37.....	94
Figura 39- Alelos encontrados para o locus L38.....	94
Figura 40- Alelos encontrados para o locus L39.....	95
Figura 41- Alelos encontrados para o locus L49.....	95

Figura 42- Alelos encontrados para o locus L57. ....	95
Figura 43- <i>Mismatches</i> nas sequências de pai (AC1) e filhote (AC16) no locus L29. A .....	107
Figura 44- Tamanho das reads dos controles negativos. ....	108
Figura 45- Reads do controle negativo do teste 2 mapeadas contra a referência do L37. ....	108
Figura 46- Biologia sintética na restauração da biodiversidade. ....	115
Figura 47- Mecanismo de edição gênica com sistema CRISPR-Cas9.....	118
Figura 48- Área de ocorrência de <i>Elephas maximus</i> . ....	119
Figura 49- Metodologia de FISH.....	126
Figura 50- Centrômeros de elefante-asiático. ....	132
Figura 51- Controle negativo (células humanas). ....	133
Figura 52- Marcação de núcleo em tecido com diferentes cores.....	134
Figura 53- Centrômeros e TO-PRO diluído em água em fígado de elefante-asiático. ....	135
Figura 54- Centrômeros e TO-PRO diluído em DAPI em fígado de elefante-asiático.....	136
Figura 55- Centrômeros e TO-PRO diluído em DAPI em células de elefante-asiático. ....	136
Figura 56-FISH em fígado de elefante-asiático após pepsina.....	137
Figura 57-Centrômeros em fígado de elefante-asiático com e sem TrueBlack. ....	138
Figura 58- FISH em fígado de elefante-asiático após glicina e tris. ....	139
Figura 59- imunofluorescência com lamin B1 (Cy5) em fígado de elefante-asiático.....	140
Figura 60- FISH após imunofluorescência com lamin B1 (pacific blue) em fígado de elefante-asiático.....	141
Figura 61- Imunofluorescência com lamin B1 (pacific blue) com fixação extra ao final do protocolo em fígado de elefante-asiático. ....	142
Figura 62- Centrômero, Cromossomo 11 e cromossomo 24 de elefante-asiático.....	143
Figura 63- Marcação do cromossomo 11 e cromossomo 24 de elefante-asiático.....	144
Figura 64- Gradiente de temperatura de anelamento para o fragmento 3 de EEHV.....	146
Figura 65- Gradiente de temperatura de anelamento para o fragmento 8 de EEHV.....	146
Figura 66- Gradiente de temperatura de anelamento para os fragmentos 12 e 15 de EEHV....	147

## Lista de tabelas

Tabela 1- Número de animais recebidos pelos Cetac sob gestão do Ibama entre 2002 e 2014.....	28
Tabela 2- Reagentes utilizados para montagem da PCR. ....	55
Tabela 3- Programação do termociclador. ....	55
Tabela 4- Valores de qualidade Phred e probabilidade de acurácia da base. ....	57
Tabela 5- Quantidade de microssatélites encontrados e primers sintetizados .....	62
Tabela 6- Conjuntos de primers sintetizados e suas características.....	64
Tabela 7- Similaridade da sequência referência de cada locus contra o banco de dados GenBank. ....	66
Tabela 8- Reads geradas no teste 1 no MiSeq. ....	69
Tabela 9- Resultado da análise do teste 1 no MiSeq. ....	72
Tabela 10- Reads geradas no teste 2 no MiSeq. ....	75
Tabela 11- Quantidade de DNA extraído das amostras de criatório e quantidade utilizada na PCR. ....	82
Tabela 12- Reads geradas no teste 3 com as amostras de criatório. ....	84
Tabela 13- Número de alelos por locus e seus tamanhos. ....	89
Tabela 14- Frequência de alelos nulos.....	97
Tabela 15- Frequências alélicas e estatísticas de todos os indivíduos sequenciados.....	98
Tabela 16- Resumo do resultado da análise de paternidade das amostras de criatório.. ..	101
Tabela 17- Genótipo de filhotes e progenitores.....	102
Tabela 18- Reagentes utilizados para montagem das PCR.....	123
Tabela 19- Programação do termociclador. ....	123
Tabela 20- Complexo CRISPR-Cas9 para knockout dos genes selecionados.....	129
Tabela 21- Conjuntos de primers desenhados para amplificação após knockout para sequenciamento.....	131



## **Lista de gráficos**

<b>Gráfico 1 - Espécies cadastradas no SISPASS.....</b>	<b>23</b>
<b>Gráfico 2- Exemplares de S. similis cadastrados no SISPASS por estado.....</b>	<b>23</b>
<b>Gráfico 3- Criatórios de S. similis cadastrados no SISPASS por estado. ....</b>	<b>24</b>
<b>Gráfico 4- Percentual de animais recebidos nos Cetas de acordo com o táxon entre 2002-2009 e 2010-2014.....</b>	<b>29</b>

## Lista de abreviaturas

%GC - conteúdo de Guanina e Citosina

°C - graus Celsius

A - adenina

Blastn - *Basic Local Alignment Search Tool- Nucleotide*

BOLD - *Barcode Of Life Database*

C - citosina

CETAS - Centro de Triagem de Animais Silvestres

CEUA - Comissão de Ética no Uso de Animais

cm - centímetros

COI - citocromo oxidase I

CONAMA - Conselho Nacional do Meio Ambiente

CRISPR - *Clustered regularly interspaced short palindromic repeats*/repetições palindrômicas curtas agrupadas intercaladas regularmente

CytB - citocromo b

DAPI - 4',6'-diamino-2-fenil-indol

DMSO - Dimetilsulfóxido

DNA - ácido desoxirribonucleico

EBOV- Ebola vírus humano

EDTA - ácido etilenodiamino tetra-acético

EEHV - *Elephant Endotheliotropic Herpesvirus*

EHW- equilíbrio de Hardy-Weinberg

FBS - soro fetal bovino

FISH - *Fluorescence in situ hybridization*

G - guanina

GB - gigabyte

HCl - ácido clorídrico

He - heterozigosidade esperada

HIV- vírus da imunodeficiência humana

Ho - heterozigosidade observada

IBAMA - Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis

IBDN - Instituto Brasileiro de Desenvolvimento Florestal

IEF - Instituto Estadual de Florestas

ISAG - Sociedade Internacional de Genética Animal

IUCN - *International Union for Conservation of Nature*

KCl- cloreto de potássio

LOD - log natural da razão de verossimilhança geral

M - molar

MAPA – Ministério da Agricultura e Pecuária

Mb - megabyte

mg - miligrama

MgCl<sub>2</sub> - cloreto de magnésio

ml - microlitro

mL – mililitro

mM - milimolar

mtDNA - DNA mitocondrial

NaCl - cloreto de sódio

NGS - Sequenciamento de Nova Geração

PAM - *Protospacer adjacent motif*

pb - pares de bases

PBS - *Phosphate buffered saline*

PCR - Reação em cadeia da polimerase

pH - Potencial Hidrogeniônico

PIC - conteúdo de informação polimórfica

pmol - picomolar

rpm – rotações por minuto

SARS-Cov-1 - coronavírus da síndrome respiratória aguda grave 1

SARS-Cov-2 - coronavírus da síndrome respiratória aguda grave 2

SDS - Dodecil sulfato de sódio

SEB -*Stain Extraction Buffer*

sgRNA – *Single guide RNA* /RNA guia

SisGen - Sistema Nacional de Gestão do Patrimônio Genético e do Conhecimento Tradicional Associado

SISPASS - Sistema de Controle e Monitoramento da Atividade de Criação Amadora de Pássaros

SIV - vírus da imunodeficiência símia

SNP - *Single Nucleotide Polymorphism*

SNV – *Single Nucleotide Variation*

SpCas9 – *Streptococcus pyogenes Cas9*

SSR - *Simple Sequence Repeats*

SSRP - *Simple Sequence Repeat Polymorphism*

STR - *Short Tandem Repeat*

T - timina

TBE 0,5% - 1.1 M Tris, 900 mM Borato, 25 mM EDTA, pH 8,3

TE - Tris 10 mM, EDTA 1 mM, pH 8

TEN9 - 50 mM Tris-HCl, 100 mM EDTA, 200 mM NaCl, pH 9

TM- temperatura de *melting*

TSA - Trichostatin A

US\$ - dólar americano

UV – Ultravioleta

## Índice

<b>Considerações iniciais</b> .....	<b>18</b>
<b>1.Capítulo 1- Ferramentas genéticas para o combate ao tráfico de <i>Saltator similis</i> (Aves, Passeriformes)</b> .....	<b>20</b>
1.1 Introdução .....	20
1.1.2 <i>Saltator similis</i> .....	20
1.1.2 Criação de <i>S. similis</i> em cativeiro.....	21
1.1.3 Tráfico de animais silvestres.....	25
1.1.4 Tráfico de <i>Saltator similis</i> .....	29
1.1.5 Tráfico de animais silvestres e surgimento novas doenças .....	31
1.1.6 Microsatélites .....	32
1.1.7 Sequenciamento massivo de DNA.....	37
1.1.8 DNA mitocondrial .....	42
1.2 Justificativa e relevância do projeto .....	45
1.3 Objetivo .....	46
1.3.1 Objetivos específicos .....	46
1.4 Metodologia .....	47
1.4.1 Fluxograma do trabalho .....	47
1.4.2 Obtenção das amostras .....	48
1.4.3 Extração de DNA.....	50
1.4.4 Sequenciamentos no MiSeq .....	52
1.4.5 Montagem de <i>contigs</i> .....	53
1.4.6 Anotação do mtDNA.....	53
1.4.7 Busca por microsatélites .....	53
1.4.8 <i>Primers</i> .....	54
1.4.9 PCRs para padronização dos pares de <i>primers</i> .....	54
1.4.10 Análise do sequenciamento de microsatélites e genotipagem .....	56
1.4.11 Análises genotípicas e de paternidade .....	58
1.5 Resultados e discussão .....	60
1.5.1 Teste de filiação de <i>S. similis</i> .....	60
1.5.2 DNA mitocondrial de <i>Saltator similis</i> .....	109
<b>1.6. Conclusão</b> .....	<b>112</b>
<b>2. Capítulo 2 - Biologia sintética: uma nova abordagem para a conservação de espécies ameaçadas de extinção</b> .....	<b>114</b>
2.1 Introdução .....	114
2.1.1 <i>De-extinction</i> .....	115
2.1.2 CRISPR-Cas9.....	117

2.1.3 <i>Elephas maximus</i> .....	118
2.1.4 Herpesvírus endoteliotrópico de elefante (EEHV) .....	119
2.2 Justificativa e relevância do projeto .....	121
2.3 Objetivo .....	122
2.3.1 Objetivos específicos .....	122
2.4 Metodologia .....	123
2.4.1 Fragmentos de EEHV .....	123
2.4.2 Edição gênica .....	124
2.4.3 <i>Fluorescence in situ hybridization</i> (FISH) .....	125
2.5 Resultados e discussão .....	129
2.5.1 Edição gênica .....	129
2.5.2 Padronização do protocolo de FISH.....	132
2.5.3 Herpesvírus endoteliotrópico de elefante .....	145
2.6 Conclusão .....	148
2.6.1 Conclusão FISH .....	148
2.6.2 Conclusão edição gênica.....	148
2.6.3 Conclusão EEHV .....	148
Considerações finais .....	149
Referências.....	150
Referências capítulo 1 .....	150
Referências capítulo 2 .....	166
ANEXO I – Tamanho das <i>reads</i> brutas geradas nos sequenciamentos.....	171

## Considerações iniciais

O presente trabalho tem como objetivo o estudo de como o uso de ferramentas genéticas pode auxiliar na identificação de espécies e combate ao tráfico de animais silvestres, em especial o *Saltator similis* (d'Orbigny & Lafresneye 1837), conhecido popularmente como trinca-ferro. Esta é uma das principais espécies traficadas no Brasil, estando frequentemente no topo das listas de espécies apreendidas nas fiscalizações e também frequentemente apresentar fraudes nas suas anilhas oficiais, indicando a ocorrência de tentativas de conferir uma aparência legal à um animal ilegalmente em cativeiro. Os testes de paternidade podem ser bastante úteis nas fiscalizações de animais em cativeiro para verificar sua legalidade, e, com isso, no primeiro capítulo desta tese é abordado o desenvolvimento de um teste de paternidade para a espécie em questão com o uso de regiões repetitivas do DNA conhecidas como microssatélites. Além disto, também é abordado o sequenciamento do DNA mitocondrial da espécie, uma vez que este pode ser utilizado para a identificação da espécie quando não é possível que esta seja feita com as características morfológicas.

Além do seu uso em testes de paternidade, os microssatélites são utilizados em estudos de genética de populações, e permitem diferenciar populações de diferentes localidades quando há estruturação genética. Também são comumente utilizados para avaliar a diversidade genética entre os indivíduos, fator de extrema importância para o planejamento de manejo de espécies e em projetos de conservação, que, infelizmente, não é raro que se iniciem quando a espécie alvo já se encontra em algum nível de ameaça, com número limitado de indivíduos e diversidade genética comprometida. Como parte dos trabalhos desenvolvidos pelo grupo de pesquisa, os microssatélites foram utilizados para avaliar a diversidade genética de *Sporophila maximiliani* (Cabanis 1851), ave popularmente conhecida como bicudo. A espécie é classificada pela lista vermelha da IUCN como ameaçada de extinção, além de que atualmente é dificilmente encontrada na natureza. Porém, ainda há um número razoável de indivíduos em cativeiro e projetos de reintrodução na natureza. Com isso, surge o questionamento sobre como avaliar os indivíduos a serem escolhidos para a reintrodução e se a diversidade genética com o uso de marcadores microssatélites seria a única abordagem genética a ser utilizada, ou se deveria ser implementado também o estudo de genes específicos.

Adicionalmente, quando se refere a projetos de conservação de espécies, estes são comumente iniciados de forma tardia, quando a espécie normalmente já tem número limitado de indivíduos e diversidade genética comprometida, podendo resultar em menor adaptação a diversidades ambientais e susceptibilidade a doenças, dentre outros. Assim, novas metodologias vêm sendo desenvolvidas visando contornar essas dificuldades, sendo a biologia sintética uma das principais abordagens. Com isso, como parte da formação de nível doutorado e visando solucionar questionamentos pertinentes do grupo de pesquisa em relação a outras espécies animais em estudo, o segundo capítulo desta tese traz a pesquisa desenvolvida no exterior, que se trata da participação em um projeto ainda em fase inicial que visa contornar algumas das dificuldades encontradas nos projetos de conservação de espécies. Dentre as abordagens do segundo capítulo, está o uso de edição gênica para posterior avaliação de fenótipo. É válido ressaltar que o projeto em questão é composto por diversas abordagens e se trata de um projeto pioneiro e de longo prazo, e que cuja participação como parte da formação ocorreu por apenas alguns meses e foi interrompido abruptamente devido à pandemia do novo coronavírus.



## 1. Capítulo 1- Ferramentas genéticas para o combate ao tráfico de *Saltator similis* (Aves, Passeriformes)

### 1.1 Introdução

#### 1.1.2 *Saltator similis*

*Saltator similis* (d'Orbigny & Lafresneye 1837), conhecido popularmente como trinca-ferro-verdadeiro, pixarro, trinca-ferro-de-asa-verde e bico-de-ferro, devido ao seu bico negro e forte, é uma ave que pertence à Ordem Passeriformes e à Família Thraupidae. Sem dimorfismo sexual, apresenta dorso verde, cauda e lados da cabeça acinzentados (figura 1), sendo estas características bastante similares ao *Saltator maximus* (temperaviola), porém, possui listra branca superciliar mais comprida e garganta toda branca. Mede cerca de 20 cm e se alimenta de insetos e frutos e habita principalmente beiras de matas e clareiras. Sua ocorrência se dá na parte central, região sul, sudeste e parte do nordeste brasileiro, alcançando também países como Argentina, Bolívia, Paraguai e Uruguai (figura 2), ocorrendo também em áreas de conservação ambiental. O estudo de genética populacional de Ferreira (2012) utilizou amostras coletadas em diferentes populações de *S. similis* nos estados de São Paulo, Mato Grosso do Sul, Minas Gerais e Bahia e não encontrou estruturação genética entre essas áreas de ocorrência. A espécie ainda não se encontra em nenhuma lista de espécies ameaçadas de extinção, porém, é sabido que sua população está em decréscimo (BirdLife International 2018). Entre os principais motivos para isto, destaca-se a perda de habitats e a captura ilegal (UFV 2009, Manhães & Loures-Ribeiro 2011). No momento, não há nenhum plano de recuperação para a espécie nem monitoramento sistemático, assim como não há nenhum programa de educação e conhecimento em andamento (BirdLife International 2018).



**Figura 1- Indivíduo da espécie *Saltator similis*.** O trinca-ferro mede cerca de 20 cm, apresenta cauda e lados da cabeça acinzentados, garganta toda branca, lista branca superciliar comprida e dorso verde. Foto: Sérgio Armelin



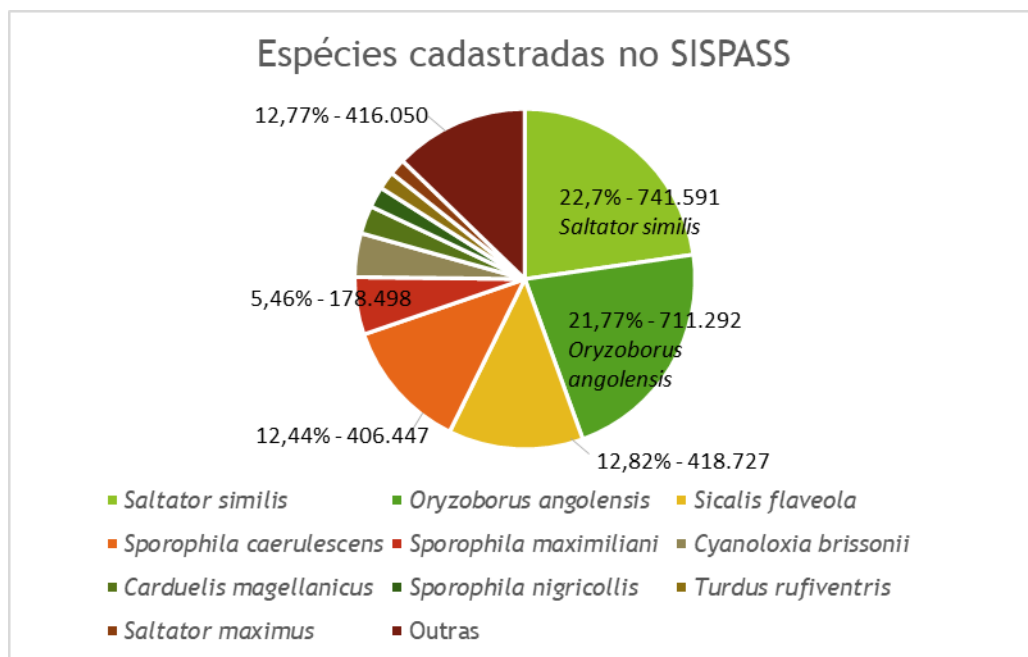
**Figura 2- Área de ocorrência de *S. similis*.** A área de ocorrência do trinca-ferro no território brasileiro abrange a parte central, região sul, sudeste e parte do nordeste, e ocorre também em outros países da América do Sul, como Argentina, Bolívia, Paraguai e Uruguai. Fonte: BirdLife International 2018.

#### 1.1.2 Criação de *S. similis* em cativeiro

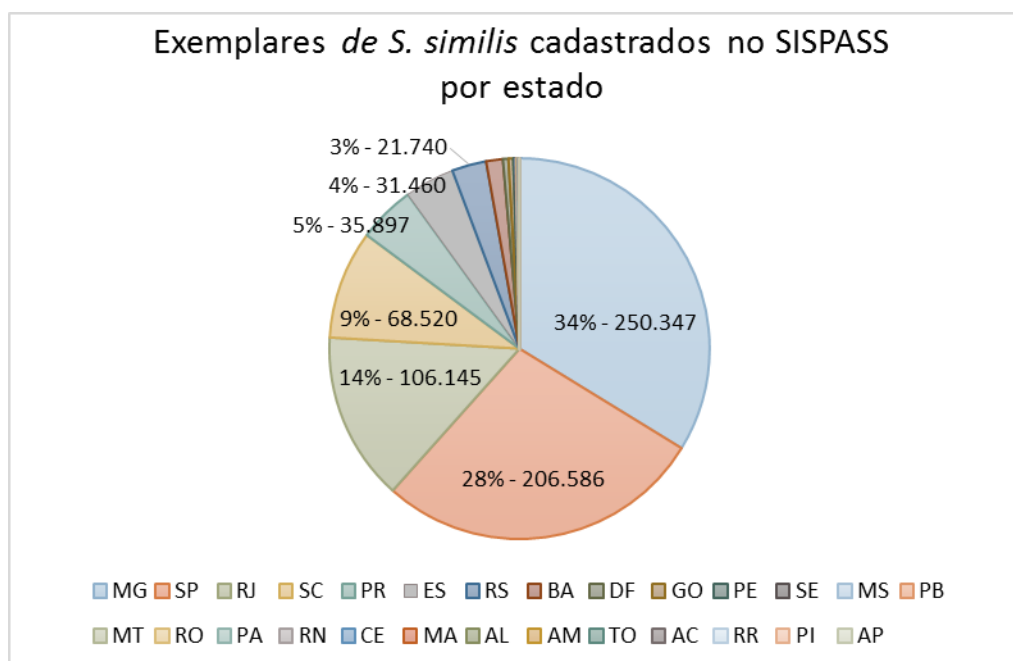
A criação de espécies de aves nativas da ordem Passeriformes teve sua primeira regulamentação no Brasil com a Portaria do Instituto Brasileiro de Desenvolvimento Florestal (IBDN) n. 3.175-DN, de 06 de dezembro de 1972. Esta portaria previa a necessidade de anilhamento, porém sem estabelecer padronização. Com a Portaria Ibama n. 631, de 18 de março de 1991, foram estabelecidas as espécies cuja criação em cativeiro era permitida. Desde então, o *Saltator similis* integra as listas em todas as normatizações referentes às espécies permitidas. Atualmente, a criação é permitida desde que os

indivíduos tenham nascido em criatórios devidamente legalizados e sejam cadastrados no Sistema de Controle e Monitoramento da Atividade de Criação Amadora de Pássaros (SISPASS), além da exigência de anilhamento com anilhas oficiais.

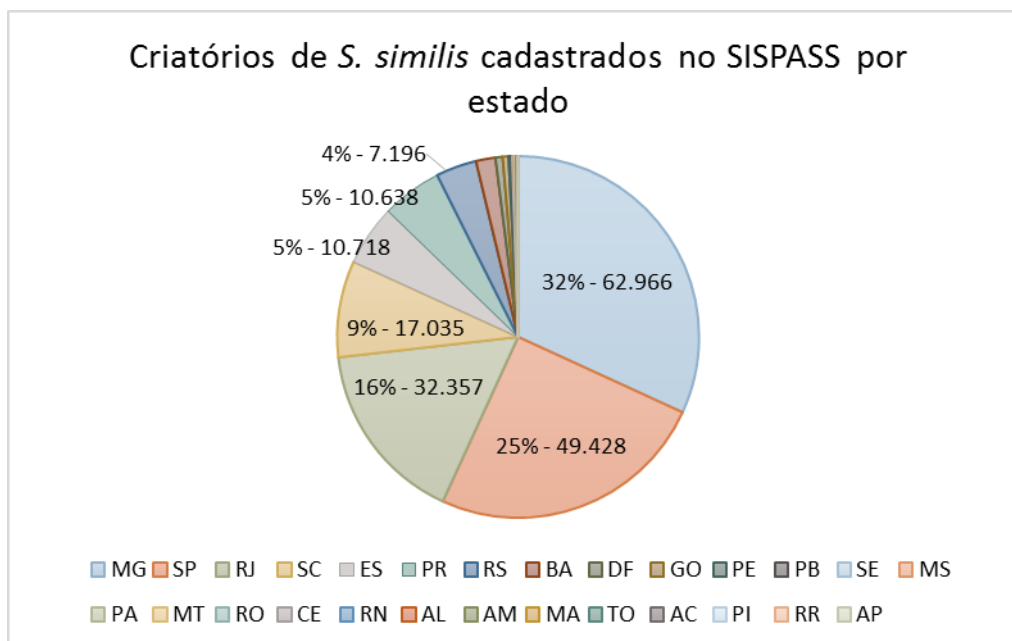
Em uma consulta realizada através do portal de sistema de informação ao cidadão (e-sic), foi verificado que, em fevereiro de 2019, constavam 3.265.973 exemplares de passeriformes cadastrados no SISPASS, sendo a maioria de *Saltator similis*, seguido por *Oryzoborus angolensis* (curió), *Sicalis flaveola* (canário-da-terra), *Sporophila caerulescens* (coleirinho) e *Sporophila maximiliani* (bicudo) (gráfico 1) (IBAMA 2019a). Em uma outra consulta realizada em janeiro de 2019 no mesmo portal, havia 741.591 exemplares e 197.736 criatórios de *S. similis* cadastrados no SISPASS nos 26 estados brasileiros mais Distrito Federal. A quase totalidade dos exemplares e criatórios encontravam-se em apenas sete estados brasileiros: Minas Gerais, São Paulo, Rio de Janeiro, Santa Catarina, Paraná, Espírito Santo e Rio Grande do Sul (gráficos 2 e 3) (IBAMA 2019b). A divergência no número de exemplares de *S. similis* entre os gráficos ocorreu devido a diferença de um mês entre as duas consultas realizadas, período no qual ocorreu a exclusão de 145 indivíduos no sistema. Apenas em Minas Gerais, são quase 63 mil criatórios e mais de 250 mil exemplares de *S. similis* cadastrados, sendo o estado com maior relevância para esta espécie em termos de criação em cativeiro (IBAMA 2019a, IBAMA 2019b).



**Gráfico 1 - Espécies cadastradas no SISPASS.** Dos 3.265.973 passeriformes cadastrados no SISPASS, 22,7% são exemplares de *S. similis*, representando 741.591 indivíduos. Exemplares de *Oryzoborus angolensis*, *Sicalis flaveola*, *Sporophila caerulescens* e *Sporophila maximiliani* também apresentam quantidades expressivas de indivíduos cadastrados no sistema (IBAMA 2019a). Na legenda, as espécies aparecem em ordem decrescente de acordo com o número de indivíduos cadastrados.

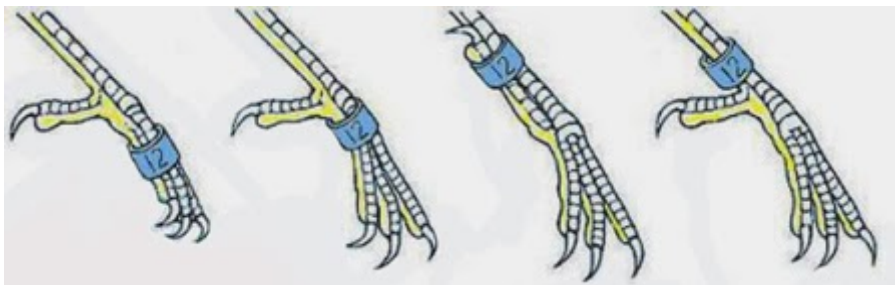


**Gráfico 2- Exemplares de *S. similis* cadastrados no SISPASS por estado.** De todos os 741.736 exemplares de *S. similis* cadastrados no SISPASS, 97% encontram-se em sete estados brasileiros, sendo a maioria nos estados de Minas Gerais e São Paulo (IBAMA 2019b). Na legenda, a ordem dos estados é de acordo com o número de exemplares cadastrados de forma decrescente.



**Gráfico 3- Criatórios de *S. similis* cadastrados no SISPASS por estado.** 96% de todos os 197.736 criatórios de *S. similis* cadastrados no SISPASS encontram-se em sete estados brasileiros, sendo a maioria nos estados de Minas Gerais e São Paulo (IBAMA 2019b). Os estados aparecem na legenda de forma decrescente em relação ao número de criatórios cadastrados.

Dentre as normas para a criação em cativeiro de passeriformes está a necessidade de identificação com o uso de anilhas (anéis metálicos fechados) (Ibama 2011), que devem ser afixadas ao membro inferior do animal até o 9º dia de vida e permitem o acompanhamento e fiscalização pelos órgãos ambientais. Uma vez afixada, não deverá machucá-lo e nem poderá ser retirada sem ser danificada (figura 3). As anilhas passaram a ser emitidas por empresas licitadas pelos órgãos públicos a partir da Instrução Normativa Ibama n. 05, de 18 de maio de 2001, que também especifica a medida de diâmetro interno da anilha para cada espécie. Para *S. similis*, foi estabelecida a anilha de diâmetro interno de 3,5mm. Tal diâmetro também foi estabelecido para outras espécies como *Gnorimopsar chopi*, *Paroaria dominicana* e *Saltator maximus*. As outras espécies para as quais este tamanho de diâmetro interno é preconizado são bem menos frequentes nos criatórios que *S. similis*.



**Figura 3- Processo de anilhamento de passeriforme.** A anilha deve ser afixada ao pé do filhote até o 9º dia de vida, e, após colocada, não deve machucar o animal e nem poderá ser retirada sem ser danificada. Imagem retirada de: <http://www.canariosbelga.com.br/anilhando-canarios-passo-a-passo/>. Acesso em 11.12.2018.

Em uma consulta no portal de sistema de informação ao cidadão realizada em abril de 2019, foi solicitado o número de nascimentos de *S. similis* dos últimos anos, porém a resposta se referiu às declarações de nascimento de animais portadores de anilhas de diâmetro interno de 3,5mm. Os registros são feitos de acordo com a temporada reprodutiva, e, portanto, vão de 1º de agosto até 31 de julho do ano seguinte. Na temporada 2014-2015 foram declarados 83.086 nascimentos, 2015-2016 foram 100.314, 2016-2017 foram 36.756 e na temporada 2017-2018 foram 24.952 declarações de nascimento (Ibama 2019c). A queda na declaração de nascimentos provavelmente ocorreu como consequência da “Operação Delivery” realizada pelo Ibama em 2016 (IMA 2016). Esta operação consistia no condicionamento da entrega das anilhas à verificação dos nascimentos por agentes ambientais, que passaram a visitar pessoalmente os criatórios e entregar o número de anilhas de acordo com o número de nascimentos. A operação foi finalizada no começo de 2020 sem justificativa técnica, apenas por pressão de criadores amadores (World Animal Protection 2020). Esta operação mostrou que alguns criadores solicitavam um número de anilhas maior que o número de filhotes nascidos, e estas anilhas então poderiam ser utilizadas posteriormente em “animais esquentados”, ou seja, animais retirados ilegalmente da natureza, mas que, por possuírem anilhas idôneas, ganhavam aspecto de animais legalizados (World Animal Protection 2020).

### 1.1.3 Tráfico de animais silvestres

Ilegalidades relacionadas ao comércio e manutenção de animais vivos e caça são englobadas na expressão “tráfico de animais silvestres”, cuja movimentação financeira

por ano é estimada anualmente entre 7 a 23 bilhões de dólares americanos para o tráfico internacional de biodiversidade, com exclusão da pesca ilegal (Nelleman 2016). A retirada ilegal de inúmeros espécimes da natureza é a principal consequência do tráfico ilegal de animais silvestres, e resulta em danos para as populações da espécie alvo e para a biodiversidade do ecossistema (Petrossian 2016). O tráfico, juntamente com a perda de habitat, introdução de espécies exóticas e poluição, pode resultar na diminuição do tamanho populacional de uma espécie, resultando, conseqüentemente, em menor diversidade genética a cada geração (Costa 2017).

No Brasil, a Lei de Crimes Ambientais (Lei 9.605/1998), aborda, em seu artigo 29, os crimes contra a fauna:

“Art. 29. Matar, perseguir, caçar, apanhar, utilizar espécimes da fauna silvestre, nativos ou em rota migratória, sem a devida permissão, licença ou autorização da autoridade competente, ou em desacordo com a obtida:

Pena - detenção de seis meses a um ano, e multa.

§ 1º Incorre nas mesmas penas:

[...]

III - quem vende, expõe à venda, exporta ou adquire, guarda, tem em cativeiro ou depósito, utiliza ou transporta ovos, larvas ou espécimes da fauna silvestre, nativa ou em rota migratória, bem como produtos e objetos dela oriundos, provenientes de criadouros não autorizados ou sem a devida permissão, licença ou autorização da autoridade competente.

[...]

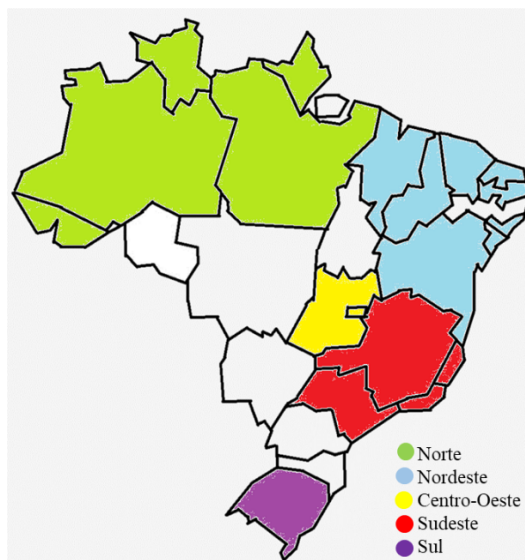
§ 3º São espécimes da fauna silvestre todos aqueles pertencentes às espécies nativas, migratórias e quaisquer outras, aquáticas ou terrestres, que tenham todo ou parte de seu ciclo de vida ocorrendo dentro dos limites do território brasileiro, ou águas jurisdicionais brasileiras” (Brasil 1998).

É importante ressaltar que estes crimes, por serem considerados de menor relevância, permitem a transação penal, garantindo ao infrator a possibilidade de ter pena não privativa de liberdade e cumprir penas alternativas (Pinto 2014). Por serem considerados de menor potencial ofensivo, possuem pena máxima não superior a dois anos, o que vai contra a recomendação da ONU de tratar estes crimes como graves, com pena privativa de liberdade de pelo menos quatro anos (ONU 2015). Já há diversos projetos de lei na Câmara dos Deputados que visam o agravamento das penas para os crimes contra a fauna silvestre.

Os Centros de Triagem de Animais Silvestres (Cetas) desempenham papel fundamental na logística de combate ao tráfico de animais no Brasil, assim como fornecem dados importantes para a dimensão desta atividade ilegal no país. De acordo com artigo 2º da Instrução normativa IBAMA nº 23, de 31 de dezembro de 2014, os Cetas são estruturas responsáveis pelo “manejo de fauna silvestre com finalidade de prestar serviço de: recepção, identificação, marcação, triagem, avaliação, recuperação, reabilitação e destinação de animais silvestres provenientes de ação fiscalizatória, resgates ou entrega voluntária de particulares; e que poderá realizar e subsidiar pesquisas científicas, ensino e extensão”. Segundo o relatório do Ibama (2016), havia 23 Cetas em funcionamento sob gestão do Ibama distribuídos nas cinco regiões do país, sendo, com exceção da Bahia (com dois Cetas – nas cidades de Salvador e Porto Seguro) e de Minas Gerais (com três Cetas – nas cidades de Belo Horizonte, Juiz de Fora e Montes Claros), um Cetas em cada um das seguintes unidades federativas: Acre, Alagoas, Amazonas, Amapá, Ceará, Distrito Federal, Espírito Santo, Goiás, Maranhão, Pará, Paraíba, Piauí, Rio de Janeiro, Rio Grande do Norte, Roraima, Rio Grande do Sul, Sergipe e São Paulo. A região Nordeste, por conter nove Cetas sob gestão do Ibama, é a que possui o maior número de unidades, seguida pelo Sudeste, com seis Cetas (figura 4). Atualmente, estas 23 unidades dos Cetas continuam em funcionamento (Ibama 2019d, Ibama 2020). Entre os anos de 2002 e 2014, os Cetas brasileiros receberam um total de 568.645 animais (tabela 1), sendo que as aves representaram cerca de 80% dos animais recebidos, seguido dos répteis, com cerca de 10% (gráfico 4) (Destro 2012, Ibama 2016). De 2010 a 2014, o Cetas de Belo Horizonte foi o que mais recebeu animais no país, recebendo mais de 6 mil animais por ano (Ibama 2016). Vale ressaltar que estes números não representam o total de animais retirados da natureza, já que boa parte dos animais é solta antes mesmo de chegar ao Cetas ou no próprio local da captura por ainda estar em estado asselvajado



(Destro 2012). Dos 261.091 animais que chegaram aos Cetas entre 2010 e 2014, 63,5% era proveniente de apreensão por órgãos de fiscalização, 14,5% de entrega voluntária (entrega de animais silvestres criados ilegalmente sem a aplicação de penalidades ou de animais cujo dono não possui mais interesse em manter em sua residência), 12,9% de resgate e 9,1% não tinham sua origem informada (Ibama 2016).



**Figura 4- Unidades da federação que possuíam Centros de Triagem de Animais Silvestres (Cetas) sob gestão do Ibama em 2016.** Os 23 Cetas em funcionamento estavam distribuídos em 20 unidades da federação, sendo apenas um Cetas por unidade federativa, com exceção da Bahia e Minas Gerais, com dois e três Cetas, respectivamente (Imagem baseada nas informações de Ibama 2016).

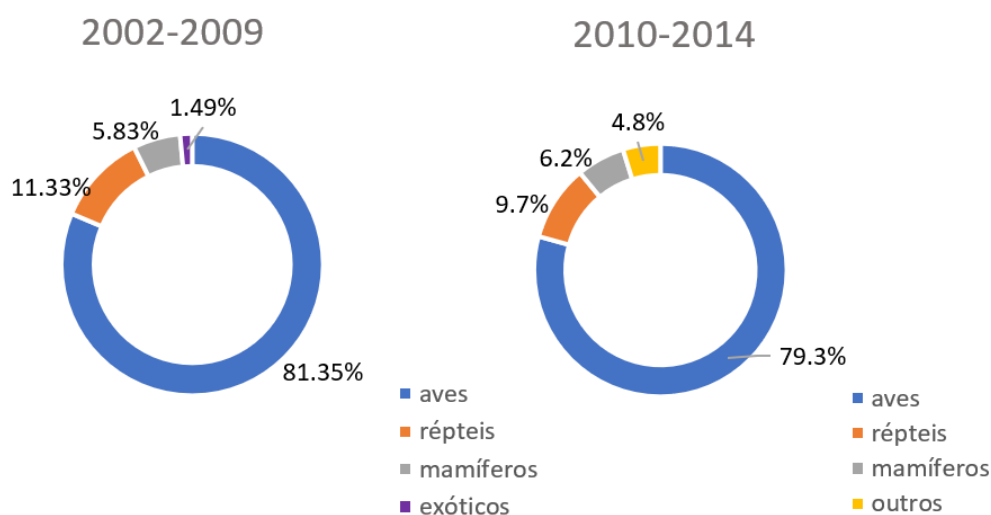
**Tabela 1- Número de animais recebidos pelos Cetas sob gestão do Ibama entre 2002 e 2014.**

Ano	Número de animais recebidos
2002	16.031
2003	57.417
2004	42.250
2005	40.309
2006	34.229
2007	28.447
2008	62.955

2009	25.916
2010	57.901
2011	53.915
2012	50.541
2013	59.097
2014	39.637
<b>Total</b>	<b>568.645</b>

Fonte: Destro 2012 e Ibama 2016.

### Percentual de espécimes recebidos nos Cetas de acordo com o táxon



**Gráfico 4- Percentual de animais recebidos nos Cetas de acordo com o táxon entre 2002-2009 e 2010-2014.** Nos dois períodos analisados, nota-se a predominância de recebimento de animais da Classe Aves, seguida pela Reptilia. As informações constantes nos gráficos foram retiradas de Destro 2012 (para o período 2002-2009) e Ibama 2016 (para o período 2010-2014).

#### 1.1.4 Tráfico de *Saltator similis*

Já há alguns anos, o *S. similis* figura no topo das listas das aves mais apreendidas nas fiscalizações. A compilação realizada por Destro (2012) trouxe o *Saltator similis* na segunda posição entre as espécies mais apreendidas pelo Ibama entre 2005 e 2009. Não diferente da escala nacional, os estados e municípios também listam o trinca-ferro no topo

das espécies mais apreendidas (Borges 2006, SEMA/PMA 2006, Gogliath 2010, Franco 2012, Nunes 2012, Felker 2013, Viana 2013, Freitas 2015, Silva 2015). No Cetas de Belo Horizonte - MG, em 2015, o trinca-ferro ocupou a primeira posição entre as espécies mais apreendidas entregues no local, totalizando 2.200 indivíduos (Érika Procópio, analista ambiental do Instituto Estadual de Florestas, comunicação pessoal).

Uma análise retrospectiva de laudos de perícia da Polícia Federal em anilhas de passeriformes do período de 2006 a 2015, realizada por Mayrink (2016), mostrou que 29,3% das anilhas oficiais de passeriformes periciadas eram de diâmetro 3,5mm (tamanho preconizado para *S. similis*), sendo 72,5% delas fraudadas. Na maioria das vezes, a fraude ocorreu por aumento do diâmetro interno da anilha. Tal alargamento é frequentemente realizado para afixar a anilha no pé de um animal adulto, e este procedimento pode resultar em fraturas no membro posterior da ave. As anilhas comumente são falsificadas por adulteração ou por contrafação, de modo que um leigo no assunto seja capaz de comprar o animal portador de anilha falsificada acreditando se tratar de um animal legalizado.

Além da exigência das anilhas, recentemente, a resolução do Conselho Nacional de Meio Ambiente n 487, de 15 de maio de 2018, trouxe a obrigatoriedade de identificação genética das matrizes reprodutoras do plantel de criatórios autorizados pelo órgão ambiental, que deverá garantir a identificação de paternidade com o uso de no mínimo dez *loci*, porém, sem maiores especificações quanto ao tipo de marcador molecular, tamanho dos marcadores ou questões estatísticas. Essa legislação também traz uma lista de espécies para genotipagem por prioridade, estando o *S. similis* em 1º lugar devido à demanda do tráfico. A genotipagem, juntamente com as anilhas, permitirá identificar de forma mais precisa os animais legalizados, incluindo casos em que a perícia da anilha é inconclusiva, em que há perda de informações das anilhas. Essa abordagem também irá garantir que os filhotes, mesmo aqueles que tenham aparência legal (animais com anilhas idôneas), são realmente descendentes de animais também legalizados. Porém, a obrigatoriedade dos testes de filiação só acontecerá à medida que os mesmos forem desenvolvidos e disponibilizados em escala comercial.

### 1.1.5 Tráfico de animais silvestres e surgimento novas doenças

Em dezembro de 2019, na China, foi descoberto um vírus que meses mais tarde veio a ser o causador da pandemia declarada pela Organização Mundial de Saúde: coronavírus da síndrome respiratória aguda grave 2 (SARS-CoV-2), causador da COVID-19 (*Coronavirus Disease* 2019). Apesar de não se ter certeza sobre sua origem, tendo o morcego e o pangolim como principais suspeitos (Cyranoski 2020), é sabido que o mercado chinês de vida silvestre teve participação na propagação do vírus (Cohen 2020). Esta não é a primeira vez, e provavelmente também não será a última, que um vírus presente em animais ultrapassa a barreira interespecífica e passa a infectar seres humanos. Neste contexto, é possível citar alguns exemplos de vírus que afetam humanos: vírus da imunodeficiência humana (HIV), que resultou da transmissão do vírus da imunodeficiência símia (SIV) de primatas não-humanos (Sharp 2010); Ebola vírus humano (EBOV), cujos surtos foram associados ao contato com homínídeos (*great apes*) (Leroy 2004); e coronavírus da síndrome respiratória aguda grave 1 (SARS-CoV-1), cujo surgimento do surto de 2003 está relacionado ao consumo de pequenos carnívoros (Bell 2004). É importante ressaltar que a barreira xenográfica pode ser ultrapassada também no sentido contrário, como por exemplo a tuberculose em elefantes asiáticos e primatas não-humanos associadas ao contato com humanos infectados (Montali 2001, Mikota&Maslow 2011), sendo a quebra da barreira de espécie pelo patógeno um problema não somente de saúde pública como também de conservação (Cook 2004).

Até recentemente, as discussões a respeito do tráfico de animais focavam, principalmente, na preservação de espécies. Com o advento da COVID-19, tal assunto ganhou relevância devido às consequências que podem trazer na população humana, gerando, inclusive, crise econômica. O comércio de animais silvestres apresenta riscos no contexto de surgimento de novas doenças humanas por aproximar pessoas à animais vivos e seus subprodutos, seja como alimentos, animais de estimação, uso medicinal ou outros fins (Broad 2020). Além disto, pode ocorrer a união de animais que carregam patógenos de diferentes ecossistemas em um local, facilitando a propagação do patógeno de uma maneira que não aconteceria normalmente na natureza (Sheikh 2020). Uma análise de produtos de animais silvestres confiscados em aeroportos americanos, feito por Smith e colaboradores (2012), destacou um potencial risco à saúde pública devido à múltiplas oportunidades de transmissão na exposição contínua deste material durante a caça, transporte e consumo. Este estudo demonstrou que a importação de produtos vindos ilegalmente de animais silvestres poderia atuar como um canal para a propagação de

patógenos e sugeriu que a implementação de vigilância de doenças no comércio de animais silvestres poderia facilitar a prevenção do surgimento de doenças.

Concentrar esforços para regular, reduzir ou, em alguns casos, eliminar o comércio de animais silvestres poderia ser uma abordagem efetiva para diminuir os riscos de doenças para seres humanos, animais domésticos, animais selvagens e ecossistemas (Karesh 2005). Recentemente e de forma pioneira, em novembro de 2020, os Estados Unidos anunciaram que traficantes de animais silvestres e madeira serão inelegíveis para o visto norte-americano, assim como seus familiares imediatos. A medida visa enfraquecer as organizações criminosas envolvidas neste tipo de crime que compromete a segurança nacional, afeta a prosperidade econômica, alimenta a corrupção e difunde doenças. Até o momento, este é o único país com tal restrição, mas há a esperança de que este modelo seja seguido por outros (U.S Department of State, 2020).

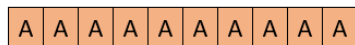
#### 1.1.6 Microssatélites

De acordo com Milach (1998), marcadores moleculares são características de DNA que diferenciam dois ou mais indivíduos e são herdadas geneticamente. Estes marcadores são utilizados com diversos propósitos, como testes de paternidade, análise de diversidade genética e estrutura populacional, genética forense, monitoramento de endocruzamentos e estudos ecológicos e evolutivos (Rafalski & Tingey 1993, Jarne & Lagoda 1996, Knapik 1998, Provan 2001, Schlotterer 2004). Dentre os principais marcadores moleculares, os do tipo microssatélites são vantajosos por serem abundantes, hipervariáveis, possuírem natureza multialélica, estarem amplamente distribuídos no genoma, serem expressos como alelos codominantes (Lima, 1998; Moreira, 1999, Schlotterer 2000) e possuírem herança mendeliana (Zane 2002). Além disto, por serem econômicos para serem desenvolvidos, são uma boa alternativa para serem utilizados em projetos com recursos limitados (Jennings 2011).

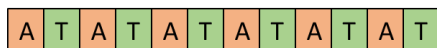
Microssatélites são sequências curtas (um a seis nucleotídeos: mono-, di-, tri-, tetra-, penta- e hexanucleotídeos) repetidas em tandem (lado a lado) (figura 5), encontradas no genoma de eucariotos e procariotos (Zane *et al.*, 2002). Apesar de serem comumente considerados “DNA lixo” (Tang 2017), são encontrados geralmente em regiões não-codificantes (Ellegren, 2000) e podem estar ligados à ocorrência de algumas doenças genéticas (Gymrek 2017) e estar envolvidos na regulação de genes (Contente 2002). De acordo com a forma na qual estão presentes no DNA, os microssatélites também podem ser classificados como perfeitos, no qual a sequência de repetição não é

interrompida; imperfeito, no qual há ocorrência de uma base que não faz parte do *motif* entre as repetições; interrompido, no qual há uma pequena sequência entre as repetições; ou composto, quando há duas repetições de sequências distintas adjacentes (Oliveira 2006) (figura 6). Também chamados de SSR (*Simple Sequence Repeats*), STR (*Short Tandem Repeats*) ou SSRP (*Simple Sequence Repeat Polymorphism*), os microssatélites são um dos marcadores moleculares mais polimórficos encontrados no genoma de plantas e animais. Este polimorfismo acontece devido à variação do número dos elementos repetidos, podendo ser consequência de mutação, erros da DNA polimerase durante o processo de replicação e reparo da molécula de DNA (Ellegren 2004). O alto grau de polimorfismo dos microssatélites faz com que eles sejam muito usados em estudos populacionais e de análise de parentesco (Dakin & Avise 2004).

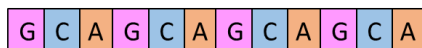
A- Mononucleotídeo



B- Dinucleotídeo



C- Trinucleotídeo



D- Tetranucleotídeo



E- Pentanucleotídeo



F- Hexanucleotídeo



**Figura 5- Exemplos da classificação dos microssatélites quanto ao motivo da repetição.** A- Mononucleotídeo: repetição em tandem de “A”. B- Dinucleotídeo: repetição em tandem da sequência de dois nucleotídeos “AT”. C- Trinucleotídeo: repetição em tandem da sequência de três nucleotídeos “GCA”. D- Tetranucleotídeo: repetição em tandem da sequência de quatro

nucleotídeos “TCGG”. E- Pentanucleotídeo: repetição em tandem da sequência de cinco nucleotídeos “ATCTT”. F- Hexanucleotídeo: repetição em tandem da sequência de seis nucleotídeos “AATTCG”.

A- Microsatélite perfeito



B- Microsatélite imperfeito



C- Microsatélite interrompido



D- Microsatélite composto



**Figura 6- Exemplos de microsatélites quanto à forma em que aparecem no DNA.** Classificação de acordo com Oliveira (2006). A- Microsatélite perfeito: os motivos das repetições (“AT”) não são interrompidos. B- Microsatélite imperfeito: há a interrupção das repetições (“AT”) por uma base que não faz parte do *motif* (“C”). C- Microsatélite interrompido: as repetições (“AT”) são interrompidas por uma pequena sequência (“CGGC”). D- Microsatélite composto: duas repetições de *motifs* diferentes (“AT” e “CT”) ocorrem de forma adjacente.

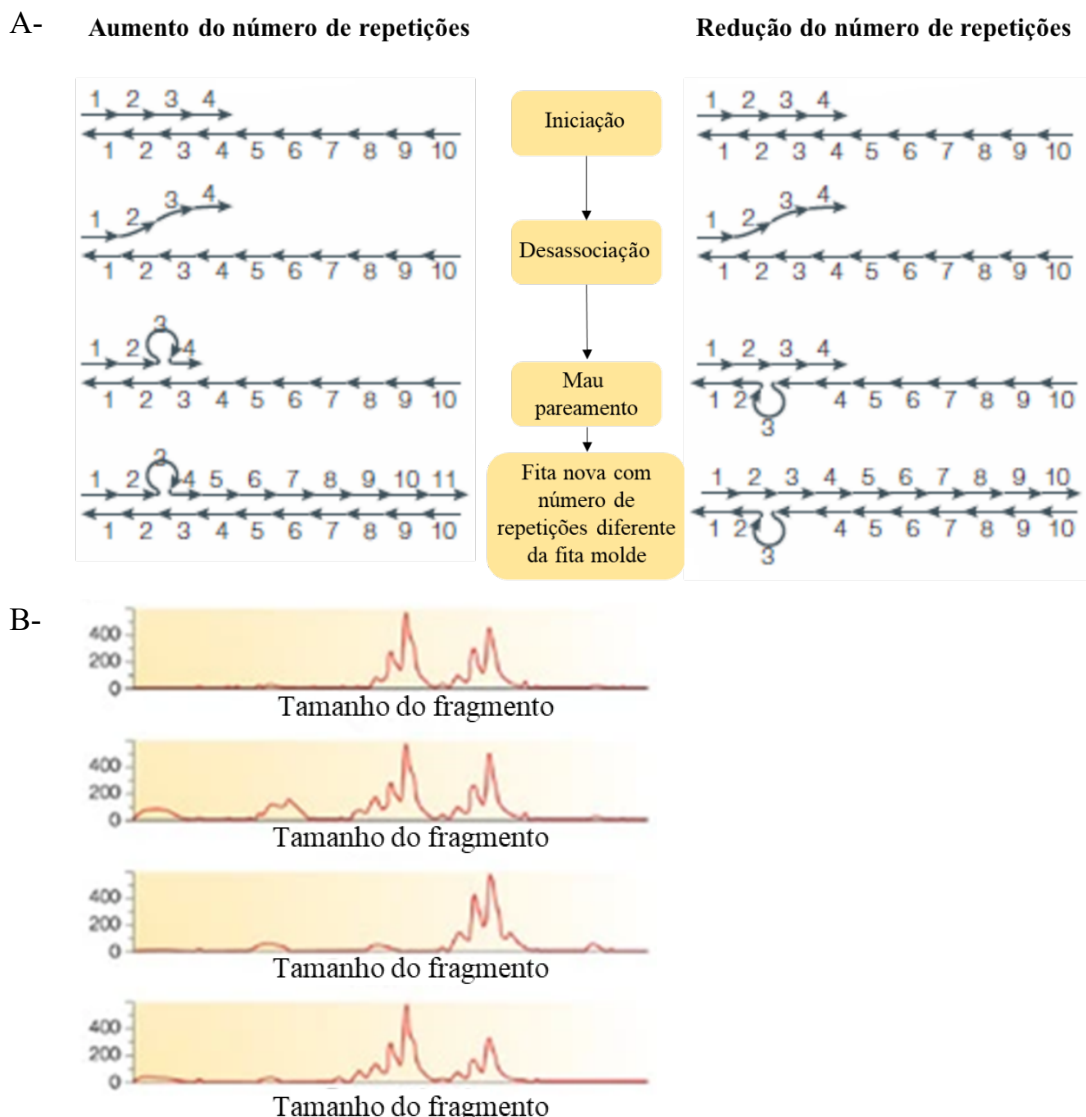
Há bastante discussão sobre a forma de evolução destes marcadores, pois apresentam padrões complexos (Ellegren 2000), e a variação no número de repetições acontece devido ao deslize da DNA polimerase durante a replicação (Umar 1996). Um dos modelos mais aceitos recentemente diz que mutações ponto quebram as longas repetições em unidades menores, explicando porque os microsatélites não expandem em enormes sequências apesar do aumento do número de repetições (Ellegren 2004). Ainda sobre questões evolutivas, aparentemente há a tendência de que a densidade de microsatélites se correlacione positivamente com o tamanho do genoma, sendo que entre os genomas eucariotos já estudados, a densidade de microsatélites é maior em mamíferos (Hancock 1996, Toth 2000, Katti 2001, Ellegren 2004); enquanto que em plantas a frequência de microsatélites se correlaciona de forma negativa em relação ao tamanho do genoma (Morgante 2002). Em humanos, repetições do tipo microsatélite representam 3% do genoma (em um genoma de 3,2 bilhões de pares de bases), sendo repetições do tipo dinucleotídeo as mais comuns, e, mais especificamente, as repetições do tipo  $(CA)_n$

(Lander 2001, International Human Genome Sequencing Consortium 2001, Ellegren 2004). Já as espécies de aves estudadas possuem pouca variação no tamanho do genoma (Organ 2011), variando entre 0,89 bilhão de pares de bases em beija-flor a 2,11 bilhões de pares de bases em avestruz (Gregory 2007, Gregory 2009), estando entre os menores tamanhos de genoma entre os vertebrados (Gregory 2002). A redução no tamanho do genoma, assim como a redução do tamanho das células de aves, pode estar relacionado com as necessidades metabólicas necessárias para o voo (Hughes 1995, Vinogradov 1997, Gregory 2009, Zhang 2012, Wright 2014), além de possivelmente estarem associadas aos mecanismos de reparo de DNA (Schubert 2016). Além do tamanho reduzido, aves também possuem baixa densidade de repetições do tipo microssatélite em seus genomas quando comparados a mamíferos, com uma estimativa de 1 microssatélite a cada 31kb, enquanto que a estimativa para humanos, anteriormente ao conhecimento da sequência do genoma humano e utilizando a mesma metodologia de hibridização, era de 1 microssatélite a cada 6kb (Primmer 1997, Neff 2007). Vale ressaltar que os estudos sobre genomas de aves vêm ampliando recentemente e, em um futuro próximo, será possível calcular a densidade de microssatélites com dados mais robustos (Peona 2018).

Por serem marcadores que utilizam de PCR para sua análise, seus *loci* podem ser amplificados com sucesso a partir de amostras de baixa qualidade ou pouca quantidade de DNA, sendo marcadores bastante úteis para estudos envolvendo DNA antigo e espécimes de museus (Wandeler 2007). Dentre as dificuldades da utilização de microssatélites para estudos genéticos está a ocorrência de bandas *stutter* durante a PCR, que ocorrem pelo deslizamento da fita molde durante a replicação seguido por mau pareamento, gerando bandas frequentemente menores que o real alelo de microssatélite (Schlotterer 2004) (figura 7), fazendo com que o tamanho do alelo observado não seja um só tamanho, mas sim uma distribuição de variação de tamanhos (Suez 2016). O deslizamento na replicação é um mecanismo de mutação de microssatélites e que também ocorre *in vitro* (Ellegren 2004), dificultando a análise de dados e, juntamente com os padrões evolutivos imprevisíveis, resulta na necessidade de extensa curadoria manual (Ellegren 2000, Guichoux 2011, Krehenwinkel 2019). A formação de *stutter* ocorre mais frequentemente em dinucleotídeos, formando fragmentos tipicamente uma repetição menor que o original (Walsh 1996, Brookes 2012) ou mais; enquanto que repetições do tipo tri, tetra- e pentanucleotídeos aparentam ser menos propensas à formação de *stutter* (Edwards 1991). A ocorrência de formação de *stutter* depende de características da



sequência, como, por exemplo, o fato de possuir maior conteúdo A-T implicar no aumento da taxa de ocorrência de *stutter* (Brookes 2012).



**Figura 7- Alterações do tamanho do microssatélite.** A– Mecanismo de aumento ou diminuição do número de repetições do microssatélite causando pela desassociação das fitas seguido por mau pareamento. B- O deslize da replicação também ocorre em ampliações *in vitro* de microssatélites, resultando na formação de sequências de tamanhos diferentes da original, conhecidas como bandas *stutter*. Imagem retirada de Ellegren 2004 e reproduzida com tradução para o português com autorização sob licença RightLink número 5076020606645, obtida em 25 de maio de 2021.

Além disto, cada tipo de microssatélite apresenta vantagens e desvantagens para serem utilizados em estudos genéticos. Dinucleotídeos costumam ser úteis por serem polimórficos (Payseur 2011) e geralmente possuem alelos de tamanhos pequenos, facilitando o sequenciamento, porém, estão mais propensos à formação de *stutter*, dificultando as análises. Trinucleotídeos possuem a vantagem de baixa ocorrência de *stutter*, porém podem estar sob seleção e possuem menos variabilidade.

Tetranucleotídeos, além de serem polimórficos, possuem menor ocorrência de formação de *stutter*, porém, podem ter alelos de tamanhos maiores. Penanucleotídeos costumam ser bastante polimórficos, porém raros e com alelos grandes. Hexanucleotídeos possuem alelos de tamanhos maiores, podem estar sob seleção e possuírem menor variabilidade (Gardner 2011).

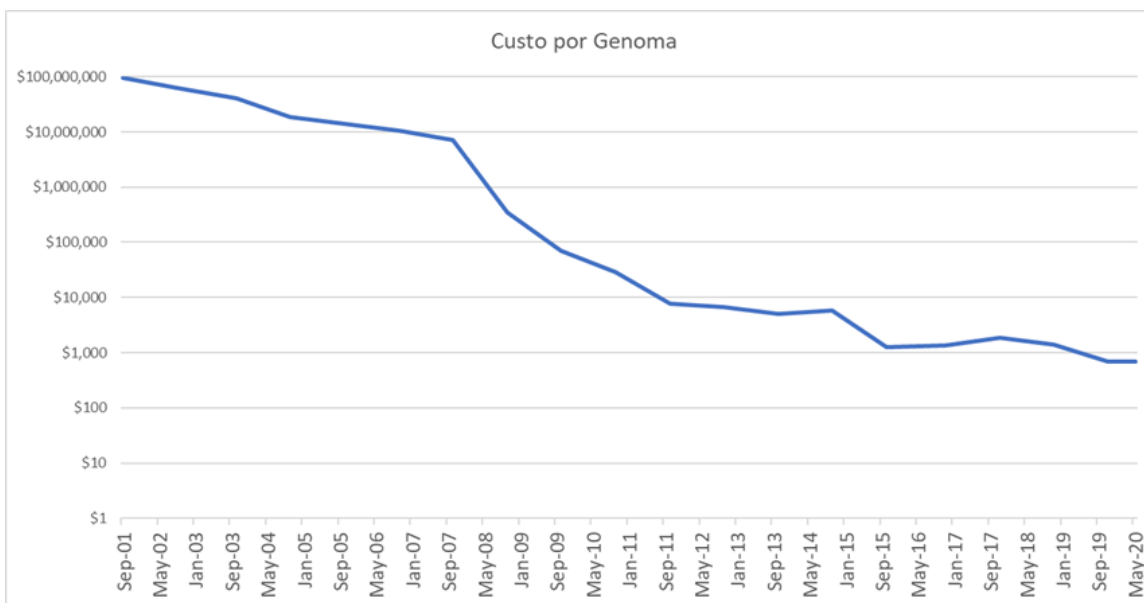
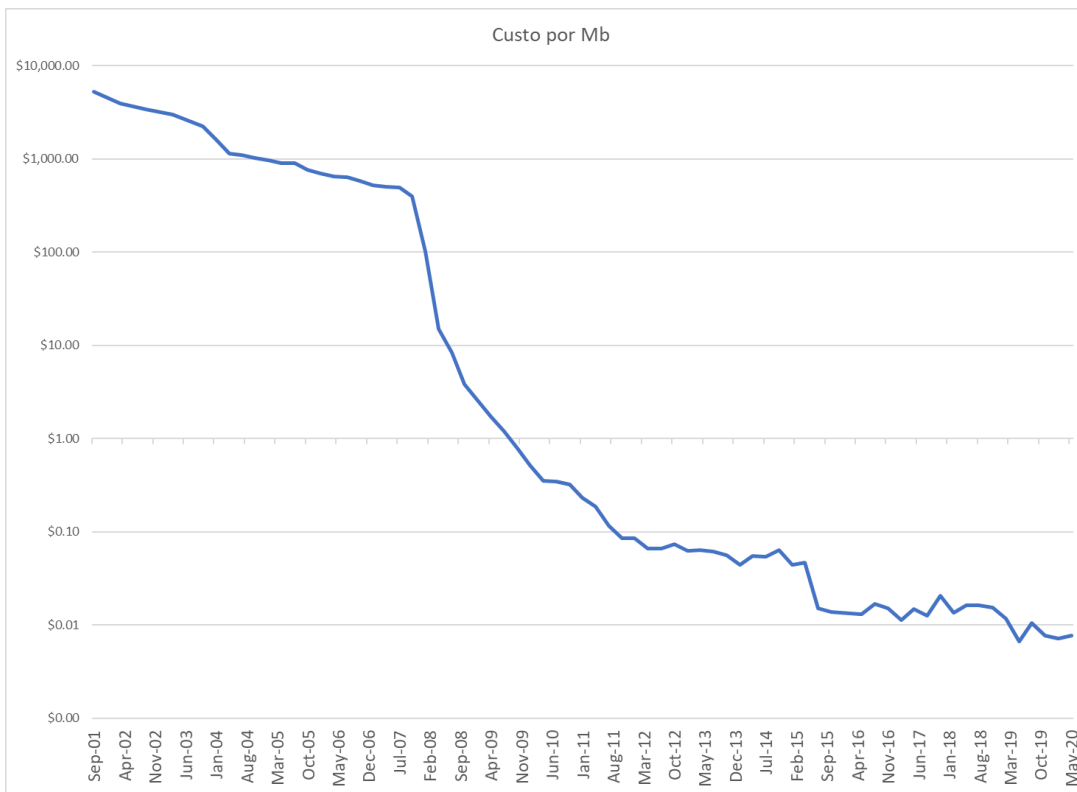
#### 1.1.7 Sequenciamento massivo de DNA

O método de Sanger de sequenciamento de DNA (sequenciamento de 1ª geração), publicado em 1977, foi o único método utilizado por cerca de 30 anos, sendo inclusive utilizado para o primeiro sequenciamento do genoma humano. Somente em 2005 que novos métodos começaram a surgir e revolucionar esta área, como o sequenciamento por síntese desenvolvido pela 454 Life Sciences (atualmente Roche) e o protocolo de sequenciamento *Multiplex Polony* desenvolvido pelo laboratório do Professor George Church, sendo que ambas as tecnologias utilizavam estratégias que reduziam o volume de reação inicial e aumentava o número reações de sequenciamento. Apesar do primeiro aparelho com tecnologia 454 gerar cerca de 50 vezes mais sequências com um custo seis vezes menor que um sequenciador por capilaridade, houve ceticismo da comunidade científica com a nova tecnologia, assim como também ocorreu quando do surgimento do método de Sanger (Schuster 2008).

Conforme ocorre com boa parte das novas tecnologias, elas são aprimoradas com o passar dos anos, e o mesmo aconteceu com o que veio a ser chamado de sequenciamento de nova geração (NGS)/sequenciamento massivo paralelo de DNA/sequenciamento de 2ª geração/sequenciamento de alta performance. Diversas plataformas foram criadas com esta metodologia (como por exemplo Solexa/Illumina, SOLiD – Sequencing by Oligo Ligation Detection e Ion Torrent), sendo que todas possuíam três pontos em comum: preparação de bibliotecas de DNA independente da clonagem de fragmentos de DNA em bactérias, milhares a milhões de sequenciamentos em paralelo em vez de centenas e o resultado do sequenciamento ser detectado sem a necessidade de eletroforese (Schuster 2008, Van Dijk 2014). Com o uso desta tecnologia de sequenciamento é possível produzir maior volume de sequências com a mesma quantidade de DNA inicial e resulta em economia de tempo e redução de custo por base quando comparado ao sequenciamento de Sanger, sendo especialmente úteis para o sequenciamento de um maior número de amostras (Carvalho & Silva 2010, Illumina 2020<sup>a</sup>, Darby 2016). Porém,

se utilizado menos que 20 alvos diferentes, o custo-benefício e o tempo necessário se tornam mais atrativos com o sequenciamento de Sanger. Outro desafio dessa tecnologia é que exige um fluxo de trabalho de tratamento de dados com utilização de ferramentas de bioinformática (Illumina 2019, Illumina 2020a), além de que, por enquanto, também há a limitação do tamanho das sequências de DNA, gerando *reads* de tamanhos inferiores às geradas pelo sequenciamento de Sanger (Van Dijk 2014, Mardis 2017).

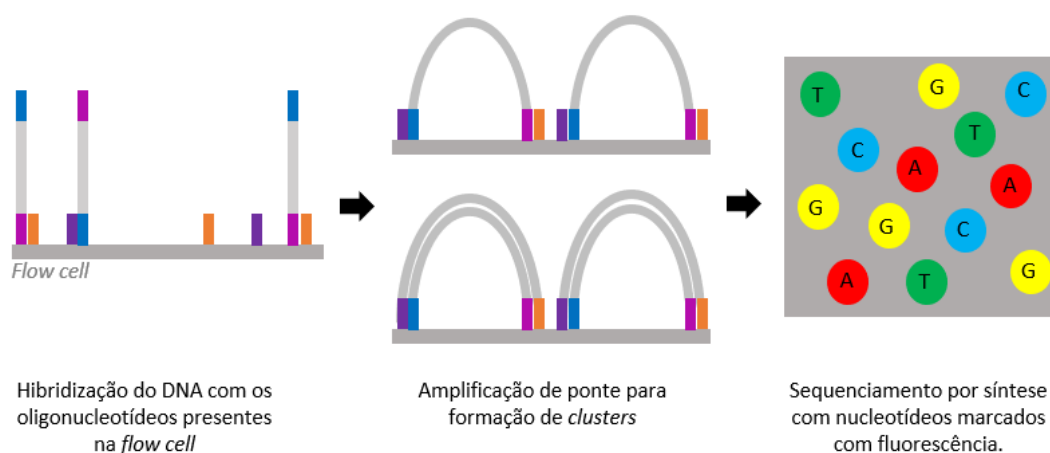
Devido ao esforço contínuo de diversos grupos de pesquisa e empresas em aprimorar técnicas de sequenciamento de DNA, o custo para se obter genomas completos vem diminuindo com o passar dos anos. Até 2008, a tendência de redução de custo de sequenciamento seguia a da Lei de Moore, que determina que a capacidade dos computadores dobra a cada dois anos. Porém, com a substituição do sequenciamento de Sanger pelo NGS, o custo do sequenciamento teve uma queda brusca e a tendência se distanciou da linha da Lei de Moore, conforme pode ser visto no gráfico 5, abaixo. O custo para o primeiro sequenciamento completo do genoma humano, que possui aproximadamente 3.2 bilhões de pares de bases e demorou mais de 10 anos para ser sequenciado (1990-2003), foi de pouco menos que 3 bilhões de dólares (valor não corrigido), e, atualmente, é possível sequenciar um genoma com tamanho similar ao genoma humano por cerca de mil dólares, e há, inclusive, a perspectiva de que estes sequenciamentos serão, em um futuro próximo, ofertados por US\$100,00 (Veritas 2019, NIH 2020).



**Gráfico 1- Histórico do custo para sequenciamento de DNA.** O custo de sequenciamento de DNA vem diminuindo com o passar dos anos devido ao desenvolvimento e melhoramento das tecnologias. Uma queda brusca pode ser observada nos gráficos no ano de 2008, que representa quando os centros de sequenciamento passaram a usar a tecnologia NGS em vez de sequenciamento de Sanger. A- Histórico do custo para sequenciamento por Mb (1Mb = 1 milhão de pares de bases) de DNA. B- Histórico do custo para sequenciamento de genoma de tamanho similar ao genoma humano (3.2 bilhões de pares de bases). Os gráficos foram construídos com os valores fornecidos pelo National Human Genome Research Institute e estes valores não refletem todos os custos envolvidos no sequenciamento, como, por exemplo, o desenvolvimento e melhoramento de ferramentas de bioinformática para análises do genoma (NIH 2020).

Para este trabalho, foi utilizado o aparelho de NGS da Illumina “MiSeq”, lançado em 2011 e especialmente conveniente para sequenciamento de *amplicons* (Liu 2012), sendo que, dentre as plataformas da Illumina disponíveis atualmente, esta é a capaz de gerar o maior tamanho de *reads* (até 300pb), fator importante também para a montagem de *contigs*. Com tempo de sequenciamento de 4 a 55 horas, é capaz de gerar até 15Gb de dados e até 25 milhões de *reads* por corrida (Illumina 2020b). Apesar do comprimento das *reads* geradas ser pequeno quando comparado a outros sequenciadores com outras metodologias (Besser 2018), as características mencionadas, juntamente com o baixo custo de sequenciamento, vem tornando o MiSeq uma das plataformas de sequenciamento mais utilizadas no mundo (Caporaso 2012, Schirmer 2015, Shokralla 2015).

De forma simplificada, à sequência do DNA alvo são adicionados os *indexes* (para identificação individual da amostra) e os adaptadores para hibridização por complementariedade de sequências com os oligonucleotídeos presentes na *flow cell*. Em seguida, há a amplificação de ponte destas sequências para a formação de *clusters* com posterior remoção das sequências *reverse*, e então o sequenciamento por síntese ocorre com o uso de nucleotídeos marcados com fluorescência, sendo adicionado apenas um por sequência em cada *round* e então é feita a leitura da base (figura 8). Esta metodologia de incorporar apenas um nucleotídeo de cada vez faz com que a tecnologia da Illumina produza significativamente menos erros do tipo *indel* quando comparado a outras técnicas de NGS, porém, é comum que mais erros aconteçam no final das *reads* (Cao 2005).



**Figura 8- Representação simplificada de sequenciamento na plataforma MiSeq.** O primeiro passo para o sequenciamento é a adição dos *indexes* ao DNA alvo. A estes *indexes* estão ligadas as sequências de adaptadores, que irão hibridizar com os oligonucleotídeos presentes na *flow cell*. As sequências contendo o DNA alvo são então amplificadas diversas vezes formando amplificação de ponte entre os adaptadores e oligonucleotídeos da *flow cell*, dando origem aos *clusters* (diversas cadeias de nucleotídeos com a mesma sequência). A cada *round*, um nucleotídeo marcado com fluorescência é adicionado ao processo de sequenciamento por síntese, e então é feita a leitura da sequência.

Anteriormente à tecnologia do NGS, a identificação de microssatélites era feita através de um trabalhoso processo de enriquecimento, clonagem e então sequenciamento de Sanger de determinado *locus*, e depois as análises costumavam ser realizadas pela comparação dos tamanhos dos fragmentos de PCR utilizando eletroforese, e o tamanho indicaria o número de repetições do STR alvo (Tang 2017). Porém, com o NGS é possível identificar centenas de microssatélites após apenas um sequenciamento contendo o DNA do organismo alvo, tornando o NGS uma metodologia comum para o desenvolvimento de microssatélites em organismos não-modelo (Chabot 2011, McCracken, Nowak 2014, Blower 2015, Pirog 2015, Maduna 2017), além de reduzir o custo em 2 a 5 vezes e também reduzir consideravelmente o tempo despendido para o desenvolvimento quando comparado a metodologia tradicional de desenvolvimento de microssatélites (Abdelkrim 2009, Santana 2009, Catoe 2010, Csencsics 2010, Mlausa 2011). Da mesma forma, as regiões microssatélites selecionadas podem ser sequenciadas por NGS para realizar as análises de parentesco, resultando em diversas *reads* para cada região amplificada. Porém, devido à limitação de tamanho das *reads* geradas pelo MiSeq, os *amplicons* precisam ter um tamanho relativamente pequeno e, por isso, neste trabalho não foi

possível utilizar os marcadores desenvolvidos por Ferreira (2012), que apesar de não terem sido desenhados especificamente para *S. similis* também foram capazes de amplificar para a espécie e aparentam ser polimórficos, porém possuíam produtos de PCR com tamanhos superiores aos indicados para o uso deste equipamento.

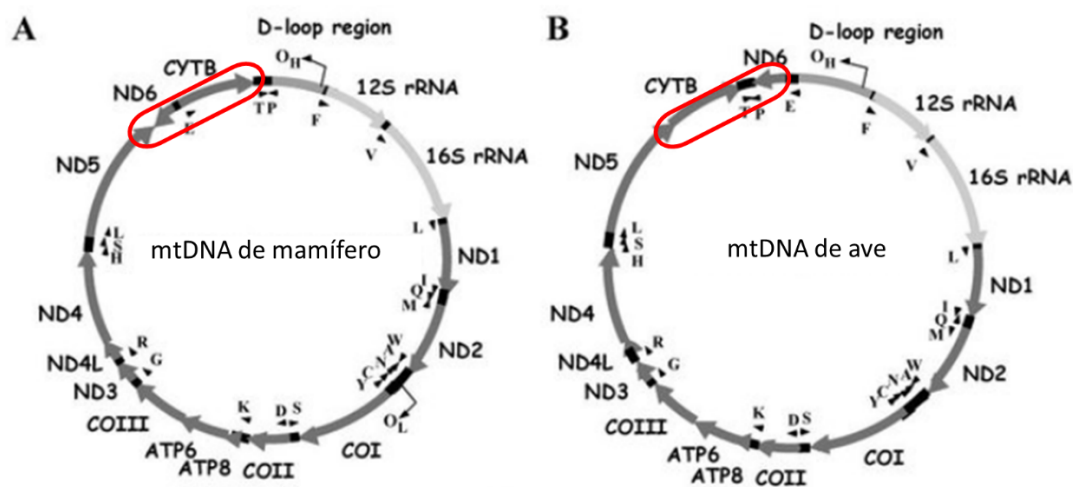
Para o sequenciamento total do genoma de um indivíduo, costuma ser necessária a realização de mais de um ciclo de sequenciamento, fazendo com que todas as bases sejam sequenciadas e possuam uma boa cobertura, que resultará em maior confiabilidade dos dados. Dentre os fatores envolvidos para determinar o tempo e número de sequenciamentos para se obter a sequência completa estão o equipamento e reagentes utilizados, assim como o tamanho do genoma a ser sequenciado. No caso do sequenciamento apenas do DNA mitocondrial, por possuir cerca de apenas 16.500pb em mamíferos (Taanman 1999), é possível determinar sua sequência completa com apenas um ciclo no NGS.

#### 1.1.8 DNA mitocondrial

O DNA mitocondrial (mtDNA) é um genoma a parte do genoma nuclear, presente no citoplasma de quase todas as células eucarióticas, e sua importância se dá principalmente nas vias de produção de energia das células. A forma de herança do mtDNA é um assunto bastante debatido, sendo que por muito tempo foi aceito que sua herança, ao menos em mamíferos, era apenas materna. Porém, cada vez mais estudos vem discutindo essa questão ao relatarem que também pode ocorrer contribuição do mtDNA paterno na formação de um novo indivíduo, inclusive humanos (Luo 2018). A ocorrência de heteroplasmia mitocondrial (ocorrência de mais de um haplótipo mitocondrial em um mesmo indivíduo) já foi demonstrada em animais como camundongos (Gyllensten 1991), gado (Sutovsky 1999), ovelhas (Zhao 2004), macacos (St John e Schatten 2004) e também em aves (Kvist 2003).

Em humanos, o mtDNA é uma molécula circular de 16.569 pb composta por 37 genes, dos quais 13 são codificadores de proteínas, dois genes de RNA ribossomal (rRNA) e 22 genes de RNA transportador (tRNA), além do D-loop, uma região controladora não-codificante, responsável pela regulação dos mecanismos de transcrição e replicação do genoma mitocondrial (Anderson 1981, Taanman 1999). Esta composição

é similar no mtDNA de aves, porém, com uma organização diferente próximo ao D-loop, conforme mostra a figura 9, abaixo (Quinn & Wilson 1993, Reyes 2004).



**Figura 9- Organização do mtDNA em mamíferos e em aves.** O genoma mitocondrial de aves e mamíferos possui composição conservada quanto ao número de genes e tamanho similar, porém, a organização dos genes próximos à região controladora D-loop possui uma pequena divergência, conforme destacado na imagem. A- mtDNA de mamíferos. B- mtDNA de aves. Imagem adaptada de Reyes (2004).

Marcadores moleculares de mtDNA têm sido frequentemente utilizados para a identificação de espécies e estudos filogenéticos por uma série de razões. Diferentemente do DNA nuclear, o mtDNA possui centenas de cópias por célula, sendo essa abundância bastante útil quando há disponibilidade de apenas uma pequena quantidade de amostra ou quando o DNA se encontra degradado (Wilson 1995 apud Tocchetto 2014, Bender 2000). Além disto, não sofre recombinação. Isso resulta na redução da diversidade dentro da espécie, o que, conseqüentemente, torna os marcadores mitocondriais adequados para a identificação de espécies (Dawnay 2007, Rojas 2011). Por não estar associado a histonas, a ausência de enzimas corretoras e a alta exposição aos radicais livres mutagênicos no ambiente oxidativo da mitocôndria faz com que o mtDNA apresente taxas de mutação mais elevadas, sendo estimada em dez vezes maior do que a do DNA nuclear (Wilson 1995 apud Tocchetto 2014). Isto resulta em um acúmulo de diferenças entre espécies, permitindo distinguir até mesmo espécies muito próximas (Budowle 2003).

Um dos bancos de dados genéticos mais famosos da atualidade é o BOLD (*Barcode Of Life Database*) (<https://www.boldsystems.org/>), que armazena sequências do citocromo oxidase C subunidade I (COI) de diversas espécies. Porém, como diferentes



partes do mtDNA evoluem com diferentes taxas de mutação (Avise 1986, Roques 2006), há a discussão sobre quais seriam os melhores genes a serem utilizados nas análises de filogenia ou identificação de espécies (Brown 1979, Moore 1995, Johns 1998). Inclusive, no sistema do BOLD há a informação de que, apesar de utilizar o COI para identificar espécies, a “validação adicional com marcadores genéticos independentes será desejável em algumas aplicações forenses” (BOLD Systems 2020). Utilizar somente a região do COI para identificação molecular de aves também pode resultar em falhas na delimitação de espécies alopátricas estreitamente relacionadas (espécies resultantes de isolamento geográfico seguido por isolamento reprodutivo) ou de espécies irmãs que hibridizaram (Aliabadian 2009). Assim, é importante ter mais informações sobre o genoma da(s) espécie(s) em estudo além do COI. Um exemplo desta importância é o projeto *Bird 10K* (<https://b10k.genomics.cn/>), que visa obter informações genéticas sobre mais de 10 mil espécies de aves ainda existentes. Recentemente, o projeto divulgou o sequenciamento do mtDNA de 336 espécies e, estas informações, juntamente com outras sequências já disponibilizadas, estão sendo utilizadas em estudos de diversidade molecular e filogenia, que poderão também lançar novas perspectivas em estudos evolutivos e auxiliar em projetos de conservação de espécies (Zhang 2015, B10K 2016, Feng 2020).

## 1.2 Justificativa e relevância do projeto

Considerando o elevado número de apreensões de *S. similis* em cativeiro de forma irregular e fraudes nas anilhas, vê-se a necessidade do desenvolvimento de um teste de filiação genético para auxiliar na fiscalização de criatórios, resultando no cumprimento da resolução do CONAMA n° 487, de 15 de maio de 2018. Os marcadores moleculares do tipo microssatélite selecionados também poderão ser úteis para avaliar a genética populacional destes animais e auxiliar nas técnicas de manejo dos criatórios e conservação, além de auxiliarem na investigação de rota de tráfico da espécie, caso sea utilizado juntamente com estudos de genética de populações. Da mesma forma, a sequência mitocondrial também poderá ser utilizada em análises forenses para detecção de tráfico de *S. similis*, assim como para estudos evolutivos.

### 1.3 Objetivo

Sequenciar e anotar o mitogenoma e buscar regiões polimórficas no genoma nuclear para desenvolvimento de teste de filiação baseado em microssatélites para *Saltator similis* (Ordem Passeriformes, Família Thraupidae) com o uso da tecnologia de NGS.

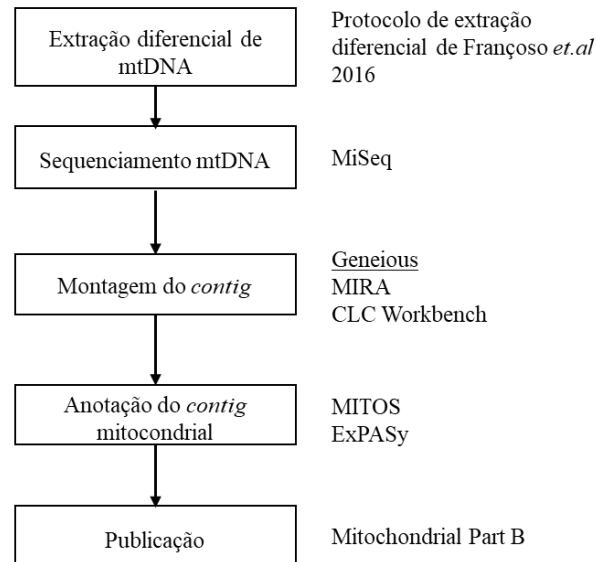
#### 1.3.1 Objetivos específicos

- Isolar, sequenciar, montar e anotar o DNA mitocondrial completo de *S. similis*;
- Buscar microssatélites nos *contigs* de DNA genômico e sintetizar conjuntos de *primers* para as regiões selecionadas;
- Padronizar as reações de PCR e avaliar o sequenciamento dos *amplicons*;
- Avaliar o grau de informatividade dos *loci* selecionados;
- Avaliar o potencial de uso dos *loci* selecionados para genotipagem;
- Realizar teste de validação com amostras de criatórios.

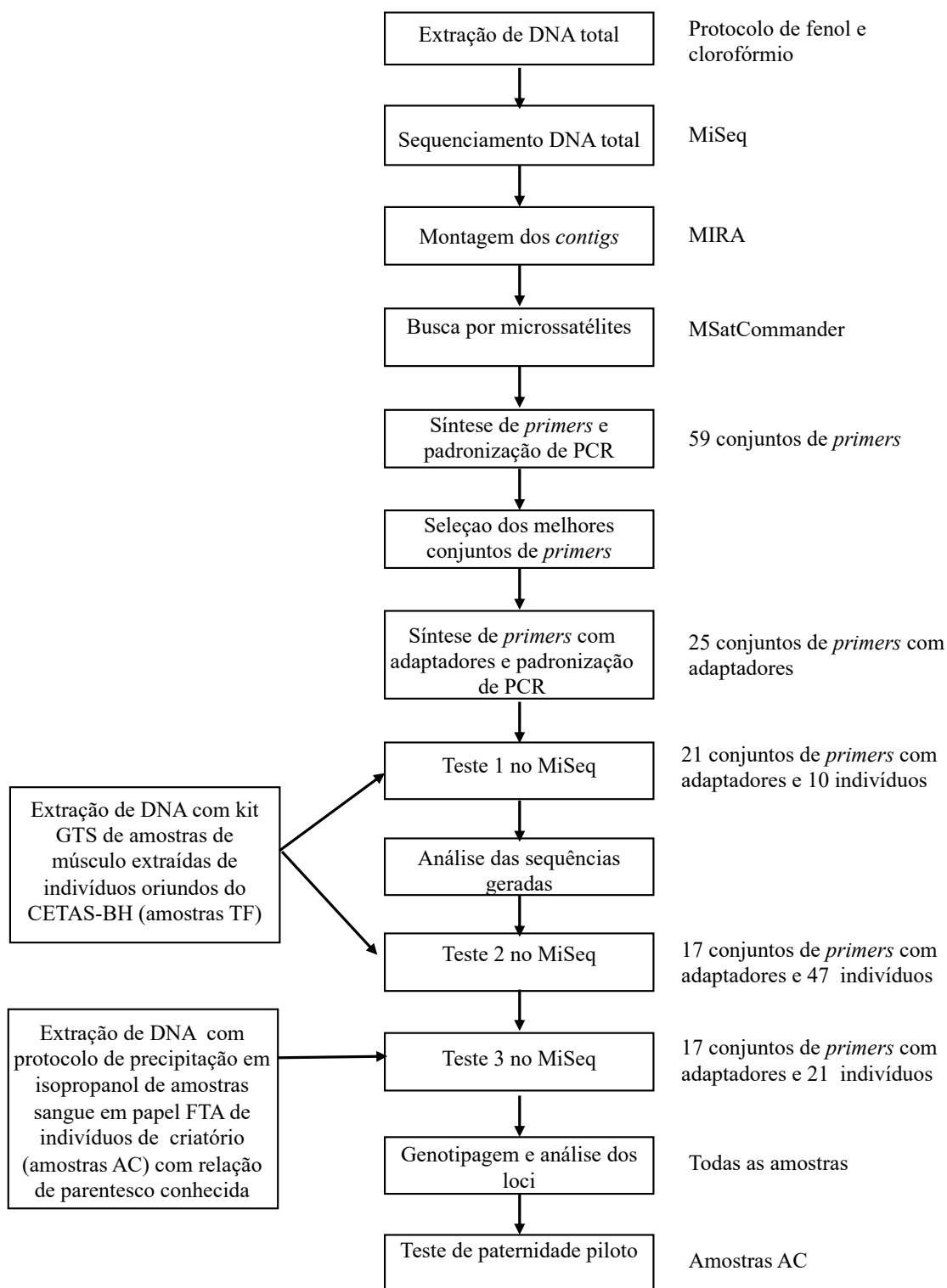
## 1.4 Metodologia

### 1.4.1 Fluxograma do trabalho

As figuras 10 e 11, abaixo, são representações esquemáticas da metodologia utilizada neste trabalho. Para facilitar a compreensão, foi adicionado, à direita do esquema, maiores informações ou resultados.



**Figura 10-** Representação esquemática da metodologia utilizada no sequenciamento do DNA mitocondrial.



**Figura 11- Representação esquemática da metodologia utilizada para desenvolvimento do teste de filiação.**

#### 1.4.2 Obtenção das amostras

As amostras de músculo utilizadas tanto para o sequenciamento do DNA mitocondrial como para o desenvolvimento do teste de paternidade foram obtidas de S.

*similis* que vieram a óbito em decorrência do tráfico de animais e tiveram a confirmação da identificação pela equipe de biólogos e médicos veterinários do Centro de Triagem de Animais Silvestres (Cetas) de Belo Horizonte-MG. As coletas foram permitidas pela Autorização para Atividades com Finalidade Científica número 024/2016, emitida pelo Instituto Estadual de Florestas (IEF) de Minas Gerais, pela autorização SISBIO (Ibama/ICMBio) número 56471 e pela autorização da Comissão de Ética no Uso de Animais (CEUA) número 37/2017. O acesso ao genoma de *Saltator similis* possui licença SisGen (Sistema Nacional de Gestão do Patrimônio Genético e do Conhecimento Tradicional Associado) número A4B3AD8. Os indivíduos foram identificados como “TF” seguido do número, sendo 52 amostras no total. Não há registro de relação de parentesco ou origem entre estas 52 amostras coletadas, podendo, portanto, conter indivíduos proximalmente relacionados assim como também de indivíduos distantes geograficamente e genealogicamente. Além disto, por ser uma espécie sem dimorfismo sexual, não havia informação sobre o sexo de cada indivíduo.

Já as amostras do criatório para a validação do teste de paternidade se tratavam de uma gota de sangue em papel FTA (figura 12), que contém reagentes que lisam as células sanguíneas e liberam o DNA, que é preservado da ação de nucleases e também inativam e inibem o crescimento de microrganismos, permitindo que as amostras sejam estocadas a temperatura ambiente por um longo período de tempo (MERCK 2018). Esta forma de coleta de amostra de DNA já é rotina nos criatórios, uma vez que este é o procedimento de envio de material para laboratórios para sexagem, realizada por PCR podendo ter como alvo a presença ou ausência do cromossomo W (presente somente nas fêmeas) (Unigen 2021) ou pelo número de bandas correspondente ao tamanho do gene CHD do cromossomo Z e cromossomo W, sendo que as fêmeas, por serem heterozigóticas, é possível observar duas bandas de tamanhos diferentes enquanto que nos machos, homozigóticos, apenas uma banda é observada (STABvida 2021). Estas amostras foram todas cedidas por um mesmo criatório do estado de Minas Gerais e foram identificadas como “AC” seguido do número, sendo 21 amostras no total. Estas amostras possuíam relação de parentesco conhecida e indicada pelo criatório que forneceu as amostras.



**Figura 12- Amostra de sangue de indivíduo de criatório em papel FTA.** Foi utilizada como referência de tamanho, à direita, marcações milimétricas.

#### 1.4.3 Extração de DNA

Diversos protocolos de extração de DNA foram testados de acordo com o tipo de amostra e o uso do DNA. Os DNAs extraídos foram testados quanto à sua qualidade e integridade em eletroforese em gel de agarose 0,7% em tampão TBE 0,5% (1.1 M Tris, 900 mM Borato, 25 mM EDTA, pH 8,3, Phoneutria) corado com brometo de etídio e revelado em luz UV. A quantificação foi realizada com Qubit (Thermo Fischer Scientific), que é mais sensível que quantificações realizadas por absorbância, pois detecta apenas fluoróforos que se ligam seletivamente ao DNA (Thermo Fischer Scientific 2021).

##### 1.4.3.1 Extração não-diferencial de DNA de amostra de músculo com Kit GTS (Phoneutria)

As amostras de músculo de 52 indivíduos oriundos do CETAS-BH tiveram seu DNA extraído com o kit GTS de acordo com as especificações do fabricante. Cerca de 40 mg de tecido muscular foram cortados em pequenos fragmentos aos quais se adicionou proteinase K, TEN9 (50 mM Tris-HCl, 100 mM EDTA, 200 mM NaCl, pH 9) e SDS (Dodecil sulfato de sódio) 20%, mantendo-se a solução a 56°C *overnight*. No dia seguinte, foi adicionado RNase por 1 hora a 56°C e em seguida iniciou-se o protocolo de extração com solução de lise, seguido de purificação com sílica e solução de lavagem A e B. Após a secagem, o DNA foi eluído em tampão TE (Tris 10mM, EDTA 1mM, pH 8). As amostras extraídas com este protocolo foram utilizadas no teste 1 e teste 2 do MiSeq.

##### 1.4.3.2 Extração não-diferencial de DNA de amostra de músculo com protocolo de fenol e clorofórmio

Este protocolo foi realizado para extração de DNA da amostra de músculo utilizada no sequenciamento total (primeiro sequenciamento realizado neste trabalho). Para isto, foi adicionado, em um tubo de 1,5ml, 60 µl de TEN9, 6 µl de SDS 20% e 4 µl de proteinase K a 10 mg/ml e incubado *overnight* a 56°C, juntamente com amostra de

músculo macerado. No dia seguinte, foram adicionados 70 µl de fenol equilibrado (pH alcalino, Phoneutria) e 70 µl de clorofórmio (Synth). Após a centrifugação por 5 minutos a 13.400 rpm, a fase superior foi coletada em novo tubo e então adicionados mais 70 µl de clorofórmio e novamente centrifugado com as mesmas condições anteriores. A fase superior foi coletada em um novo tubo e adicionados 50 µl de *beads* magnéticas (Beckman Coulter) por 5 minutos. Em seguida, o tubo foi colocado em uma placa magnética por 5 minutos e depois teve seu sobrenadante retirado. Ainda na placa magnética, foram adicionados 200 µl de álcool 80% (preparado no momento da extração com água ultrapura) e logo removido. Esta etapa foi repetida duas vezes. Em seguida, o álcool foi evaporado e então adicionados 25 µl de tampão TE (Tris 10 mM, EDTA 1 mM, pH 8, Phoneutria), que foi homogeneizado vagarosamente com o *pellet*. Após 3 minutos, o tubo retornou à placa magnética, onde permaneceu por mais 5 minutos. O sobrenadante, contendo o DNA purificado, foi transferido para um novo microtubo.

#### *1.4.3.3 Extração não-diferencial de DNA de amostra de sangue em papel FTA com protocolo de precipitação em isopropanol*

O pedaço de papel contendo a amostra de sangue foi levado ao vortex em um tubo com 450 µl de SEB (*Stain Extraction Buffer* – 10 mM Tris-HCl, 10 mM EDTA, 100 mM NaCl, 2% SDS, pH 7,5) e 20 µl de proteinase K 10mg/mL e em seguida incubado em banho-maria a 56°C por 18 horas. O pedaço de papel foi então retirado e o líquido foi transferido para uma coluna (spin basket) e centrifugado a 10.000 rpm por 2 minutos. Ao filtrado foi adicionado 500 µl de clorofórmio, homogeneizado e então centrifugado a 15.000 rpm por 3 minutos. O sobrenadante foi transferido para um novo tubo, ao qual foi adicionado 500 µl de isopropanol e 50 µl de NaCl 5M, e então mantido a -20°C por 30 minutos. Posteriormente houve a centrifugação a 14.000 rpm por 4 minutos, adição de 200 µl de etanol 70% ao *pellet* e em seguida uma nova centrifugação de 15.000 rpm por 3 minutos. Estas últimas etapas foram realizadas duas vezes. Após a secagem em temperatura ambiente, foi adicionado 40 µl de tampão TE, homogeneizado e incubado a 56°C por 10 minutos. Este protocolo foi cedido pelo laboratório de DNA da Polícia Civil de Minas Gerais.

#### *1.4.3.4 Extração diferencial de mtDNA*

Foi utilizado o protocolo de França et. al (2016), no qual o tecido foi homogeneizado com adição de 500 µl de solução contendo 0,32M de sacarose, 1mM de EDTA, 10mM de Tris-HCl com pH 7,8 e posteriormente centrifugado a 3.500 rpm por 10 minutos a 4°C. O sobrenadante foi coletado e novamente centrifugado para remover



restos de DNA nuclear. O sobrenadante foi então centrifugado a 12.000 rpm por 20 minutos. No *pellet* resultante estava o mtDNA, e, portanto, o sobrenadante foi descartado. Adicionou-se 100 µl de Tris 10mM com pH 8, 2 µl de tampão 10X *Reaction Buffer* with MgCl<sub>2</sub> for DNase I (Thermo Scientific) e 1 µl de DNase (Thermo Scientific). Após 30 minutos a 37°C, a DNase foi inativada por 10 minutos a 65°C e adição de 1 µl de EDTA 100mM (Thermo Scientific). Em seguida, foram adicionados 5 µl de proteinase K a 10 mg/mL e 5 µl de SDS 20% e deixado a 56°C *overnight*. No dia seguinte, foi adicionado 1 µl de RNase e mantida a mesma temperatura. Em seguida, foi adicionado 100 µl de fenol e 100 µl de clorofórmio e centrifugado a 13.400 rpm por 5 minutos. A fase superior foi coletada em novo tubo ao qual foram adicionados 100 µl de clorofórmio e novamente centrifugado. A fase superior foi coletada, adicionados 200 µl de solução de precipitação de DNA (Phoneutria) e mantido no freezer por 30 minutos e em seguida centrifugado a 13.400 rpm por 10 minutos. Ao sobrenadante foram adicionados 300 µl de álcool 80% (preparado com água ultrapura e no momento da extração) e centrifugado a 13.400 rpm por 2 minutos. Esta etapa foi realizada duas vezes. O *pellet* ficou por 10 minutos a 37°C e depois foram adicionados 15 µl de tampão TE (Tris 10mM, EDTA 1mM, pH 8) e homogeneizado, e então mantido a 37°C por 15 minutos. O espécime utilizado para extração do mtDNA foi depositado no centro de coleções taxonômicas da Universidade Federal de Minas Gerais (voucher DZ7289).

#### 1.4.4 Sequenciamentos no MiSeq

DNA total de um indivíduo de *S. similis* extraído pelo protocolo de fenol e clorofórmio foi sequenciado utilizando estratégia *paired-end* 301/301. Essas etapas foram realizadas por membros do Laboratório de Biotecnologia e Marcadores Moleculares em 2016 e não foi suficiente para se obter a sequência completa do mtDNA, mas foi suficiente para a detecção de regiões microssatélites. Para o DNA mitocondrial e teste 2 e 3 com microssatélites, o sequenciamento foi realizado utilizando a estratégia de *single-read* de 300 ciclos. Para o teste 1 no MiSeq com os microssatélites, foi utilizada estratégia *paired-end* 181/121. A estratégia utilizada nos testes 1 a 3 no MiSeq foi a de *pseudo-multiplex*, no qual cada par de *primer* é utilizado individualmente para amplificar o DNA alvo de cada indivíduo e depois os *amplicons* de todos os *loci* do mesmo indivíduo são agrupados para a PCR *index*, conforme descrito por Pimentel (2018).

No teste 1 do MiSeq, foram sequenciadas 10 amostras escolhidas ao acaso oriundas do CETAS-BH (amostras TF3, TF9, TF14, TF16, TF17, TF18, TF21, TF23, TF25 e TF26) amplificadas para 25 conjuntos de *primers*. O Objetivo deste primeiro teste com número reduzido de amostras foi fazer uma seleção prévia dos *loci* (formação de dímeros de *primers*, falhas no sequenciamento ou polimorfismo muito baixo), reduzindo, assim, o número de *loci* para a próxima etapa com todas as amostras TF. O teste 2 no MiSeq foi realizado com 21 *loci* previamente selecionados e 47 indivíduos TF (os indivíduos TF9, TF16, TF18, TF21 e TF23 foram sequenciados novamente nesta etapa), totalizando 52 indivíduos TF sequenciados. Como as amostras de criatório (amostras AC1-AC21) foram recebidas posteriormente, estas foram então sequenciadas no teste 3 do MiSeq.

As bibliotecas, com exceção do DNA total (*indexado* com o kit Nextera), foram *indexadas* com o kit Nextera XT (Illumina). Todos os sequenciamentos foram analisados quanto à qualidade das *reads* geradas utilizando FastQC 0.11.8 ([www.bioinformatics.babraham.ac.uk/projects/](http://www.bioinformatics.babraham.ac.uk/projects/)).

#### 1.4.5 Montagem de *contigs*

Foram utilizados os seguintes programas para montagem dos *contigs* neste trabalho: MIRA 4.0 (Chevreux 1999), CLC Genomics Workbench 9.0.1 (CLC bio QIAGEN, Alemanha, <https://www.qiagenbioinformatics.com/>) e Geneious 11.1.15 (<https://www.geneious.com>).

#### 1.4.6 Anotação do mtDNA

A anotação automática do *contig* mitocondrial foi feita com MITOS WebServer (Bernt 2013). A anotação foi corrigida manualmente com auxílio do *software* ExPASy (Gasteiger 2003) e comparação com sequências disponíveis no GenBank.

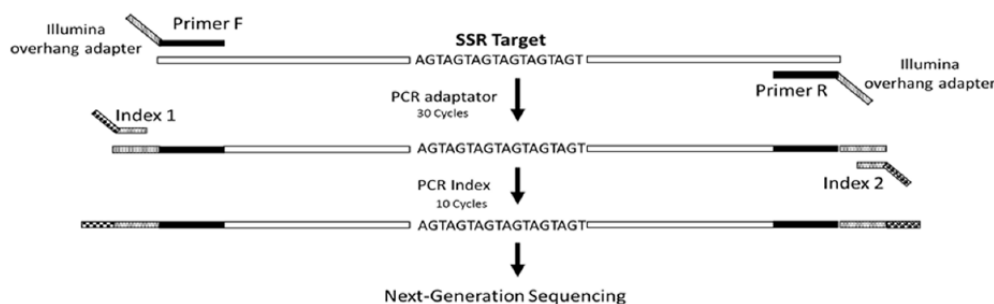
#### 1.4.7 Busca por microssatélites

Os *contigs* de DNA total gerados utilizando o programa MIRA 4.0 (Chevreux 1999) foram submetidos ao programa MSatCommander 1.0.8 (Faircloth 2008) para encontrar sequências do tipo microssatélites (parâmetro: mínimo de seis repetições) e também para desenhar *primers* que amplificassem estas regiões (parâmetros: 40-60% de

conteúdo GC, tamanho mínimo do *primer* de 18pb e máximo de 25pb, tamanho do *amplicon* entre 100 e 450pb e temperatura de *melting* entre 53 e 65°C).

#### 1.4.8 Primers

Todos os *primers* selecionados foram sintetizados pela empresa Integrated DNA Technologies (Coralville, Iowa, EUA). Os *primers* foram diluídos em tampão Tris-EDTA pH 8,0 (Phoneutria) para a concentração de 200 pmol/μl para a solução estoque. A solução de uso foi diluída em água ultrapura para biologia molecular na concentração de 5 pmol/μl. Os *primers* que demonstraram bons resultados foram sintetizados novamente, porém, contendo as sequências dos adaptadores nas suas extremidades de acordo com o Protocolo 16S Metagenomic Sequencing Library Preparation (15044223 B) desenvolvido pela Illumina (*Amplicon*, Clean-Up, and *Index* 2013). Ao *primer forward* foi adicionada a seguinte sequência: 5' TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-[sequência do *primer*], enquanto que ao *primer reverse* foi adicionada a seguinte sequência: 5'GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-[sequência do *primer*]. Estas sequências denominadas adaptadores permitem que as sequências *index* (que individualizam as amostras de NGS) se liguem ao fragmento de interesse, conforme mostrado na figura de Pimentel (2018), abaixo (figura 13).



**Figura 13- Representação esquemática do protocolo de amplificação das sequências alvo para sequenciamento.** Aos *primers* selecionadas foram adicionadas sequências de adaptadores a suas extremidades, que permitem que as sequências *index* (utilizadas para individualização da amostra) se liguem ao fragmento alvo. Imagem retirada de Pimentel 2018 e reproduzida com autorização de Creative Commons Attribution License.

#### 1.4.9 PCRs para padronização dos pares de *primers*

A padronização das PCRs para os conjuntos de *primers* sintetizados foi realizada com diferentes temperaturas de anelamento (50°C, 55°C e 60°C), quantidades de *primer* (5 μl, 2 μl e 1 μl), adição de DMSO (3% e 5%), número de ciclos na etapa de anelamento

e extensão (25, 30 e 35 ciclos) e pré-mix IC, IB e IVB 10x concentrado (Phoneutria). A composição do pré-mix IVB não é divulgada pelo fabricante, enquanto que a composição do pré-mix IC é 500 mM KCl; 100 mM Tris-HCl pH 8,4; 1% Triton X-100 20 mM; MgCl<sub>2</sub> e a do pré-mix IB é similar à do pré-mix IC, porém com 15 mM de MgCl<sub>2</sub>. A quantidade de cada reagente utilizado na PCR está detalhada na tabela 2. Inicialmente as reações foram padronizadas com 5 ng de DNA. As reações foram feitas em termociclador Applied Biosystems, modelo Veriti. A programação do termociclador está reproduzida no quadro abaixo (tabela 3). Para visualizar os resultados das PCRs, 5 µl de *amplicon* com 1 µl de tampão de amostra foram aplicados em gel de agarose 1,5% ou 2% em tampão TBE 0,5% (1.1 M Tris, 900 mM Borato, 25 mM EDTA, pH 8,3, Phoneutria) corado com brometo de etídio e revelado em luz UV.

**Tabela 2- Reagentes utilizados para montagem da PCR.** As quantidades destacadas em negrito foram alteradas em algumas reações durante a padronização. Quando adicionado DMSO ou alterado a quantidade de algum outro reagente, havia alteração no volume de água para que o volume final de 25 µl fosse mantido.

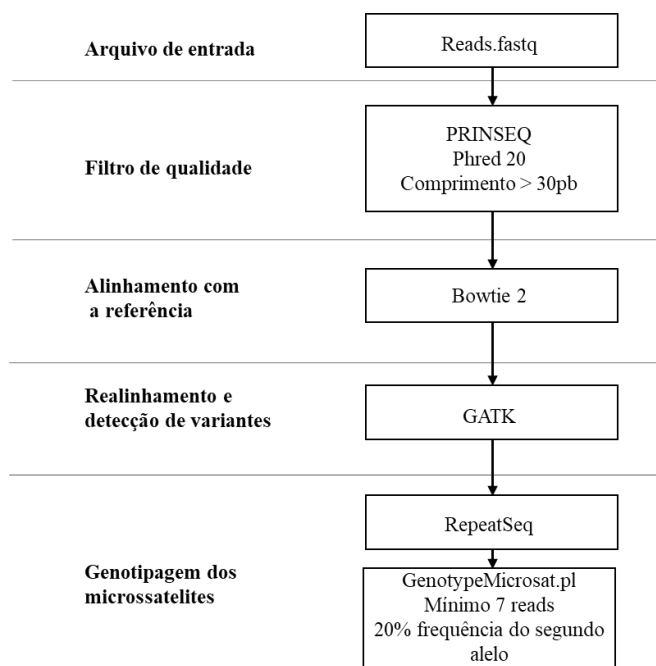
Reagente	Quantidade por reação (µl)
Água ultrapura	<b>9,5</b>
Pré-mix 2X (Phoneutria)	12,5
<i>Primer Forward</i> (5 pmol/µl)	<b>1</b>
<i>Primer Reverse</i> (5 pmol/µl)	<b>1</b>
DNA	<b>1</b>
<b>Volume total</b>	25

**Tabela 3- Programação do termociclador.** Os números destacados em negrito foram alterados em algumas reações durante a padronização.

Temperatura (°C)	Tempo	
94	2 min	Desnaturação
95	30 seg	Desnaturação
<b>60</b>	45 seg	Anelamento
72	40 seg	Extensão
		<b>30 ciclos</b>
72	5 min	Extensão final
4	∞	

#### 1.4.10 Análise do sequenciamento de microssatélites e genotipagem

As *reads* geradas pelos sequenciamentos no MiSeq foram analisadas com o *pipeline* utilizado por Pimentel et al. (2018 e 2020) e esquematizado na figura 14, desenvolvido pela Doutora Izinara Rosse, que também colaborou com a utilização deste *pipeline* neste trabalho. A primeira etapa realizada no processamento dos dados de sequenciamento foi a filtragem e trimagem das *reads* com a ferramenta PRINSEQ (*Preprocessing and Information of SEquences*) (Schmieder and Edwards, 2011). Este passo é importante para garantir que as análises posteriores não serão comprometidas por seqüências de baixa qualidade que podem comprometer as conclusões. Os arquivos de entrada foram submetidos em formato FASTQ (formato baseado em texto que contém as seqüências de nucleotídeos e seus índices de qualidade correspondentes) e foram filtradas as *reads* que possuíam qualidade  $Q < 20$  e tamanho menor que 30pb. O valor de qualidade (Q Score), ou Phred score, é um valor inteiro que representa a probabilidade estimada de erro, ou seja, a chance de a base estar incorreta, de acordo com a tabela 4.



**Figura 14- Representação esquemática da estratégia de genotipagem utilizada.** O *pipeline* utilizado e a representação esquemática foram baseados em Pimentel 2018 e 2020.

**Tabela 4- Valores de qualidade Phred e probabilidade de acurácia da base.**

Valor de qualidade Phred	Probabilidade de a base estar incorreta	Acurácia da base
10	1 em 10	90%
20	1 em 100	99%
30	1 em 1.000	99.9%
40	1 em 10.000	99.99%
50	1 em 100.00	99.999%

Informações da tabela retiradas de Illumina 2011.

Após a filtragem das *reads*, foi feita a *indexação* e alinhamento das *reads* utilizando o Bowtie2 (Langmead e Salzberg 2012), etapa em que as *reads* são alinhadas contra um genoma de referência (no caso deste trabalho, as sequências referências utilizadas nesta etapa são as sequências nas quais foram identificadas as regiões microsatélites, contendo desde a região do *primer forward* até a região do *primer reverse*). Conforme especificado na linha de comando, os arquivos gerados pelo Bowtie2 foram do tipo SAM (Sequence Alignment/Map) e, para melhorar a performance, foram convertidos, utilizando o SAMtools (Li 2009), em arquivos do tipo BAM (Binary Alignment/Map). Este tipo de arquivo é a representação binária de SAM e contém exatamente as mesmas informações (Li 2009). O SAMtools também é utilizado para a detecção de variantes.

Como parte das análises, foram utilizados também os *softwares* Picard (<http://broadinstitute.github.io/picard/>) e GATK (*Genome Analysis Toolkit*) (McKenna et al. 2010). Posteriormente foi utilizado o programa RepeatSeq (Highnam 2013), que utiliza alinhamentos pré-existentes e utiliza arquivos do tipo vcf para identificação de variações diretamente pelas diferenças entre a sequência da *read* e a sequência referência (Cao 2015). Por fim, foi utilizado um *script* em linguagem Perl denominado GenotypeMicrosat.pl para converter os resultados em formato que comumente é utilizado nos programas de genética de populações e também para evitar falsos negativos. Para cada *locus* foi determinado o genótipo seguindo os seguintes critérios: máximo de 2 alelos por indivíduo por *locus*, mínimo de 7 *reads* por *locus* por indivíduo contendo a sequência completa e mínimo de 20% das *reads* correspondente ao segundo alelo para ser considerado heterozigoto.

Além do *pipeline* mencionado acima, também foi utilizado o programa CHIIMP (*Computational High-throughput Individual Identification through Microsatellite Profiling*) (Barbian 2018), que utiliza o pacote R para realizar a análise de genótipos de

microsatélites. Os parâmetros utilizados foram: *nrepeats* 5, *stutter.count.ratio\_max* 0.7, *artifact.count.ratio\_max* 0.7, *fraction.min* 0.2 e *counts.min* 7.

Conjuntamente com os programas mencionados, as análises também foram realizadas com os programas Geneious (<https://www.geneious.com>) e MEGA 6 (Tamura 2013). Nestes, as *reads* de cada *locus* foram selecionadas de acordo com a sequência referência, e em seguida as mesmas foram alinhadas e então foi feita a análise dos microsatélites de forma manual, de forma a verificar os resultados obtidos nas outras análises bioinformáticas. No mapeamento contra a referência no programa Geneious os parâmetros utilizados foram: máx. *mismatches* por *read* 30%, máx. *gaps* por *read* 60%, máx. tamanho de *gap* 180.

#### 1.4.11 Análises genotípicas e de paternidade

Após obter o genótipo de cada alelo para cada *locus* e indivíduo, foi utilizado o programa Cervus 3.0.7 (Marshall 1998) com parâmetros padrão definidos pelo programa, e que, através dos dados da genotipagem, calcula o número de alelos por *locus*, a heterozigosidade esperada ( $H_e$ ) e observada ( $H_o$ ) (equação 8.4, Nei 1987), frequência de alelos nulos (Summers e Amos 1997) e conteúdo de informatividade polimórfica (PIC) (Botstein 1980). Estas análises também foram realizadas com o programa STRAF 1.0.5 (*STR Analysis for Forensics*) (Gouy 2017). O teste de desvio do equilíbrio de Hardy-Weinberg (equação 7.8 Nei 1987) também foi realizado com o programa Cervus com avaliação da significância com correção de Bonferroni, correção de Yates para o valor de Chi-quadrado e frequência mínima esperada de 5.

A análise de paternidade também foi realizada com o programa Cervus, cujo primeiro passo é realizar uma simulação de paternidade para estimar o poder de resolução de uma série de *loci* codominantes levando em consideração as suas frequências alélicas e também para estimar os valores verossimilhança estatística LOD (log natural da razão de verossimilhança geral) para permitir a avaliação estatística das relações de parentesco estabelecidas. Para a simulação de paternidade foram utilizados os genótipos das famílias oriundas de criatório com os parâmetros: par de pais (sexos desconhecidos), 100.000 simulações de genótipos da prole, proporção de erros de genótipos da prole de 0,01, número mínimo de *loci* de 4, nível de confiança estrito de 95% e relaxado de 80%. Com o intervalo de confiança estrito estabelecido, a simulação identifica o valor crítico de LOD para que 19 a cada 20 LOD com valor superior a este são atribuídos aos progenitores candidatos que são os verdadeiros progenitores, enquanto que apenas um a cada 20 LOD

com valor superior sejam referente a possíveis progenitores que não são os verdadeiros progenitores, resultando em uma taxa de falsos positivos de 1 em 20. A escolha de se utilizar a opção de sexos desconhecidos foi devido ao fato de que a espécie não possui dimorfismo sexual e, assim, animais sem sexagem ou com sexagem errônea poderiam interferir nas análises. Para a análise de paternidade, todos os possíveis progenitores foram comparados com todos os filhotes (escolhendo a opção do programa “progenitor aparece como uma coluna para toda a prole”) e o resultado continha os dois progenitores mais prováveis de serem os pais de determinado filhote de acordo com a pontuação de LOD.

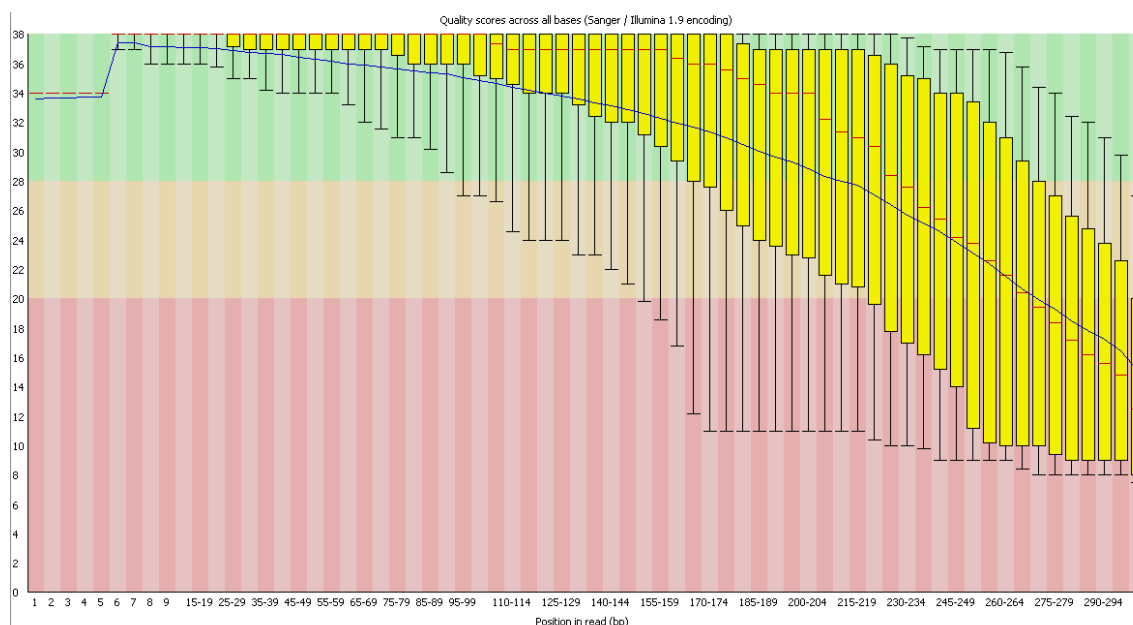


## 1.5 Resultados e discussão

### 1.5.1 Teste de filiação de *S. similis*

#### 1.5.1.1 Sequenciamento do DNA total

O sequenciamento do DNA de *S. similis* gerou 8,81 GB de dados e resultou em 35.521.090 *reads* (R1 e R2). A figura 15, abaixo, mostra a qualidade das *reads* para R1. Assim como pode ser observado nessa imagem, é comum a queda na qualidade com o passar dos ciclos do sequenciamento. Isso é causado pelo fato de que moléculas únicas em um *cluster* ficam fora de sincronia, e em cada ciclo do sequenciamento, moléculas individuais do *cluster* podem não conseguir obter uma base recém-sintetizada (Dohm 2008, Tan 2019). A montagem utilizando o programa MIRA 4.0 resultou em 387.395 *contigs*, que foram então submetidos ao programa MSatCommander para realizar a busca por microssatélites.



**Figura 15-** Qualidade das bases das *reads* R1 do sequenciamento do DNA total. O eixo Y mostra o valor de qualidade e o eixo X a posição da base na *read*. O *background* do gráfico é dividido em qualidade muito boa (verde), qualidade razoável (laranja) e qualidade ruim (vermelho). A linha vermelha representa a mediana e a linha azul a média da qualidade. A caixa amarela representa o inter-quartil (25-75%). O gráfico foi gerado pelo programa FastQC 0.11.8 ([www.bioinformatics.babraham.ac.uk/projects/](http://www.bioinformatics.babraham.ac.uk/projects/)).

#### 1.5.1.2 Seleção de pares de primers para microssatélites

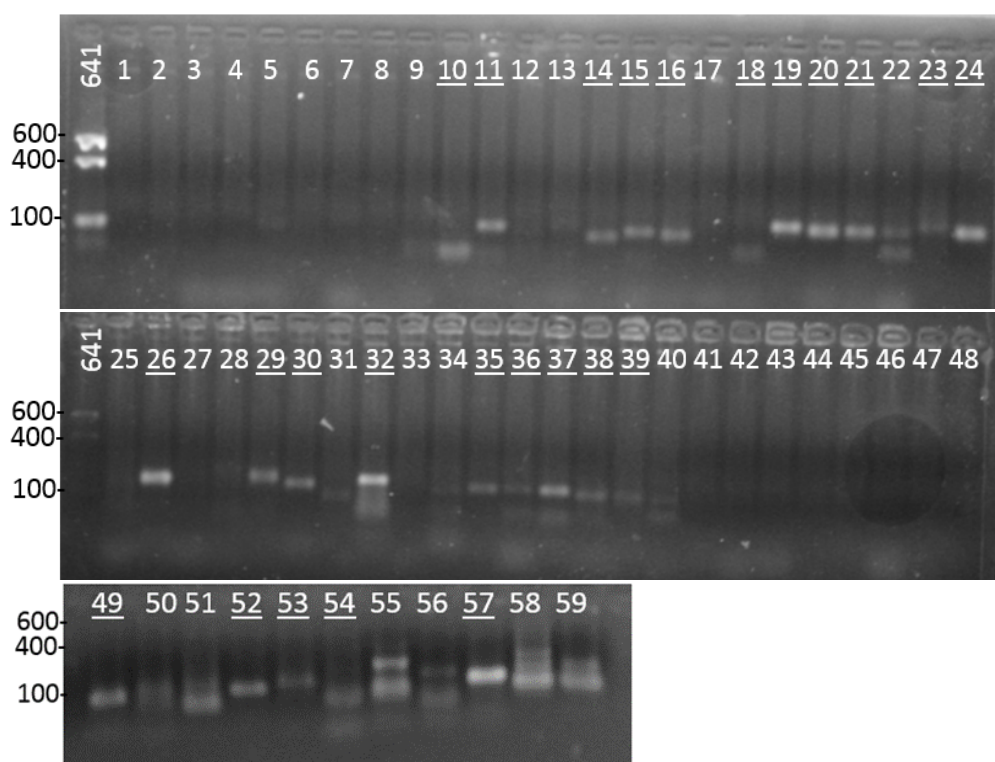
A análise do MSatCommander encontrou 3.827 regiões microssatélites e foram desenhados 2.231 conjuntos de *primers* (tabela 5). Estes números mostram que, atualmente, com o NGS, comparado às metodologias tradicionais, a descoberta de

microssatélites e desenho de marcadores não é mais uma etapa limitante, sendo estas etapas a triagem e validação dos *loci* (Wei 2014). Vale ressaltar também que o número de regiões microssatélite encontradas nesta análise é dependente do *software* e parâmetros utilizados na busca, e que, além do MSatCommander (Faircloth 2008) utilizado, existem outras ferramentas similares para este tipo de análise, como MISA (MicroSAtellite identification tool) (Thiel 2003), PAL\_FINDER (Castoe 2012), Phobos (plugin do *software* Geneious) (Keatse 2012), SSR Locator (Da Maia 2008) e GMATo (Wang 2013), e que podem trazer resultados ligeiramente diferentes (Hodel 2016). Nesta busca, foram excluídos os mononucleotídeos, uma vez que estes são mais difíceis de se trabalhar com precisão (Sun 2006) e, portanto, são frequentemente eliminados dos estudos (Kim 2008), sendo raramente incluídos em trabalhos que utilizam microssatélites (Guichoux 2011).

Dentre os conjuntos de *primers* desenhados que possuíam *amplicon* esperado entre 100 e 160pb, foram selecionados 59 conjuntos de *primers* para serem sintetizados, conforme mostrado na tabela 5 e detalhado na tabela 6. Este tamanho máximo de fragmento foi definido devido à limitação de tamanho das *reads* do MiSeq e queda da qualidade após um certo número de bases sequenciadas. A grande maioria dos *primers* escolhidos possuíam temperatura de *melting* entre 58 e 60°C, temperatura considerada ótima (Butler 2005, Hill 2009). Tais conjuntos de *primers* foram testados na PCR com os pré-mix IB, IC e IVB, temperaturas de anelamento de 60°C e 65°C e 30 ciclos, sendo os melhores resultados obtidos quando utilizado o pré-mix 1B e 60°C (figura 16). Com base nestes resultados, optou-se por continuar o desenvolvimento do teste de paternidade apenas com aqueles conjuntos de *primers* em que foi possível visualizar a banda no gel de agarose, resultando na escolha de 25 conjuntos de *primers* para a próxima etapa de síntese com adaptadores (cujos números aparecem sublinhados na figura 17 e em negrito e com fundo acinzentado na tabela 6), sendo que nenhum destes possuía motivo de repetição de hexanucleotídeos.

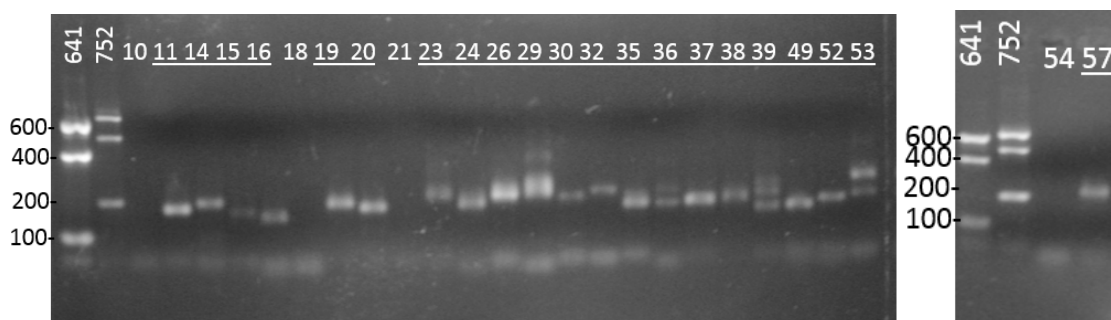
**Tabela 5- Quantidade de microssatélites encontrados e *primers* sintetizados.** Os parâmetros da busca pelos microssatélites no MSatCommander (Faircloth 2008) foram: 40-60% de conteúdo GC, tamanho mínimo do *primer* de 18pb e máximo de 25pb, tamanho do *amplicon* entre 100 e 450pb e temperatura de *melting* entre 53 e 65°C. Os *primers* sintetizados foram escolhidos dentre os que possuíam tamanho de *amplicon* entre 100 e 160pb.

	<b>Microssatélites encontrados pelo MSatCommander</b>	<b>Pares de <i>primers</i> sintetizados</b>
Dinucleotídeos	2.081	21
Trinucleotídeos	472	18
Tetranucleotídeos	726	14
Pentanucleotídeos	433	4
Hexanucleotídeos	114	2
<b>Total</b>	<b>3.827</b>	<b>59</b>



**Figura 16- Resultado da PCR com os pares de *primers* sintetizados sem adaptadores em gel de agarose.** A figura mostra gel de agarose 1,5% em TBE 0,5X corado com brometo de etídio e revelado em luz UV referente a 5µl de produtos da PCR com todos os conjuntos de *primers* sintetizados. As reações foram realizadas utilizando pré-mix IB, 30 ciclos e temperatura de anelamento de 60°C. Foi utilizado o padrão de peso molecular 641 (Phoneutria), que apresenta as bandas de 600,400 e 100 pb. Os conjuntos selecionados para serem posteriormente sintetizados com adaptadores têm seus respectivos números sublinhados na imagem.

Com o intuito de tornar a realização do teste de filiação mais ágil e uma possível padronização *multiplex* futuramente, optamos por utilizar as mesmas condições para todos os conjuntos de *primers* com adaptadores. Para os 25 conjuntos, obteve-se resultados melhores quando utilizados com o pré-mix IVB e 30 ciclos. Os resultados foram semelhantes quando utilizadas as temperaturas de anelamento de 55°C, 60°C ou 65°C. Assim, optou-se por utilizar 60°C como padrão. Conforme mostrado na figura 17, nestas condições, houve amplificação de todos, exceto dos conjuntos 10, 18, 21 e 54. Estes quatro conjuntos não seguiram para os testes com o MiSeq, restando apenas 21 conjuntos de *primers* com adaptadores (cujos números aparecem sublinhados na figura 8). A adição das sequências dos adaptadores nas extremidades dos *primers* resulta em um aumento de 67 pb nos *amplicons* esperados, e, portanto, para melhor visualização no gel de agarose, foi utilizado o padrão de peso molecular 641 e 752 (Phoneutria), para que constasse as referências de 100 e de 200 pb.



**Figura 17- Resultado da PCR com os conjuntos de *primers* com adaptadores.** A figura mostra gel de agarose 2% em TBE 0,5X corado com brometo de etídio e revelado em luz UV referente a 5µl de produto de PCR com todos os conjuntos de *primers* com adaptadores. As reações foram realizadas utilizando o pré-mix IVB (Phoneutria), 30 ciclos e temperatura de anelamento de 60°C. Foi utilizado o padrão de peso molecular 641 e 752 (Phoneutria), que apresentam as bandas de 600, 400 e 100 pb e 700, 500 e 200 pb, respectivamente. Ocorreu amplificação de todos os conjuntos, exceto 10, 18, 21 e 54.

**Tabela 6- Conjuntos de *primers* sintetizados e suas características.** As sequências dos *primers* estão sem os adaptadores. Os conjuntos cujos números aparecem em negrito e com fundo acinzentado são aqueles que foram posteriormente sintetizados com adaptadores e, após padronização da PCR, seguiram para o teste 1 no MiSeq.

Tabela a disposição sob consulta com o  
Prof. Dr. Evanguedes Kalapothakis.

TM= temperatura de *melting*, %GC= conteúdo de Guanina e Citosina.

Não possuímos informações sobre a localização dos microssatélites ao longo do genoma, podendo ocorrer de duas ou mais regiões estarem próximas no genoma e serem herdadas conjuntamente. Além disto, se a região microssatélite em estudo estiver em um cromossomo sexual, o padrão de herança será diferente. Vale ressaltar que, em aves, a fêmea que é o sexo heterogamético, possuindo os cromossomos W e Z, enquanto os machos possuem duas cópias do cromossomo Z (Gianaroli 2013).

Em uma tentativa de obter alguma informação sobre a localização no genoma das regiões selecionada e também verificar se poderia ter ocorrido contaminação e consequentemente amplificação de genoma humano no sequenciamento a partir do qual foram extraídas as regiões microssatélites de *Saltator similis*, foi realizado o alinhamento de cada sequência referência dos 21 *loci* selecionados para a próxima etapa contra o banco de dados GenBank utilizando o Blastn (*Basic Local Alignment Search Tool- Nucleotide*) (NCBI 1988). O resultado destes alinhamentos (mostrados na tabela 7, abaixo) indicou que não houve contaminação de DNA humano, uma vez que este tem sua sequência completa conhecida depositada e, portanto, se houvesse similaridade, o alinhamento retornaria esta informação. Das 21 sequências alinhadas, apenas oito retornaram com alguma similaridade, e estas sempre com espécies de aves (nenhuma de ocorrência no território brasileiro). Para L11, L15, L24 e L30, o alinhamento mostrou variação no número de repetições dos microssatélites e bastante similaridade na região de ligação dos *primers*, indicando que estes *loci* poderiam ser considerados para o estudo de microssatélites de outras espécies. O alinhamento de L32 também trouxe diferença na quantidade de repetições do microssatélite, porém também ocorreram diferenças nas regiões de ligação dos *primers*, o que poderia dificultar a utilização destes mesmos *primers*.

O alinhamento no GenBank retornou apenas algumas informações e não foi possível verificar se alguma região selecionada está em cromossomo sexual. Possivelmente isso ocorreu devido à existência de relativamente poucas sequências gênicas de aves no banco de dados, ressaltando, mais uma vez, a importância de projetos que visam o sequenciamento de DNA de diversas espécies, como, por exemplo, o projeto Bird 10K, mencionado anteriormente, que pretende sequenciar o genoma mitocondrial e/ou o genoma nuclear de diversas espécies.

**Tabela 7- Similaridade da sequência referência de cada *locus* contra o banco de dados GenBank.**

<i>Locus</i>	Quantidade de sequências retornadas	Max identidade (%)*	Espécie com máxima identidade	Gene/cromossomo associado
L11	13	97,18	<i>Taeniopygia guttata</i>	<i>mitochondrial elongation factor 1 (MIEF1)</i>
L14	sem similaridade			
L15	16	98	<i>Geospiza fortis</i>	<i>insulin like growth factor 1 receptor (IGF1R)</i>
L16	sem similaridade			
L19	sem similaridade			
L20	1	97,87	<i>Erithacus rubecula</i>	cromossomo 4
L23	sem similaridade			
L24	5	99,09	<i>Zonotrichia albicollis</i>	<i>SET binding protein 1 (SETBP1)</i>
L26	sem similaridade			
L29	sem similaridade			
L30	70	99,3	<i>Geospiza fortis</i>	<i>BCL6 corepressor like 1 (BCORL1)</i>
L32	1	82,5	<i>Corvus brachyrhynchos</i>	não determinado
L35	sem similaridade			
L36	sem similaridade			
L37	4	100	<i>Apteryx australis mantelli</i>	não determinado
L38	sem similaridade			
L39	sem similaridade			
L49	12	98,59	<i>Zonotrichia albicollis</i>	<i>cat eye syndrome critical region protein 2 homolog</i>
L52	sem similaridade			
L53	sem similaridade			
L57	sem similaridade			

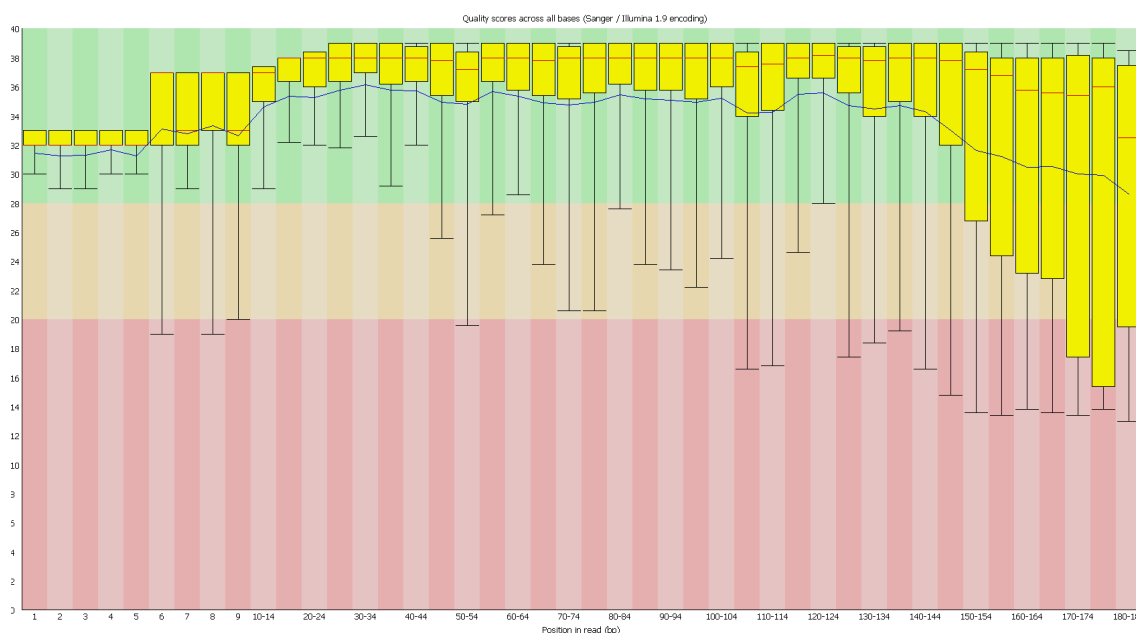
\*similaridade apenas de fragmentos e não sequência referência completa.

#### 1.5.1.3 Teste 1 no MiSeq

A proposta deste teste 1 no MiSeq foi avaliar a qualidade dos *loci* selecionados quando submetidos ao NGS e também obter informação prévia sobre o polimorfismo das regiões selecionadas, com o intuito de se reduzir custos e tempo de manipulação no teste 2 no MiSeq ao se retirar dos experimentos os *loci* que não apresentaram boas perspectivas.

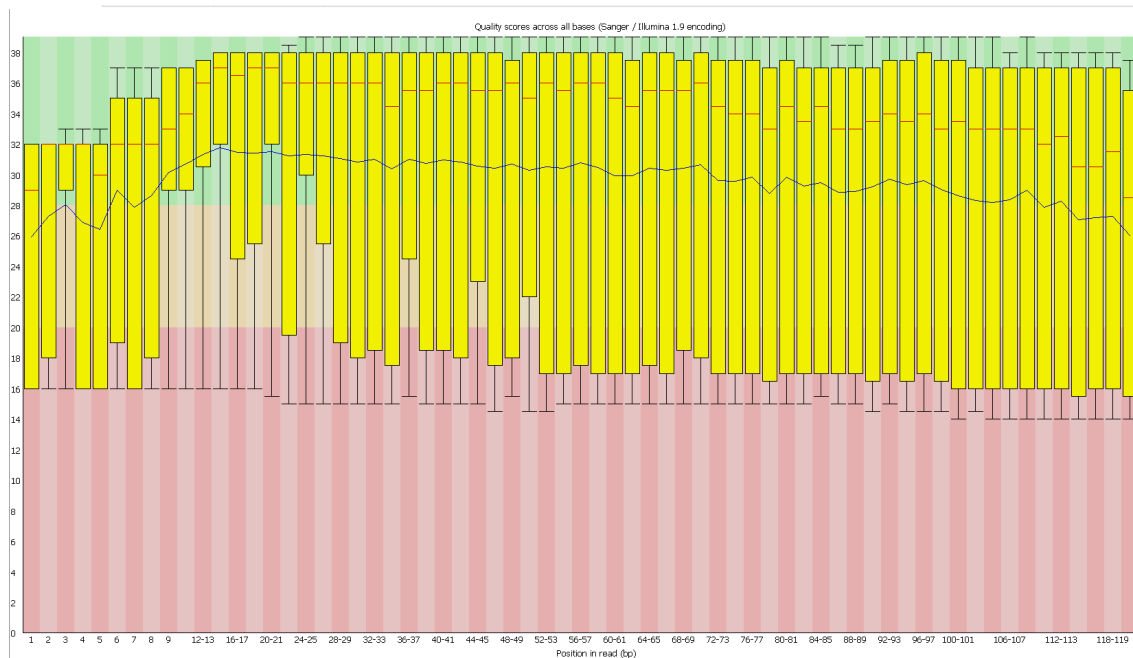
O sequenciamento dos *amplicons* dos 21 conjuntos de *primers* selecionados para dez indivíduos resultou em 55.900 *reads* de tamanhos de 35 a 181 pb. Conforme mostrado na figura 18, a qualidade das bases foi majoritariamente considerada muito boa até cerca

de 150pb, ocorrendo um decréscimo na qualidade das bases nas posições mais distantes do começo da sequência. Apesar de ter sido utilizada a estratégia de sequenciamento *paired-end*, apenas as *reads* R1 foram utilizadas nas análises, uma vez que a qualidade do sequenciamento das *reads* R2 foi ruim (figura 19). De forma geral, os dados gerados continham uma boa proporção de *reads* por indivíduo. Com o mapeamento das *reads* contra a referência, utilizando o programa Geneious 11.1.15 (<https://www.geneious.com>), obtivemos o número de *reads* de cada indivíduo para cada *locus*, e, em alguns casos, poucas *reads* foram geradas para determinados *loci* resultando em comprometimento das análises, como, por exemplo, a existência de apenas duas *reads* do L49 do indivíduo TF18 (tabela 8). Este mapeamento também trouxe a informação de quantas *reads* possuíam qualidade superior a Q30 e quantas não foram mapeadas contra nenhuma referência. O número de *reads* presentes no controle negativo não se trata necessariamente de contaminação e são abordadas no tópico 1.5.1.8 - controle negativo.



**Figura 18- Qualidade das bases das *reads* R1 do sequenciamento do teste 1 no MiSeq.** As diferentes amostras apresentaram distribuição de qualidade da base sequenciada similar, sendo a mostrada na imagem referente às *reads* R1 do indivíduo TF14. O eixo Y mostra o valor de qualidade e o eixo X a posição da base na *read*. O *background* do gráfico é dividido em qualidade muito boa (verde), qualidade razoável (laranja) e qualidade ruim (vermelho). A linha vermelha representa a mediana e a linha azul a média da qualidade. A caixa amarela representa o interquartil (25-75%). O gráfico foi gerado pelo programa FastQC 0.11.9 ([www.bioinformatics.babraham.ac.uk/projects/](http://www.bioinformatics.babraham.ac.uk/projects/)).





**Figura 19- Qualidade das bases das *reads* R2 do sequenciamento do teste 1 no MiSeq.** As diferentes amostras apresentaram distribuição de qualidade da base sequenciada similar, sendo a mostrada na imagem referente às *reads* R2 do indivíduo TF14. O eixo Y mostra o valor de qualidade e o eixo X a posição da base na *read*. O *background* do gráfico é dividido em qualidade muito boa (verde), qualidade razoável (laranja) e qualidade ruim (vermelho). A linha vermelha representa a mediana e a linha azul a média da qualidade. A caixa amarela representa o interquartil (25-75%). O gráfico foi gerado pelo programa FastQC 0.11.9 ([www.bioinformatics.babraham.ac.uk/projects/](http://www.bioinformatics.babraham.ac.uk/projects/)).

**Tabela 8- Reads geradas no teste 1 no MiSeq.** Número de *reads* geradas por indivíduo no teste 1 do Miseq, porcentagem de *reads* com qualidade superior a Q30, número de *reads* mapeadas, número de *reads* não mapeadas e número de *reads* por *locus*. O controle negativo passou por todas as reações e etapas das demais amostras, sendo a única diferença a ausência de DNA na PCR

indivíduo	≥Q30(%)	número total de <i>reads</i>	número de <i>reads</i> mapeadas	número de <i>reads</i> não mapeadas	L11	L14	L15	L16	L19	L20	L23	L24	L26
					número de <i>reads</i>								
controle negativo	81,3	3686	3522	164	13	43	107	0	4	1652	1118	6	0
TF3	77,3	5856	5419	437	499	12	542	7	180	519	2	170	33
TF9	71,9	5352	4622	730	396	2	367	15	209	308	313	89	19
TF14	78,6	5290	5087	203	1184	10	153	19	303	249	38	42	12
TF16	74,8	5174	4337	837	849	6	125	14	278	124	201	172	61
TF17	68,1	4818	3357	1461	226	8	120	8	478	539	36	68	10
TF18	79,7	7978	6399	1579	270	25	337	38	139	1887	104	504	62
TF21	77,4	6800	6533	267	399	95	184	43	1359	2096	55	543	62
TF23	75,6	6550	6270	280	585	207	499	37	425	639	200	290	359
TF25	72	6822	4780	2042	729	174	296	30	417	586	114	213	45
TF26	74,5	8604	5569	3035	637	322	298	79	14	422	237	95	54

Os dados da tabela foram gerados a partir do mapeamento contra a referência utilizando o programa Geneious 11.1.15 (<https://www.geneious.com>) de acordo com os parâmetros padrões.

<b>indivíduo</b>	<b>L29</b>	<b>L30</b>	<b>L32</b>	<b>L35</b>	<b>L36</b>	<b>L37</b>	<b>L38</b>	<b>L39</b>	<b>L49</b>	<b>L53</b>	<b>L57</b>
controle											
negativo	2	1	55	2	0	12	168	2	24	247	66
TF3	54	1287	62	60	10	240	385	554	640	30	133
TF9	17	115	98	392	264	646	457	582	130	27	176
TF14	32	375	104	324	18	674	567	241	625	9	108
TF16	34	123	135	193	166	463	424	242	191	378	158
TF17	21	144	86	127	8	330	267	172	307	212	190
TF18	71	436	16	132	237	788	296	398	2	178	479
TF21	44	246	0	64	12	529	188	176	131	16	291
TF23	42	256	214	90	6	1124	204	397	549	2	145
TF25	26	282	96	58	84	418	390	168	373	63	218
TF26	335	380	28	20	46	885	313	421	457	23	503

Os dados da tabela foram gerados a partir do mapeamento contra a referência utilizando o programa Geneious 11.1.15 (<https://www.geneious.com>) de acordo com os parâmetros padrões.

Na tabela 9 constam os resultados das análises dos *loci* sequenciados após genotipagem completamente manual e com valores de heterozigozidade superestimados. Para o *locus* 52, os indivíduos ou não possuíam *reads*, possuíam poucas ou as *reads* eram curtas e não constavam a região microssatélite e, portanto, não consta nenhuma análise deste *locus* na tabela. Este resultado pode estar relacionado ao fato de que a sequência *Forward* do *primer* L52 já continha parte da região repetitiva. Para L53, as *reads* eram apenas de dímeros de *primers* e, portanto, assim como L52, também foi excluído das análises posteriores. É válido ressaltar que os valores apresentados nesta tabela foram calculados com base em genotipagem realizada de forma manual e considerando apenas o número de repetições dos microssatélites, uma vez que no momento das análises os servidores da UFMG apresentaram indisponibilidade e, portanto, esta análise prévia não contava com os resultados do *pipeline* pretendido e ainda não contávamos com outros programas de análises. Além disto, estes valores também foram calculados com genótipos de um número reduzido de indivíduos (10 indivíduos).

O conteúdo de informação polimórfica (PIC) é um indicador da qualidade do marcador em estudos genéticos, sendo útil para identificação de populações, segregação e controle de paternidade. São considerados muito informativos os marcadores que apresentam valores de PIC superiores a 0.5, medianos aqueles que apresentam valores entre 0.25 e 0.5 e pouco informativos aqueles que apresentam valores inferiores a 0.25 (Botstein 1980). Dentre os 21 *loci* analisados, apenas três apresentaram valor de PIC menor que 0.5. Para o *locus* 30, todos os indivíduos analisados possuíam exatamente oito repetições, e, portanto, esse *locus* não é informativo.

A heterozigosidade observada ( $H_o$ ) representa o número de heterozigotos de um determinado *locus* dividido pelo número total de indivíduos analisados, enquanto que a heterozigosidade esperada ( $H_e$ ) indica a probabilidade de um indivíduo ser heterozigoto em um *locus* de acordo com as suas frequências alélicas. Quando a  $H_o$  é inferior a  $H_e$ , significa que há excesso de homozigotos. Nesta etapa do estudo, cerca de 50% das regiões analisadas mostraram valor de  $H_o$  inferior ao de  $H_e$ .

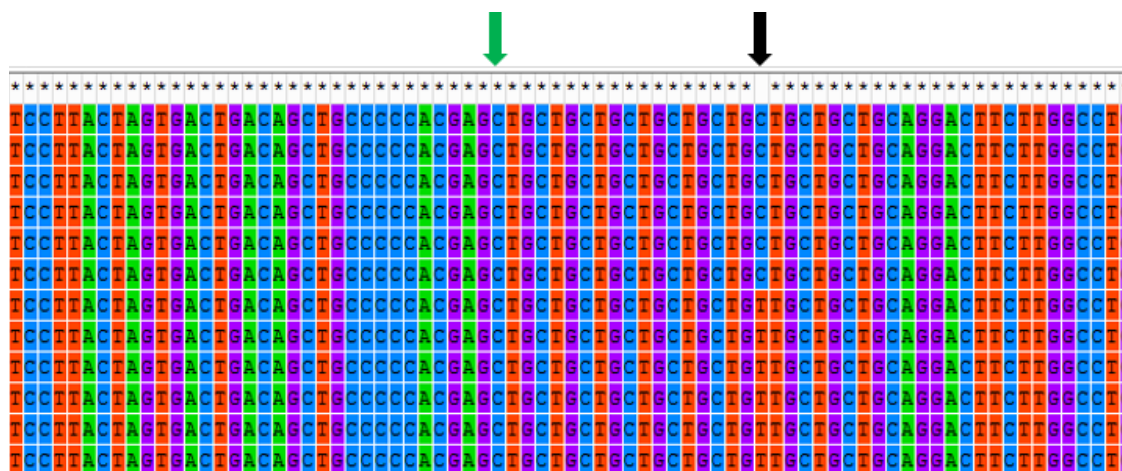
**Tabela 9- Resultado da análise do teste 1 no MiSeq.** Os valores apresentados na tabela foram calculados com base em genótipos identificados manualmente considerando *reads* brutas.

<i>Locus</i>	<b>PIC</b>	<b>Ho</b>	<b>He</b>
L11	0.55	0.889	0.641
L14	0.492	1	0.644
L15	0.671	1	0.758
L16	0.645	1	0.778
L19	0.79	0.889	0.863
L20	0.7	1	0.784
L23	0.786	0.75	0.923
L24	0.786	0.667	0.856
L26	0.51	0.75	0.592
L29	0.825	1	0.964
L30	0	0	0
L32	0.354	0.25	0.425
L35	0.758	0.667	0.837
L36	0.693	0.5	0.783
L37	0.544	0.556	0.614
L38	0.835	0.75	0.908
L39	0.784	1	0.856
L49	0.592	0.625	0.692
L53	0.794	0.2	0.911
L57	0.733	0.889	0.817

PIC = Conteúdo de Informação Polimórfica. Ho= Heterozigosidade observada. He= Heterozigosidade esperada. Análises realizadas com o programa Cervus 3.0.7 (Marshall 1998) a partir de genotipagem manual.

Apesar de trabalhosa, a análise majoritariamente manual das *reads* permitiu fazer algumas observações. Um dos exemplos é a verificação de que cerca de 50% das *reads* do L24 do indivíduo TF14 apresentaram uma substituição de citosina por timina, transformando um microssatélite perfeito em imperfeito (figura 20). Devido à frequência, esta diferença não aparenta ser resultado de erro de sequenciamento. Este exemplo mostra a importância da análise manual e escolha do programa de bioinformática escolhido, pois dependendo do algoritmo utilizado, poderíamos ter como resultado um alelo com nove

repetições e outro com apenas seis repetições ou dois alelos com nove repetições e menção à substituição. A informação desta mutação pode ser importante nos casos de paternidade, uma vez que pode ter sido herdada.



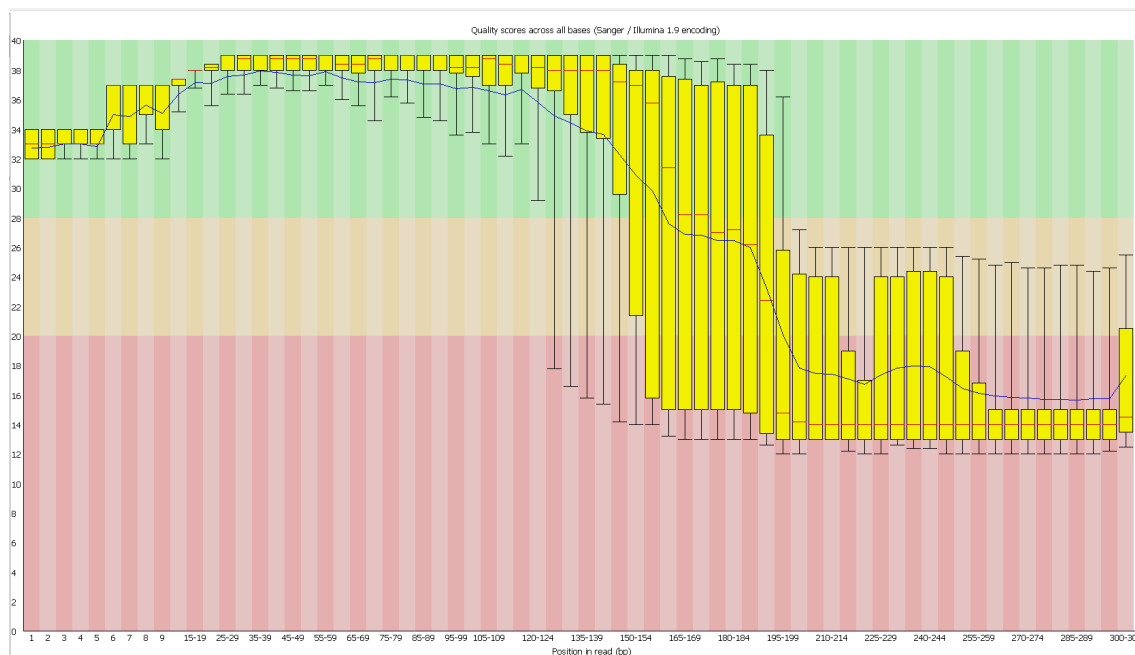
**Figura 20- Exemplo de mutação na região microsatélite.** A imagem mostra algumas *reads* do L24 do indivíduo TF14 alinhadas no programa MEGA 6.0. A seta verde indica o início do microsatélite  $(CTG)_n$  e a seta preta indica a posição na qual foi observada uma substituição de citosina por timina. Esta mutação foi observada apenas neste indivíduo entre os nove analisados.

Como o objetivo deste primeiro teste no MiSeq com poucos indivíduos e sequenciamento de todos os *loci* foi fazer uma breve análise das regiões escolhidas, de acordo com os resultados obtidos neste teste optou-se por prosseguir com os experimentos apenas com 17 dos 21 conjuntos de *primer*, excluindo os conjuntos L30, L32, L52 e L53 das próximas etapas.

#### 1.5.1.4 Teste 2 no MiSeq

Para esta etapa foram utilizados os 17 conjuntos de *primers* com adaptadores selecionados após o teste 1 no MiSeq (L11, L14, L15, L16, L19, L20, L23, L24, L26, L29, L35, L36, L37, L38, L39, L49 e L57) e amostras de 47 espécimes de *S. similis* provenientes do CETAS-BH, totalizando 52 indivíduos sequenciados, além de um controle negativo. Neste sequenciamento não foram incluídas amostras de TF3, TF14, TF17, TF25 e TF26 (sequenciadas no teste 1), e foram repetidos os sequenciamentos de TF9, TF16, TF18, TF21 E TF23 (já previamente sequenciados no teste 1). O teste 2 no MiSeq foi realizado utilizando a estratégia de *single-read* de 300 ciclos, e, conforme mostra a imagem 21, abaixo, a qualidade das bases se manteve muito alta até por volta de 150 pb. A imagem mostra apenas a qualidade das *reads* de TF6 como exemplificação,

uma vez que os demais indivíduos apresentaram gráficos de qualidades similares. A tabela 10, abaixo, traz o número de *reads* geradas para cada indivíduo e cada *loci*. As *reads* geradas para o controle negativo são tratadas no tópico 1.5.1.8.



**Figura 21- Qualidade das bases das *reads* do sequenciamento do teste 2 no MiSeq.** As diferentes amostras apresentaram distribuição de qualidade da base sequenciada similar, sendo a mostrada na imagem referente às *reads* do indivíduo TF6. O eixo Y mostra o valor de qualidade e o eixo X a posição da base na *read*. O *background* do gráfico é dividido em qualidade muito boa (verde), qualidade razoável (laranja) e qualidade ruim (vermelho). A linha vermelha representa a mediana e a linha azul a média da qualidade. A caixa amarela representa o interquartil (25-75%). O gráfico foi gerado pelo programa FastQC 0.11.9 ([www.bioinformatics.babraham.ac.uk/projects/](http://www.bioinformatics.babraham.ac.uk/projects/)).

**Tabela 10- Reads geradas no teste 2 no MiSeq.**

<b>indivíduo</b>	<b>≥Q30 (%)</b>	<b>número total de reads</b>	<b>número de reads mapeadas</b>	<b>número de reads não mapeadas</b>	<b>L11</b>	<b>L14</b>	<b>L15</b>	<b>L16</b>	<b>L19</b>	<b>L20</b>	<b>L23</b>	<b>L24</b>	<b>L26</b>
Controle negativo	80,1	17279	16427	852	489	327	172	7	32	5440	1182	22	13
TF1	89,9	10192	10118	74	1470	82	1321	48	1340	332	20	406	413
TF2	89,4	7922	7821	101	2288	76	405	45	310	142	28	232	434
TF4	90,1	17516	17250	266	2736	46	864	117	2666	326	66	757	1149
TF5	90,2	18451	18165	286	3777	128	2300	336	1463	264	45	210	70
TF6	90,8	19020	18751	269	3337	57	1048	179	1823	310	57	1361	120
TF7	90,6	23215	22682	533	3544	221	1043	358	2151	337	38	1412	1682
TF8	88,4	22570	21894	676	2799	217	1952	83	3082	689	1692	906	1078
TF9	88,3	14542	14036	506	1616	110	1744	236	1255	375	30	433	768
TF10	89,3	21246	21005	241	5334	320	1477	192	2761	1092	129	271	933
TF11	90	14033	13963	70	1920	84	823	148	1764	487	67	375	539
TF12	89,9	14227	14074	153	2258	109	1033	113	997	532	60	532	658
TF13	89,8	17061	16791	270	3017	159	1169	242	2840	556	40	405	438
TF15	90,2	12770	12687	83	2611	135	812	93	2032	398	50	464	407
TF16	89,4	22267	21993	274	3769	199	1371	150	3220	885	60	550	603
TF18	88,5	15608	15433	175	2568	226	1543	175	1955	611	96	804	607
TF19	89,4	26595	26271	324	4348	215	2543	333	3643	653	222	1669	808
TF20	90,1	21031	20815	216	3342	137	1032	84	2283	696	115	247	409
TF21	89,7	20431	20229	202	3110	153	1668	112	4210	544	99	583	638

Número de *reads* geradas por indivíduo no teste 2 do MiSeq, porcentagem de *reads* com qualidade superior a Q30, número de *reads* mapeadas, número de *reads* não mapeadas e número de *reads* por *locus*. O controle negativo passou por todas as reações e etapas das demais amostras, sendo a única diferença a ausência de DNA adicionado na PCR. Os dados da tabela foram gerados a partir do mapeamento contra a referência utilizando o programa Geneious 11.1.15 (<https://www.geneious.com>) de acordo com os parâmetros padrões.



<b>indivíduo</b>	<b>≥Q30 (%)</b>	<b>número total de reads</b>	<b>número de reads mapeadas</b>	<b>número de reads não mapeadas</b>	<b>L11</b>	<b>L14</b>	<b>L15</b>	<b>L16</b>	<b>L19</b>	<b>L20</b>	<b>L23</b>	<b>L24</b>	<b>L26</b>
TF28	90,2	19245	19075	170	1611	111	207	187	1731	727	95	1038	252
TF29	89	22432	22220	212	2939	406	1921	115	3208	382	97	618	687
TF30	89,1	15126	15008	118	68	142	803	81	2818	910	82	854	359
TF31	88,2	18419	18232	187	4580	236	952	33	2869	115	74	898	610
TF32	88,9	15723	15525	198	2740	107	563	69	3313	343	53	979	384
TF33	88,6	18443	18284	159	1889	89	1512	62	2344	289	75	1535	386
TF34	88,4	13977	13820	157	1777	178	1129	68	2925	149	47	929	446
TF35	85,6	19561	18981	580	3268	382	1015	122	1992	214	97	608	403
TF36	88,6	22501	22207	294	2213	195	1351	153	2825	360	113	1374	491
TF37	88,7	26943	26602	341	2524	206	1754	113	4026	986	135	1288	680
TF38	90,8	22289	22170	119	1970	70	1764	166	3921	338	40	1205	402
TF39	89,6	18580	18436	144	2775	82	1026	70	3218	196	58	1848	520
TF40	88,4	5008	4894	114	792	34	355	18	710	119	18	306	128
TF41	90	16498	16315	183	1921	43	1213	119	3259	145	31	1336	375
TF42	90,3	15047	14952	95	2012	48	886	58	2936	84	40	780	483
TF43	89,9	22413	22185	228	2845	93	1368	182	3512	182	60	1584	337
TF44	89	24661	23834	827	3227	126	2326	58	5662	239	69	646	621
TF45	89,5	26176	25889	287	2712	106	1488	124	4000	379	72	2147	550
TF46	90,6	9502	9450	52	832	55	726	52	1403	85	35	583	325
TF47	90,2	7030	6990	40	552	84	593	72	1094	45	21	502	293
TF48	90,9	30905	30635	270	2895	337	2565	269	5637	397	138	2490	2432
TF49	91	35088	34907	181	3539	191	2339	219	6812	177	443	1720	1705
TF50	91,2	22443	22034	409	1866	112	1790	214	2312	185	115	1463	342
TF51	89,1	8773	8636	137	877	58	551	80	831	54	21	553	381
TF52	90,7	21721	21526	195	1652	175	1379	144	3151	159	115	1628	1126

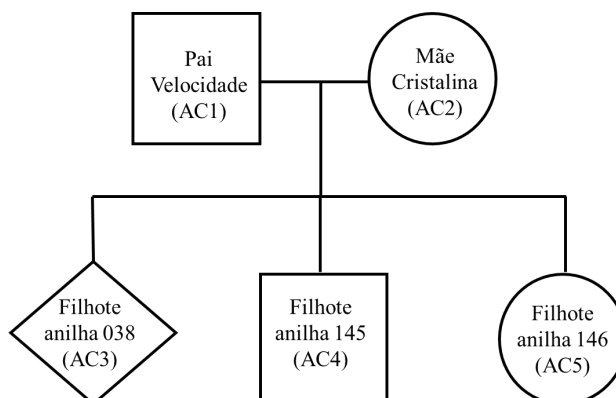
<b>indivíduo</b>	<b>L29</b>	<b>L35</b>	<b>L36</b>	<b>L37</b>	<b>L38</b>	<b>L39</b>	<b>L49</b>	<b>L57</b>
Controle								
negativo	296	485	2924	70	1454	2724	511	279
TF1	745	618	450	996	536	287	652	402
TF2	371	377	70	688	595	1037	581	142
TF4	1462	851	343	1333	723	1866	1520	425
TF5	1319	868	209	1869	1879	1314	1742	372
TF6	1114	1072	41	2041	2308	1253	2188	442
TF7	1549	947	356	2292	1443	2260	2432	617
TF8	1740	705	379	1596	354	2831	1308	483
TF9	1042	667	182	1164	2183	903	1088	240
TF10	309	1779	232	1550	767	1956	1352	551
TF11	493	1048	98	1684	938	1343	1521	631
TF12	698	1474	205	1279	567	1811	1174	574
TF13	462	1594	33	1653	1558	1030	985	610
TF15	484	816	81	1315	730	427	1368	464
TF16	1313	1014	817	1474	2815	1737	1040	976
TF18	833	1020	106	1955	937	807	117	1073
TF19	1239	1313	1006	1800	3070	1172	1069	1168
TF20	1028	1218	421	2355	1615	2462	2417	954
TF21	965	1260	99	1483	1237	1294	1660	1114
TF22	898	1714	251	1852	2711	2542	1401	720
TF23	309	1886	499	1520	827	2782	2064	686
TF24	818	1132	641	2284	1787	1429	1933	958
TF27	116	66	197	2752	1186	2704	76	160
TF28	1078	1307	632	2357	2269	2945	1435	1093
TF29	1008	1048	387	1814	2308	2060	1793	1429

<b>indivíduo</b>	<b>L29</b>	<b>L35</b>	<b>L36</b>	<b>L37</b>	<b>L38</b>	<b>L39</b>	<b>L49</b>	<b>L57</b>
TF30	636	1038	244	2115	1363	1642	1385	468
TF31	1088	875	224	1236	697	2442	753	550
TF32	640	895	122	1385	1853	423	1091	565
TF33	685	1634	53	899	1935	2265	1905	727
TF34	541	912	196	1194	1406	440	1158	325
TF35	792	1416	318	2675	1482	2357	1161	679
TF36	813	1326	377	2886	1704	2738	2192	1096
TF37	929	1835	393	2431	3409	1934	2518	1441
TF38	782	1329	128	2460	2093	2056	2680	766
TF39	279	1082	161	1993	2051	1714	805	558
TF40	213	263	26	549	501	486	257	119
TF41	657	1316	1060	2048	922	376	1338	156
TF42	743	1014	595	653	2577	829	667	547
TF43	930	1694	96	1719	4088	859	1697	939
TF44	970	1325	45	1696	2507	2015	1775	527
TF45	971	1451	179	1830	3793	2296	2474	1317
TF46	413	598	297	656	1287	974	893	236
TF47	467	473	339	916	668	673	21	177
TF48	1854	2082	832	3228	1200	965	2738	576
TF49	2048	2928	124	4216	2354	1330	3137	1628
TF50	1607	1433	46	2503	2910	2509	1736	891
TF51	661	528	67	828	1127	1000	665	354
TF52	1240	1580	249	2003	2059	1811	2450	605

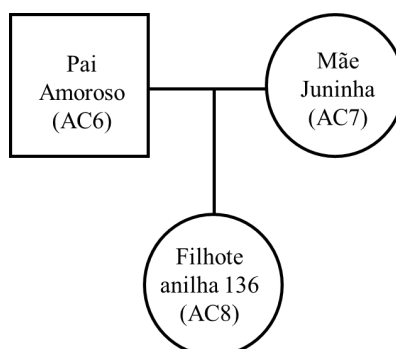
### 1.5.1.5 Teste 3 no MiSeq - amostras de criatório

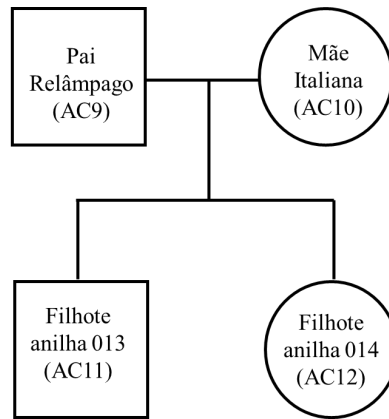
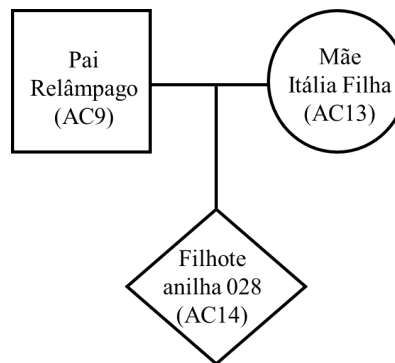
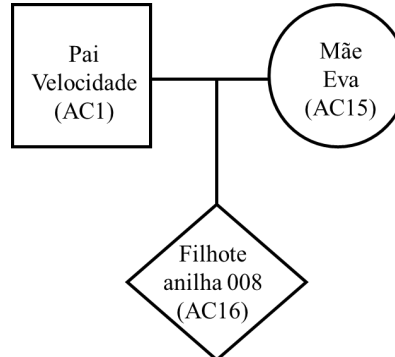
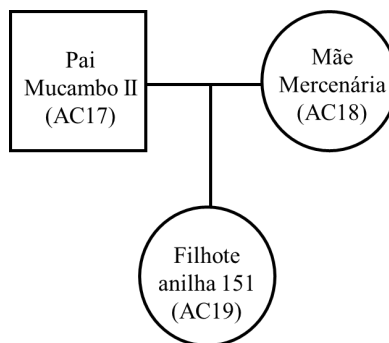
Amostras de 21 indivíduos com relação de parentesco conhecida foram cedidas para este trabalho por um criatório do estado de Minas Gerais. Diferentemente das amostras de músculo utilizadas nos experimentos anteriores, neste caso as amostras se tratavam de gota de sangue em papel FTA e formavam 6 famílias completas (filhote(s), pai e mãe) e uma família composta apenas por dois irmãos. Conforme demonstrado na tabela abaixo com os heredogramas das famílias (figura 22), os animais foram identificados como “AC” seguido do número. Os filhotes que possuíam a informação sobre o nascimento eram datados entre dezembro de 2011 e fevereiro de 2019. Foi realizado contato com diversos criatórios brasileiros de trinca-ferro na tentativa de se obter mais amostras com informações de parentesco por um período de três anos, porém não obtivemos sucesso.

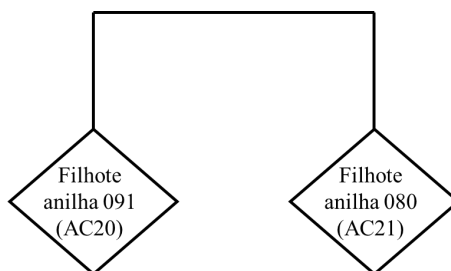
#### Família 1



#### Família 2



**Família 3****Família 4****Família 5****Família 6**

**Família 7**

**Figura 22- Heredogramas com as amostras de famílias de *S. similis* oriundas de criatório.** Os heredogramas presentes na figura estão separados por família, sendo seis famílias completas (composta por filhote(s), pai e mãe) e uma família composta apenas por dois irmãos. Nos heredogramas, o quadrado representa indivíduo do sexo masculino, o círculo representa indivíduo do sexo feminino, o losango representa indivíduo com sexo não determinado e as linhas representam a relação de parentesco. As informações de parentesco, nome e número de anilha foram fornecidas pelo criatório. Entre parênteses está o código com o qual a amostra foi identificada para o sequenciamento e análises, sendo um total de 21 amostras provenientes de criatório.

Diferente das outras amostras utilizadas neste estudo nos demais sequenciamentos (amostras TF), a disponibilidade das amostras oriundas de criatório era bastante restrita quanto à quantidade. Anteriormente à extração de DNA destas amostras, foram realizados testes de extração de DNA de sangue em papel FTA com diversos protocolos e adaptações, incluindo kit GTS (Phoneutria), protocolo de fenol-clorofórmio e o protocolo com precipitação com isopropanol. Estes dois últimos apresentaram melhores resultados quanto à quantidade de DNA extraído, e optou-se por realizar o último por já ser o protocolo de praxe do laboratório de DNA da Polícia Civil de Minas Gerais. Ao final do protocolo de extração, as amostras foram eluídas em 40 µl de tampão TE, e apresentaram as concentrações de DNA apresentadas na tabela 11, abaixo. Para as amostras cuja quantificação foi superior a 2,5 ng/µl, foi possível utilizar 5ng de DNA para cada PCR, conforme padronizado. Já para as amostras cuja concentração foi inferior a 2,5 ng/µl, foi utilizado 2 µl de amostra para cada PCR e, mesmo com concentração inferior, foi possível sequenciar todas as amostras. Mesmo com quantidades de DNA molde inferiores ao padronizado, ocorreu amplificação de todos os *loci* de todas as amostras, que foi observado com aplicação de 9µl dos produtos das PCR em gel de agarose 1,5%.

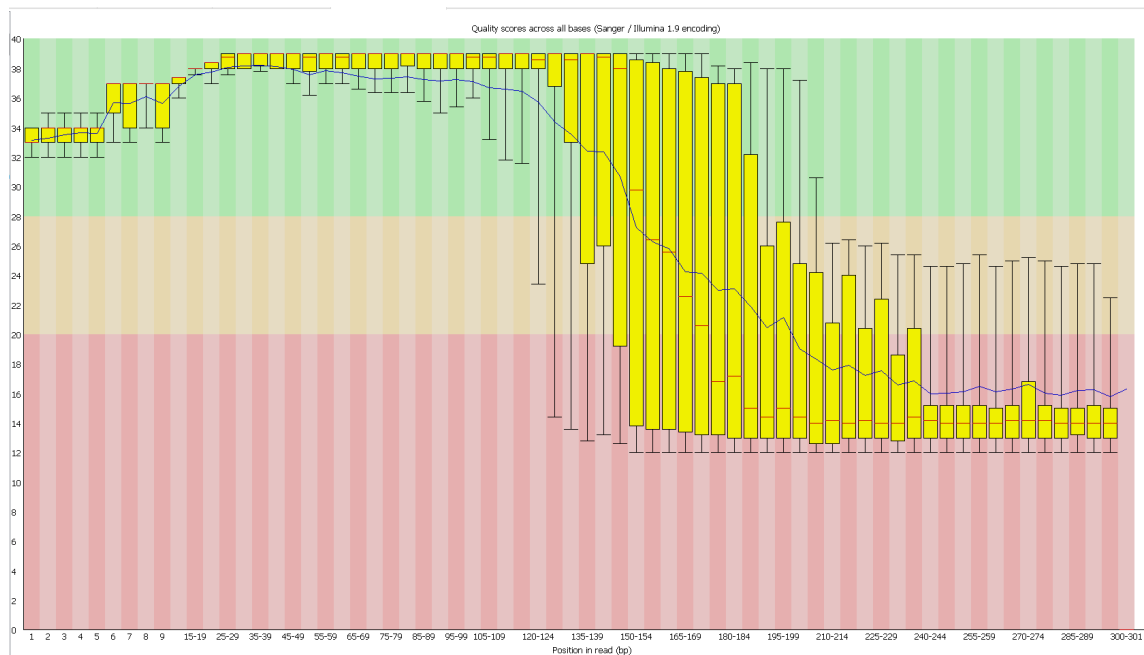
**Tabela 11- Quantidade de DNA extraído das amostras de criatório e quantidade utilizada na PCR.**

Indivíduo	Concentração de DNA (ng/ $\mu$ l)	Quantidade de DNA utilizado na PCR (ng)
AC1	3,12	5
AC2	0,76	1,52
AC3	2,62	5
AC4	2,14	4,28
AC5	6,5	5
AC6	1,13	2,26
AC7	12,9	5
AC8	3,62	5
AC9	3,82	5
AC10	5,14	5
AC11	5,20	5
AC12	2,34	4,68
AC13	1,68	3,36
AC14	1,21	2,42
AC15	5,0	5
AC16	2,24	4,48
AC17	0,85	1,70
AC18	4,74	5
AC19	2,10	4,20
AC20	1,11	2,22
AC21	0,85	1,70

Ao final do processo de extração de DNA as amostras foram eluídas em 40  $\mu$ l de tampão TE e suas concentrações são as mostradas na tabela. Para as amostras cuja concentração de DNA foi inferior 2,5 ng/ $\mu$ l, não foi possível utilizar os 5ng conforme padronizado para as reações e utilizou-se as quantidades conforme mostrado na última coluna da tabela.

O sequenciamento das amostras de criatório foi realizado utilizando a estratégia de *single-read* de 300 ciclos, e a qualidade das bases se manteve de alta qualidade até cerca de 135pb conforme mostrado na figura 23. A qualidade das bases ao longo do sequenciamento de todos os indivíduos de criatório foi similar, e, portanto, a imagem se refere apenas ao AC8 para exemplificação. Para as amostras de criatório foram sequenciados, assim como no teste 2 no MiSeq, os 17 *loci*: L11, L14, L15, L16, L19, L20, L23, L24, L26, L29, L35, L36, L37, L38, L39, L49 e L57. A tabela 12 traz o número de *reads* brutas geradas por indivíduo e por *locus*, e nota-se que alguns possuem poucas *reads*, comprometendo a genotipagem. O número menor de *reads* geradas neste sequenciamento quando comparado aos demais não é devido à menor quantidade de DNA nas PCRs, mas sim à proporção com que estas amostras foram inseridas no cartucho para o sequenciamento. Como as amostras de criatório eram diferentes das amostras

previamente testadas, neste teste 3 do MiSeq foi incluso também um controle positivo, cujo DNA utilizado foi extraído de uma amostra de músculo de *S. similis*. O controle negativo passou por todas as etapas de sequenciamento e possuía todos os reagentes, com exceção do DNA. As *reads* que constam na tabela como controle negativo não é resultado de contaminação e é melhor abordado no item 1.5.1.8.



**Figura 23- Qualidade das bases das *reads* do sequenciamento do teste com amostras de criatórios.** As diferentes amostras apresentaram distribuição de qualidade da base sequenciada similar, sendo a mostrada na imagem referente às *reads* do indivíduo AC8 para exemplificação. O eixo Y mostra o valor de qualidade e o eixo X a posição da base na *read*. O *background* do gráfico é dividido em qualidade muito boa (verde), qualidade razoável (laranja) e qualidade ruim (vermelho). A linha vermelha representa a mediana e a linha azul a média da qualidade. A caixa amarela representa o inter-quartil (25-75%). O gráfico foi gerado pelo programa FastQC 0.11.9 ([www.bioinformatics.babraham.ac.uk/projects/](http://www.bioinformatics.babraham.ac.uk/projects/)).



**Tabela 12- Reads geradas no teste 3 com as amostras de criatório.**

indivíduo	≥Q30(%)	número total de reads	Número de reads mapeadas	Número de reads não mapeadas	L11	L14	L15	L16	L19	L20	L23	L24
controle negativo	80,2	973	332	660	2	1	13	1	2	1	1	16
controle positivo	90,3	2129	1811	318	275	14	105	24	112	9	3	251
AC1	82,6	1813	1038	794	98	4	17	5	106	1	1	150
AC2	86	2132	1767	365	92	1	112	47	140	0	4	126
AC3	88,4	1920	1508	412	145	7	71	54	101	5	4	134
AC4	88,3	1702	1304	398	193	1	50	17	82	0	4	145
AC5	91,3	2496	2141	355	238	14	67	52	174	0	6	188
AC6	87,3	2097	1563	534	133	6	118	42	72	3	0	63
AC7	90,6	5017	3850	1167	635	18	137	63	295	2	16	386
AC8	89,3	1833	1437	415	226	26	64	33	27	30	7	134
AC9	89,6	1743	1462	300	217	6	46	25	119	5	21	151
AC10	88,8	1797	1358	458	241	8	71	36	78	1	6	155
AC11	91,1	1315	1107	227	211	8	40	20	87	1	7	107
AC12	50,2	1122	957	184	172	3	33	18	73	2	4	49
AC13	89,9	1176	982	213	134	14	41	10	78	8	8	18
AC14	90,2	1857	1578	298	166	10	106	27	262	1	6	123
AC15	90,5	1904	1593	311	264	20	119	16	90	2	18	140
AC16	90,2	898	720	197	69	4	37	6	92	1	5	71

Número de *reads* geradas por indivíduo no teste com as amostras de criatório (AC), porcentagem de *reads* com qualidade superior a Q30, número de *reads* mapeadas, número de *reads* não mapeadas e número de *reads* por *locus*. O controle negativo passou por todas as reações e etapas das demais amostras, sendo a única diferença a ausência de DNA adicionado na PCR. A amostra utilizada no controle positivo era proveniente de tecido muscular, assim como àquelas utilizadas nos testes 1 e 2. Os dados da tabela são referentes às *reads* brutas a partir do mapeamento contra a referência utilizando o programa Geneious 11.1.15 (<https://www.geneious.com>) de utilizando os parâmetros padrões.

<b>indivíduo</b>	<b>≥Q30(%)</b>	<b>número total de reads</b>	<b>Número de reads mapeadas</b>	<b>Número de reads não mapeadas</b>	<b>L11</b>	<b>L14</b>	<b>L15</b>	<b>L16</b>	<b>L19</b>	<b>L20</b>	<b>L23</b>	<b>L24</b>
AC17	84,6	1213	783	449	95	6	50	13	51	1	4	49
AC18	85,4	1631	1223	427	101	18	51	32	143	1	1	68
AC19	89	1393	1107	305	180	10	96	20	140	1	8	107
AC20	89,4	2083	1284	799	168	17	148	23	180	1	31	219
AC21	86,4	2067	1495	572	215	11	133	28	204	0	0	134

<b>indivíduo</b>	<b>L26</b>	<b>L29</b>	<b>L35</b>	<b>L36</b>	<b>L37</b>	<b>L38</b>	<b>L39</b>	<b>L49</b>	<b>L52</b>	<b>L53</b>	<b>L57</b>
controle negativo	2	4	1	3	114	1	2	14	18	39	97
controle positivo	111	46	43	0	78	407	80	124	23	67	39
AC1	21	31	21	11	153	98	38	67	22	25	169
AC2	69	32	251	35	168	113	118	99	28	119	213
AC3	66	51	61	76	119	192	79	108	46	64	125
AC4	75	63	13	5	140	159	26	88	23	102	118
AC5	109	22	84	119	329	175	140	140	57	88	139
AC6	86	24	11	8	219	173	40	336	48	122	59
AC7	100	52	206	36	575	418	175	291	109	138	198
AC8	76	49	2	36	197	103	134	101	28	37	127
AC9	35	25	143	20	98	156	118	110	51	45	71
AC10	50	34	117	11	95	166	61	79	27	28	94
AC11	43	10	81	2	150	148	56	40	25	26	45
AC12	61	41	1	3	120	156	55	83	35	15	33
AC13	71	75	2	5	61	149	64	96	54	21	73
AC14	47	50	43	5	208	201	137	102	17	22	45
AC15	94	11	3	34	358	54	151	110	18	48	43
AC16	29	20	13	13	87	80	50	65	13	36	29

<b>individuo</b>	<b>L26</b>	<b>L29</b>	<b>L35</b>	<b>L36</b>	<b>L37</b>	<b>L38</b>	<b>L39</b>	<b>L49</b>	<b>L52</b>	<b>L53</b>	<b>L57</b>
AC17	32	20	10	3	77	78	2	66	39	95	92
AC18	29	35	44	41	142	135	48	84	37	95	118
AC19	46	25	60	7	137	89	16	89	28	19	29
AC20	44	59	52	3	116	29	9	24	35	29	97
AC21	21	53	41	15	132	134	34	35	48	66	191

#### 1.5.1.6 Análises dos loci

A análise de genótipos de regiões microssatélites sequenciados com a técnica de NGS ainda vem sendo aperfeiçoada, e neste trabalho cogitou-se utilizar diversos programas para a análise das amostras sequenciadas, dentre eles: *pipeline* de Pimentel (2018 e 2020), *Repeat Finder* e *Phobos* (plugins do programa Geneious), *MISA-web* (Beier 2017), *MEGASAT* (Zhan 2016), *Tandem Repeats Finder* (Benson 1999), *ReviSTER* (Tae 2013), *STRetch* (Dashnow 2018) e *CHIIMP* (Barbian 2018). Por fim, as análises dos genótipos foram realizadas utilizando o *pipeline* de Pimentel (2018 e 2020) mencionado na metodologia e o programa *CHIIMP*, sendo os resultados conferidos também com o auxílio dos programas *Geneious* e *MEGA 6* (Peterson 2013). As duas formas de análise tiveram pontos positivos e negativos. A dificuldade em se utilizar o *pipeline* de Pimentel (2018 e 2020) se dá pela necessidade de utilizar *clusters* de maior capacidade de processamento e não somente um computador comum. Além disto, várias etapas em vários programas são necessários para se chegar ao resultado final e exigem maior conhecimento na área de bioinformática, além de ser preciso realizar ajustes no arquivo referência para cada *locus*. Porém, este *pipeline* também viabiliza mais alterações nos critérios de análise e o resultado é uma tabela já no modelo utilizado em boa parte dos programas de genética de populações. Já o *CHIIMP* possui a facilidade de ser apenas um programa necessário e poder ser processado em computador comum, além de considerar também variações de nucleotídeo único (SNV, do inglês “*Single Nucleotide Variation*”) na identificação dos alelos e não somente o tamanho destes, e também gerar, dentre os arquivos output, representações esquemáticas dos alelos alinhadas para cada *locus*, que facilita a observação. Porém, o *CHIIMP* é um pouco mais limitado na quantidade de parâmetros que podem ser alterados e o resultado final não está pronto para ser utilizado em programas de genéticas de populações.

Para ambas as formas de análise, diversos parâmetros foram testados e optou-se por utilizar o mínimo de sete *reads* para cada *locus* para a análise e mínimo de 20% de frequência do segundo alelo para ser considerado heterozigoto. Os resultados foram comparados e foi observado compatibilidade entre os resultados do *pipeline* de Pimentel (2018 e 2020) e do *CHIIMP*. A porcentagem mínima do segundo alelo para ser considerado heterozigoto e o número mínimo de *reads* para genotipagem ainda são valores bastante discutidos nos artigos científicos e podem variar conforme o tipo de sequência alvo, sendo encontrados, portanto, diversos parâmetros em artigos já publicados. Elwick (2018), por exemplo, utiliza em suas análises o mínimo de duas *reads*,

Zavodna (2014) o mínimo de cinco *reads*, Van der Gaag (2016) o mínimo de oito *reads* e Pimentel (20018) e Tibihika (2019) utilizam o mínimo de dez *reads*. Quanto à frequência do segundo alelo, os valores também são bastantes variáveis, como, por exemplo, análises que utilizam 8% de frequência para determinados alelos (Van der Gaag 2016), 10% (Zavodna 2014), 20% (Pimentel 2018), 30% (Hoogenboom 2017) e 40% (Gettings 2016). Os dados deste trabalho, quando analisados com frequências iguais ou inferiores a 15%, ocorria um aumento do número de heterozigotos que não correspondia às observações feitas, enquanto que com valores de 20 e 25% os resultados eram mais compatíveis. Alguns indivíduos, para alguns *loci*, não foi possível determinar o genótipo mesmo havendo *reads* para determinado indivíduo e *loci*. Isso ocorre porque as *reads* brutas geradas nos sequenciamentos não são todas utilizadas para a genotipagem, pois podem possuir baixa qualidade, serem dímeros de *primers*, não conter a sequência completa do *amplicon* ou não possuir o número mínimo de *reads* estabelecido para a determinação do genótipo.

As figuras 24 à 42, abaixo, foram geradas pelo CHIIMP e trazem o alinhamento dos diferentes alelos encontrados para cada *locus* para todos os indivíduos sequenciados neste trabalho, sendo possível observar que diferentes alelos possuem o mesmo número de repetições microssatélites mas possuem SNV em alguma parte da sequência. Com exceção de L30 e L32, que foram sequenciados apenas no teste 1, os demais *loci* possuem de 3 a 53 alelos por *locus* quando considerados o tamanho de fragmento e SNV, sendo a média de alelo por *locus* de 19,3. Se os alelos encontrados levassem em consideração apenas o tamanho do fragmento e não considerassem suas sequências, o número de alelos por *locus* seria reduzido, variando de 3 a 20, e a média de alelo por *locus* seria de 8,5. Apesar de possuírem um poder discriminatório por *locus* menor que os STR, adicionar SNV às análises melhora o índice de paternidade do *loci* e são bastante úteis para distinguir indivíduos com grau de parentesco em alguns casos (Phillips 2008) e incompatibilidades (Ye 2013). Estes dados, assim como a variação no tamanho dos alelos, são mostrados na tabela 13, e exemplificam como a utilização das sequências, e não somente o uso do tamanho do fragmento, são úteis na identificação dos alelos.

Tabela 13-Número de alelos por *locus* e seus tamanhos.

<i>Locus</i>	Número de alelos considerando tamanho do fragmento e SNV	Varição no tamanho dos alelos (pb)	Número de alelos considerando apenas o tamanho do fragmento
L11	8	109 a 175	7
L14	11	110 a 131	7
L15	18	96 a 112	9
L16	14	81 a 101	8
L19	12	115 a 129	7
L20	11	109 a 113	3
L23	29	118 a 160	14
L24	20	101 a 119	7
L26	3	143 a 152	2
L29	6	145 a 166	4
L30*	1	143	1
L32*	8	158 a 161	2
L35	14	100 a 136	10
L36	45	112 a 184	20
L37	21	116 a 134	7
L38	53	104 a 149	14
L39	19	97 a 136	10
L49	6	105 a 113	3
L57	38	120 a 180	12
Média de alelos por <i>locus</i> *	19,3	-	8,5

Os dados compilados na tabela foram gerados pelo programa CHIIMP (Barbian 2018). \*Os *loci* L30 e L32 participaram apenas do teste 1 no MiSeq. Os cálculos de média de alelos por *locus* não contabilizam os dados de L30 e L32.

Figura a disposição sob consulta com o Prof. Dr. Evanguedes Kalapothakis.

**Figura 24- Alelos encontrados para o *locus* L11.** Foram encontrados oito alelos para L11 em todas as amostras sequenciadas e genotipadas de *S. similis*. No alinhamento mostrado na figura é possível observar a variação no número de repetições de microssatélites e também SNV (*Single Nucleotide Variation*) nas sequências. No eixo X há uma referência de tamanho de pares de bases e à direita há a informação do tamanho do alelo, além do número de vezes que o respectivo alelo apareceu na tabela de genotipagem. A imagem foi gerada pelo programa CHIIMP (Barbian 2018).

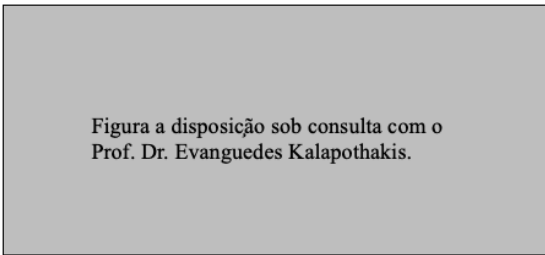


Figura a disposição sob consulta com o Prof. Dr. Evanguedes Kalapothakis.

**Figura 25- Alelos encontrados para o locus L14.** Foram encontrados 11 alelos para L14 em todas as amostras sequenciadas e genotipadas de *S. similis*. No alinhamento mostrado na figura é possível observar a variação no número de repetições de microssatélites e também SNV nas sequências. Nota-se que alguns alelos (de tamanhos 100 a 106 pb) não possuem uma sequência entre as repetições do microssatélite (em torno de 40pb). No eixo X há uma referência de tamanho de pares de bases e à direita há a informação do tamanho do alelo, além do número de vezes que o respectivo alelo apareceu na tabela de genotipagem. A imagem foi gerada pelo programa CHIIMP (Barbian 2018).

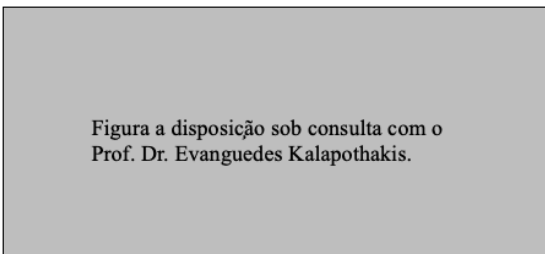


Figura a disposição sob consulta com o Prof. Dr. Evanguedes Kalapothakis.

**Figura 26- Alelos encontrados para o locus L15.** Foram encontrados 18 alelos para L15 em todas as amostras sequenciadas e genotipadas de *S. similis*. No alinhamento mostrado na figura é possível observar a variação no número de repetições de microssatélites e também SNV nas sequências. No eixo X há uma referência de tamanho de pares de bases e à direita há a informação do tamanho do alelo, além do número de vezes que o respectivo alelo apareceu na tabela de genotipagem. A imagem foi gerada pelo programa CHIIMP (Barbian 2018).

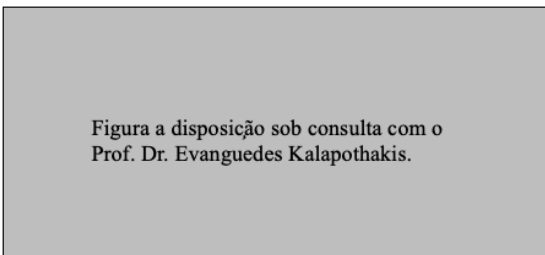


Figura a disposição sob consulta com o Prof. Dr. Evanguedes Kalapothakis.

**Figura 27- Alelos encontrados para o locus L16.** Foram encontrados 14 alelos para L16 em todas as amostras sequenciadas e genotipadas de *S. similis*. No alinhamento mostrado na figura é possível observar a variação no número de repetições de microssatélites e também SNV nas sequências. No eixo X há uma referência de tamanho de pares de bases e à direita há a informação do tamanho do alelo, além do número de vezes que o respectivo alelo apareceu na tabela de genotipagem. A imagem foi gerada pelo programa CHIIMP (Barbian 2018).

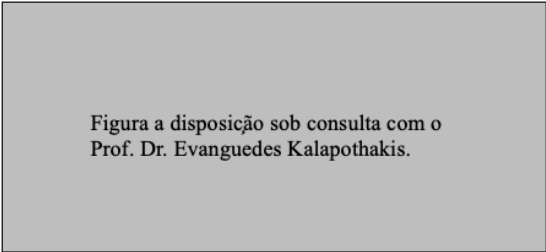


Figura a disposição sob consulta com o Prof. Dr. Evanguedes Kalapothakis.

**Figura 28- Alelos encontrados para o locus L19.** Foram encontrados 12 alelos para L19 em todas as amostras sequenciadas e genotipadas de *S. similis*. No alinhamento mostrado na figura é possível observar a variação no número de repetições de microssatélites e também SNV nas sequências. No eixo X há uma referência de tamanho de pares de bases e à direita há a informação do tamanho do alelo, além do número de vezes que o respectivo alelo apareceu na tabela de genotipagem. A imagem foi gerada pelo programa CHIIMP (Barbian 2018).

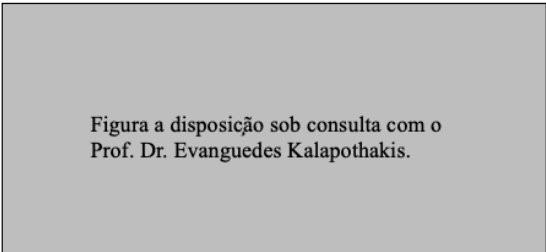


Figura a disposição sob consulta com o Prof. Dr. Evanguedes Kalapothakis.

**Figura 29- Alelos encontrados para o locus L20.** Foram encontrados 11 alelos para L20 em todas as amostras sequenciadas e genotipadas de *S. similis*. No alinhamento mostrado na figura é possível observar a variação no número de repetições dos dois *motifs* de microssatélites e também SNV nas sequências. No eixo X há uma referência de tamanho de pares de bases e à direita há a informação do tamanho do alelo, além do número de vezes que o respectivo alelo apareceu na tabela de genotipagem. A imagem foi gerada pelo programa CHIIMP (Barbian 2018).

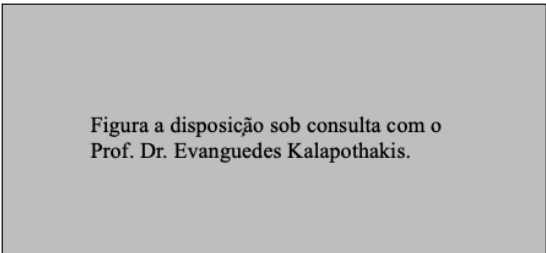


Figura a disposição sob consulta com o Prof. Dr. Evanguedes Kalapothakis.

**Figura 30- Alelos encontrados para o locus L23.** Foram encontrados 29 alelos para L23 em todas as amostras sequenciadas e genotipadas de *S. similis*. No alinhamento mostrado na figura é possível observar a variação no número de repetições de microssatélites e também SNV nas sequências. No eixo X há uma referência de tamanho de pares de bases e à direita há a informação do tamanho do alelo, além do número de vezes que o respectivo alelo apareceu na tabela de genotipagem. A imagem foi gerada pelo programa CHIIMP (Barbian 2018).



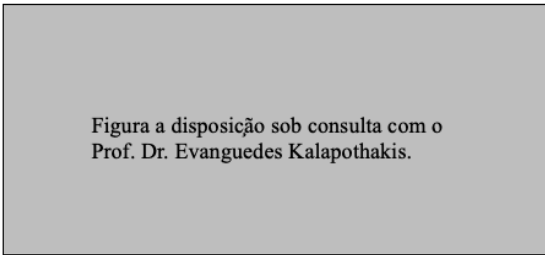


Figura a disposição sob consulta com o Prof. Dr. Evanguedes Kalapothakis.

**Figura 31- Alelos encontrados para o locus L24.** Foram encontrados 20 alelos para L24 em todas as amostras sequenciadas e genotipadas de *S. similis*. No alinhamento mostrado na figura é possível observar a variação no número de repetições de microssatélites e também SNV nas sequências. No eixo X há uma referência de tamanho de pares de bases e à direita há a informação do tamanho do alelo, além do número de vezes que o respectivo alelo apareceu na tabela de genotipagem. A imagem foi gerada pelo programa CHIIMP (Barbian 2018).

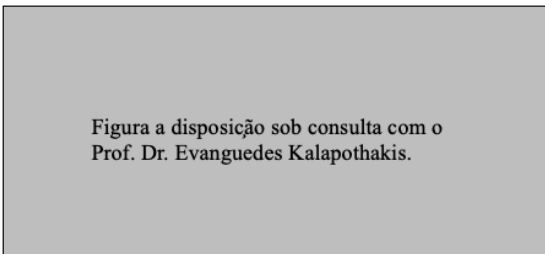


Figura a disposição sob consulta com o Prof. Dr. Evanguedes Kalapothakis.

**Figura 32-Alelos encontrados para o locus L26.** Foram encontrados três alelos para L26 em todas as amostras sequenciadas e genotipadas de *S. similis*. No alinhamento mostrado na figura é possível observar a variação no número de repetições de microssatélites e também SNV nas sequências. No eixo X há uma referência de tamanho de pares de bases e à direita há a informação do tamanho do alelo, além do número de vezes que o respectivo alelo apareceu na tabela de genotipagem. A imagem foi gerada pelo programa CHIIMP (Barbian 2018).

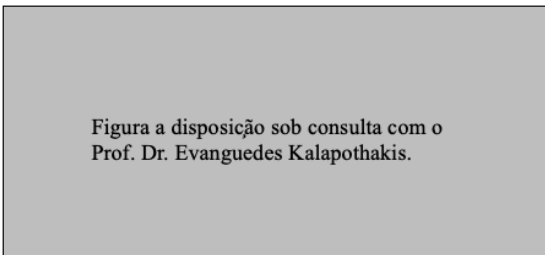


Figura a disposição sob consulta com o Prof. Dr. Evanguedes Kalapothakis.

**Figura 33- Alelos encontrados para o locus L29.** Foram encontrados seis alelos para L29 em todas as amostras sequenciadas e genotipadas de *S. similis*. Poucas amostras tiveram seu genótipo estabelecido para este locus. No alinhamento mostrado na figura é possível observar a variação no número de repetições de microssatélites e também SNV nas sequências. No eixo X há uma referência de tamanho de pares de bases e à direita há a informação do tamanho do alelo, além do número de vezes que o respectivo alelo apareceu na tabela de genotipagem. A imagem foi gerada pelo programa CHIIMP (Barbian 2018).

Figura a disposição sob consulta com o Prof. Dr. Evanguedes Kalapothakis.

**Figura 34- Alelos encontrados para o locus L30.** Foi encontrado apenas 1 alelo para L30 em todas as amostras sequenciadas e genotipadas de *S. similis* (locus sequenciado apenas no teste 1 do MiSeq), não havendo variação no número de repetições do microssatélites e nem ocorrência de SNV. Devido à ausência de polimorfismo detectada logo no primeiro sequenciamento, este locus não foi utilizados nos testes 2 e 3. A imagem foi gerada pelo programa CHIIMP (Barbian 2018).

Figura a disposição sob consulta com o Prof. Dr. Evanguedes Kalapothakis.

**Figura 35- Alelos encontrados para o locus L32.** Considerando SNV e variação no número de repetições de microssatélites, foram encontrados oito alelos para L32 em todas as amostras sequenciadas e genotipadas de *S. similis* (locus sequenciado apenas no teste 1 do MiSeq). Porém, considerando apenas o número de repetições de microssatélites, há apenas dois alelos. Devido ao baixo polimorfismo detectado logo no primeiro sequenciamento, este locus não foi utilizados nos testes 2 e 3. No eixo X há uma referência de tamanho de pares de bases e à direita há a informação do tamanho do alelo, além do número de vezes que o respectivo alelo apareceu na tabela de genotipagem. A imagem foi gerada pelo programa CHIIMP (Barbian 2018).

Figura a disposição sob consulta com o Prof. Dr. Evanguedes Kalapothakis.

**Figura 36- Alelos encontrados para o locus L35.** Foram encontrados 14 alelos para L35 em todas as amostras sequenciadas e genotipadas de *S. similis*. No alinhamento mostrado na figura é possível observar a variação no número de repetições de microssatélites e também SNV nas seqüências. No eixo X há uma referência de tamanho de pares de bases e à direita há a informação do tamanho do alelo, além do número de vezes que o respectivo alelo apareceu na tabela de genotipagem. A imagem foi gerada pelo programa CHIIMP (Barbian 2018).

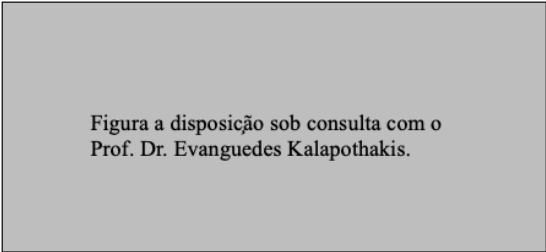


Figura a disposição sob consulta com o Prof. Dr. Evanguedes Kalapothakis.

**Figura 37- Alelos encontrados para o locus L36.** Foram encontrados 45 alelos para L36 em todas as amostras sequenciadas e genotipadas de *S. similis*. No alinhamento mostrado na figura é possível observar a variação no número de repetições de microssatélites de dois *motifs* diferentes e também SNV nas sequências. No eixo X há uma referência de tamanho de pares de bases e à direita há a informação do tamanho do alelo, além do número de vezes que o respectivo alelo apareceu na tabela de genotipagem. A imagem foi gerada pelo programa CHIIMP (Barbian 2018).

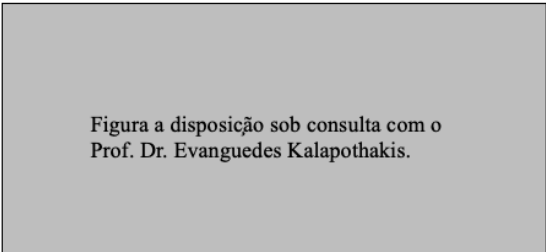


Figura a disposição sob consulta com o Prof. Dr. Evanguedes Kalapothakis.

**Figura 38- Alelos encontrados para o locus L37.** Foram encontrados 21 alelos para L37 em todas as amostras sequenciadas e genotipadas de *S. similis*. No alinhamento mostrado na figura é possível observar a variação no número de repetições de microssatélites e também SNV nas sequências. No eixo X há uma referência de tamanho de pares de bases e à direita há a informação do tamanho do alelo, além do número de vezes que o respectivo alelo apareceu na tabela de genotipagem. A imagem foi gerada pelo programa CHIIMP (Barbian 2018).

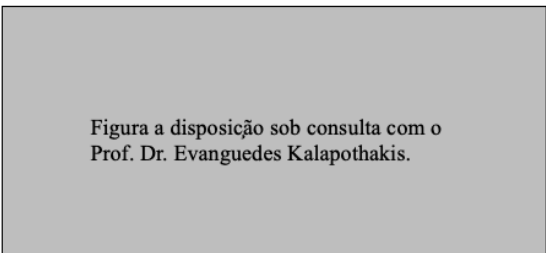


Figura a disposição sob consulta com o Prof. Dr. Evanguedes Kalapothakis.

**Figura 39- Alelos encontrados para o locus L38.** Foram encontrados 53 alelos para L38 em todas as amostras sequenciadas e genotipadas de *S. similis*. No alinhamento mostrado na figura é possível observar a variação no número de repetições de microssatélites e também SNV nas sequências. No eixo X há uma referência de tamanho de pares de bases e à direita há a informação do tamanho do alelo, além do número de vezes que o respectivo alelo apareceu na tabela de genotipagem. A imagem foi gerada pelo programa CHIIMP (Barbian 2018).

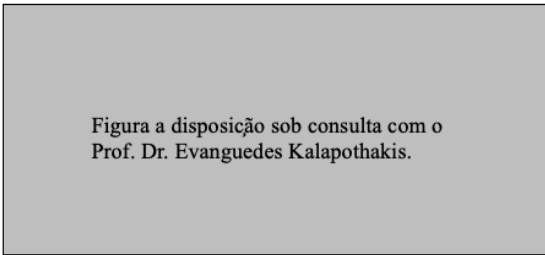


Figura a disposição sob consulta com o Prof. Dr. Evanguedes Kalapothakis.

**Figura 40- Alelos encontrados para o locus L39.** Foram encontrados 19 alelos para L39 em todas as amostras sequenciadas e genotipadas de *S. similis*. No alinhamento mostrado na figura é possível observar a variação no número de repetições de microssatélites e também SNV nas sequências. No eixo X há uma referência de tamanho de pares de bases e à direita há a informação do tamanho do alelo, além do número de vezes que o respectivo alelo apareceu na tabela de genotipagem. A imagem foi gerada pelo programa CHIIMP (Barbian 2018).

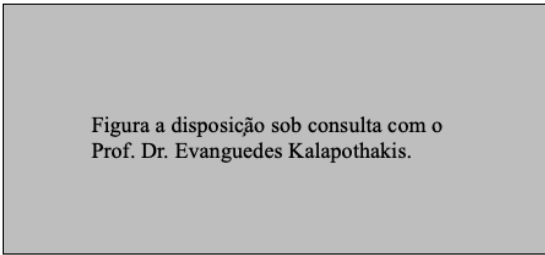


Figura a disposição sob consulta com o Prof. Dr. Evanguedes Kalapothakis.

**Figura 41- Alelos encontrados para o locus L49.** Foram encontrados seis alelos para L49 em todas as amostras sequenciadas e genotipadas de *S. similis*. No alinhamento mostrado na figura é possível observar a variação no número de repetições de microssatélites e também SNV nas sequências. No eixo X há uma referência de tamanho de pares de bases e à direita há a informação do tamanho do alelo, além do número de vezes que o respectivo alelo apareceu na tabela de genotipagem. A imagem foi gerada pelo programa CHIIMP (Barbian 2018).

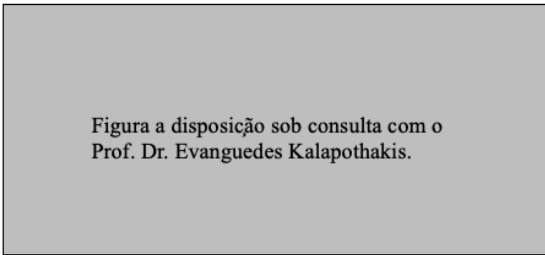


Figura a disposição sob consulta com o Prof. Dr. Evanguedes Kalapothakis.

**Figura 42- Alelos encontrados para o locus L57.** Foram encontrados 38 alelos para L57 em todas as amostras sequenciadas e genotipadas de *S. similis*. No alinhamento mostrado na figura é possível observar a variação no número de repetições de microssatélites e também SNV nas sequências. No eixo X há uma referência de tamanho de pares de bases e à direita há a informação do tamanho do alelo, além do número de vezes que o respectivo alelo apareceu na tabela de genotipagem. A imagem foi gerada pelo programa CHIIMP (Barbian 2018).

Os genótipos das amostras de criatório (AC) obtidos foram posteriormente submetidos ao programa Cervus 3.0.7 (Marshall 1998) para cálculo de frequência de alelos nulos (tabela 14). De acordo com Dakin (2004), alelos nulos são alelos que costumam não amplificar em níveis detectáveis via PCR por três principais motivos: 1- mal anelamento dos *primers* devido à divergência de sequência de nucleotídeos quando utilizado *primers* não específicos para a espécie (Kwok 1990, Primmer 1995); 2- amplificação diferencial de alelos de tamanhos diferentes, sendo que alelos menores geralmente amplificam de forma mais eficiente do que alelos maiores (Wattier 1998) e 3- falha na PCR devido à baixa qualidade ou quantidade do DNA usado como molde (Gagneux 1997). Alguns fatores podem dar a falsa impressão de que os alelos nulos estão presentes, como, por exemplo, os cruzamentos consanguíneos, que causam uma redução nos heterozigotos em relação ao equilíbrio de Hardy-Weinberg e podem dar uma falsa evidência de alelos nulos (Chakraborty 1992), assim como os alelos de *loci* nos cromossomos sexuais, que se não forem reconhecidos podem gerar uma redução nos heterozigotos e serem falsos indicativos de alelos nulos. Quando presentes em heterozigotos, podem resultar como homozigotos; e quando presentes em homozigotos podem resultar em ausência de amplificação (Guichoux 2011). O estudo de Dakin (2004) concluiu que alelos nulos provavelmente não terão muito impacto nas probabilidades médias de exclusão na maioria dos estudos de parentesco, considerando que sejam raros e as análises de parentesco são baseadas em um conjunto de *loci* de microssatélites. Já alelos nulos muito frequentes podem causar desvios em algumas estimativas e, portanto, costumam ser excluídos das análises genéticas. Chapuis (2007) estabelece que alelos nulos com frequência  $\geq 0.2$  são considerados muito frequentes, e, portanto, neste trabalho, os *loci* L11, L14, L19, L24 e L35 (frequência de alelos nulos: 0.2, 0.5, 0.4, 0.6 e 0.3, respectivamente) não são muito indicados para estudos de paternidade e foram removidos das análises posteriores. Adicionalmente, para L19 e L24, os dois *loci* com maior frequência de alelos nulos, os genótipos dos progenitores e filhotes tiveram incompatibilidades em todas as famílias analisadas. Nota-se que a frequência de alelos nulos para L36 não foi realizada pelo programa por Cervus, porém, uma análise dos genótipos de pais e filhotes mostrou inconsistência de dois filhotes (AC3 – filhote de AC1 e AC2, e AC8- filhote de AC6 e AC7) compatível com a alta frequência de alelos nulos, já que os filhotes foram genotipados como homozigotos, sendo que os pais de AC8 também eram homozigotos para L36 porém de alelos diferentes. Por precaução, o L36 também foi retirado das análises posteriores.

**Tabela 14- Frequência de alelos nulos.**

<i>Locus</i>	<b>k</b>	<b>N</b>	<b>F(Nulo)</b>
L11	5	21	0.2142
L14	6	10	0.5419
L15	9	17	0.0552
L16	8	17	-0.002
L20	1	1	ND
L23	2	1	ND
L26	2	21	-0.0046
L29	4	7	ND
L35	5	16	0.3592
L36	6	7	ND
L37	7	13	0.0955
L38	16	19	0.0091
L39	8	20	0.0331
L49	3	21	0.1046
L57	5	4	ND
L19	6	21	0.4463
L24	8	20	0.6514

K= número de alelos no *locus*, N=número de indivíduos genotipados para o *locus*, F(Nulo)= Frequência de alelos nulos, ND= Não realizado. Dados da tabela gerados pelo programa Cervus 3.0.7 (Marhsall 1998) utilizando os genótipos das amostras de criatório (AC).

Os genótipos de todos os indivíduos sequenciados foram submetidos ao programa Cervus 3.0.7 (Marshall 1998) para cálculos de frequência alélica juntamente com estatísticas como  $H_o$ ,  $H_e$ , PIC e equilíbrio de Hardy-Weinberg (EHW) e os resultados para cada *locus* são mostrados na tabela 15, abaixo. Nota-se que os valores obtidos quando fornecidos apenas os dados das amostras de criatório foram, de modo geral, similares (com exceção daqueles em que o número de indivíduos AC genotipados foi muito baixo), indicando que a população do criatório reflete a diversidade genética da espécie. Os valores de  $H_o$ ,  $H_e$  (também chamado de GD – diversidade genética) e PIC também foram calculados com o programa STRAF 1.0.5 (Gouy 2017) e apresentaram resultados iguais aos gerados pelo Cervus 3.0.7. A vantagem da utilização do STRAF se dá por ser uma plataforma de uso online, que não necessita a realização de download, e também realiza outras estatísticas populacionais e forenses não abordadas pelo Cervus.

**Tabela 15- Frequências alélicas e estatísticas de todos os indivíduos sequenciados.**

<i>Locus</i>	Todos os indivíduos (TF+AC)						Indivíduos de criatório (AC)				
	k	N	Ho	He	PIC	HW	k	N	Ho	He	PIC
L15	18	69	0.739	0.796	0.766	NS	9	17	0.706	0.806	0.757
L16	14	67	0.627	0.761	0.731	NS	8	17	0.824	0.859	0.813
L20	11	46	0.587	0.769	0.73	NS	1	1	0	0	0
L23	29	48	0.479	0.944	0.931	ND	2	1	1	1	0.375
L26	3	71	0.028	0.028	0.028	ND	2	21	0.048	0.048	0.045
L29	5	8	0.125	0.775	0.682	ND	4	7	0	0.747	0.641
L37	21	65	0.769	0.829	0.805	NS	7	13	0.615	0.742	0.679
L38	53	71	0.761	0.976	0.968	ND	16	19	0.895	0.937	0.906
L39	19	72	0.681	0.824	0.798	NS	8	20	0.75	0.815	0.766
L49	6	63	0.317	0.644	0.574	*	3	21	0.476	0.61	0.527
L57	38	56	0.786	0.962	0.952	ND	5	4	0.5	0.857	0.712

K= número de alelos no *locus*, N=número de indivíduos genotipados para o *locus*, Ho= Heterozigosidade observada, He= Heterozigosidade esperada, PIC= Conteúdo de Informatividade Polimórfica, HW= Desvio do equilíbrio de Hardy-Weinberg com correção de Bonferroni (NS= Não significativo, \*=significativo no nível de 5%, ND= Não realizado).O HW não foi calculado quando somente os genótipos dos indivíduos de criatório foram fornecidos. Dados da tabela gerados pelo programa Cervus 3.0.7 (Marhsall 1998) utilizando os genótipos de todos os indivíduos sequenciados.

Em relação ao valor de PIC, a grande maioria dos *loci* estudados apresentaram valores superiores a 0.5, o que significa que são muito informativos, com exceção do L26. Assim, o L26, juntamente com os outros seis *loci* mencionados anteriormente, foram excluídos das análises posteriores. Por possuir baixo PIC, o L26 também possui valores altos para as probabilidades de não exclusão, ou seja, são pouco informativos para análises de parentesco.

#### 1.5.1.7 Teste de paternidade piloto

Após as análises discutidas no tópico anterior, apenas dez *loci* foram mantidos para a análise de paternidade (L15, L16, L20, L23, L29, L37, L38, L39, L49 e L57). Os genótipos destes *loci* das amostras de famílias completas de criatório (AC1-AC19) foram submetidos ao teste de paternidade do programa Cervus 3.0.7, que utiliza a razão de verossimilhança (do inglês “*likelihood ratio*”), também conhecido como índice de paternidade, para identificar os pais mais prováveis. Neste cenário, há duas hipóteses possíveis: o suposto pai é o verdadeiro pai ou o suposto pai não é o verdadeiro pai, sendo que a razão de verossimilhança é calculada pela probabilidade de que o suposto pai é o verdadeiro pai dividido pela probabilidade do suposto pai não ser o verdadeiro pai, probabilidades essas que são calculadas a partir da probabilidade de se obter os genótipos

observados. Para diferenciar genótipos frequentes que gerariam alta chance de compatibilidade de genótipos raros que possuem menor chance de compatibilidade, são utilizadas as frequências alélicas de cada *locus*, e assume-se que estes possuem herança mendeliana. As razões de verossimilhança de cada *locus* são multiplicadas para se obter a razão de verossimilhança total, expressa no Cervus como valor de LOD, que é obtido tomando o log natural da razão de verossimilhança. Uma pontuação positiva de LOD indica que há mais chances de o possível progenitor ser o verdadeiro pai do que de não ser o verdadeiro pai; enquanto que se as chances de um possível pai ser o verdadeiro pai e de não ser o verdadeiro pai são iguais, o valor de LOD será igual a zero. Já nos casos que o valor de LOD é negativo, a chance de que o suposto pai é o pai verdadeiro é menos provável do que a chance de não ser o verdadeiro pai.

Os *loci* L20, L23, L29 e L57 tiveram muitos indivíduos de criatório sem genótipo definido, e, portanto, na maioria dos casos, não foram utilizados neste teste de paternidade, fazendo com que a análise fosse realizada com menos de dez *loci*. O resultado da análise de paternidade do programa Cervus, mostrado na tabela abaixo (tabela 16), foi condizente com o observado manualmente, em especial ao filhote AC14 não ser filhote de AC9 e AC13, uma vez que este filhote apresentou incompatibilidade com os pais sugeridos em todos os *loci* analisados. Possivelmente ocorreu um erro no armazenamento das informações de parentesco deste filhote no criatório que forneceu as amostras. O filhote AC16 apresentou *mismatch* em um *locus* com o possível pai (AC1), e uma análise manual mostra que essa incompatibilidade ocorreu no *locus* L29 (genótipos mostrados na tabela 17), no qual ambos possuem alelos com 148pb de tamanho, porém possuem quatro SNVs ao longo da sequência (figura 43), sendo que nenhuma diferença é encontrada na região dos microssatélites. Apesar deste *mismatch*, AC1 ainda foi considerado o pai de AC16 com o nível de confiança relaxado. Vale ressaltar que caso as sequências não fossem levadas em consideração na análise, mas somente o tamanho dos fragmentos, este *mismatch* não teria sido observado. Com exceção dos progenitores 1 e 2 de AC14 e do progenitor 2 de AC16, todos os demais foram considerados os verdadeiros progenitores dos respectivos filhotes com nível de confiança de 95%. Conforme esperado, os irmãos AC20 e AC21 não foram compatíveis com nenhum dos progenitores.

Como os testes de paternidade para espécies de aves traficadas ainda são um assunto relativamente novo no contexto de regulação, ainda não foi, no momento, definido como estes testes serão analisados. Para humanos, por exemplo, as estatísticas



de verossimilhança e poder de exclusão podem ser utilizados, enquanto que para animais outro sistema é utilizado. A Instrução Normativa do Ministério da Agricultura e Pecuária (MAPA) nº45, de 15 de dezembro de 2017, traz de forma detalhada os *loci* que devem ser utilizados na genotipagem de animais das espécies bovina, equina, ovina, caprina, bubalina e asinina, sendo todos descritos e validados pela Sociedade Internacional de Genética Animal (ISAG). Para estes, devem ser genotipados no mínimo 12-14 *loci* de acordo com a espécie, e caso ocorra uma discrepância, é necessário a genotipagem de mais alguns *loci* de microssatélites para compor o painel.

**Tabela 16- Resumo do resultado da análise de paternidade das amostras de criatório.** Resumo do arquivo de saída do programa Cervus 3.0.7 para a análise de parentesco das amostras de criatório mostrando os possíveis pais de cada filhote.

<b>Prole</b>	<b>Possível progenitor 1</b>	<b>Mismatches entre prole e possível progenitor 1</b>	<b>LOD prole x progenitor 1</b>	<b>Confiança 1</b>	<b>Possível progenitor 2</b>	<b>Mismatches entre prole e possível progenitor 2</b>	<b>LOD prole x progenitor 2</b>	<b>Confiança 2</b>	<b>Número de loci comparados no trio</b>	<b>Mismatches do trio</b>	<b>LOD trio</b>	<b>Confiança trio</b>
AC3	AC2	0	2.68E+00	*	AC1	0	3.92E+00	*	6	0	7.76E+00	*
AC4	AC2	0	1.79E+00	*	AC1	0	8.17E-01	*	4	0	3.98E+00	+
AC5	AC2	0	2.68E+00	*	AC1	0	1.71E+00	*	5	0	5.54E+00	+
AC8	AC7	0	3.23E+00	*	AC6	0	3.88E+00	*	6	0	8.21E+00	*
AC11	AC10	0	4.13E+00	*	AC9	0	4.51E+00	*	6	0	1.21E+01	*
AC12	AC10	0	1.60E+00	*	AC9	0	3.61E+00	*	5	0	7.94E+00	*
AC14	AC13	4	#####	-	AC9	3	#####	-	6	4	#####	-
AC16	AC15	0	3.94E+00	*	AC1	1	#####	+	6	1	2.49E+00	+
AC19	AC18	0	3.38E+00	*	AC17	0	1.14E+00	*	5	0	5.60E+00	+

Trio= filhote e os dois possíveis progenitores

\*Nível estrito de confiança: 95% de confiança no resultado positivo

+Nível relaxado de confiança: 80% de confiança no resultado positivo

- Provável não atribuição de parentesco

**Tabela 17- Genótipo de filhotes e progenitores.** Os códigos dos alelos foram gerados automaticamente pelo programa CHIIMP, e a parte anterior ao hífen corresponde ao tamanho do fragmento analisado. Os destaques com fundo acinzentado são os loci que apresentaram incompatibilidade entre progenitor(es) e filhote.

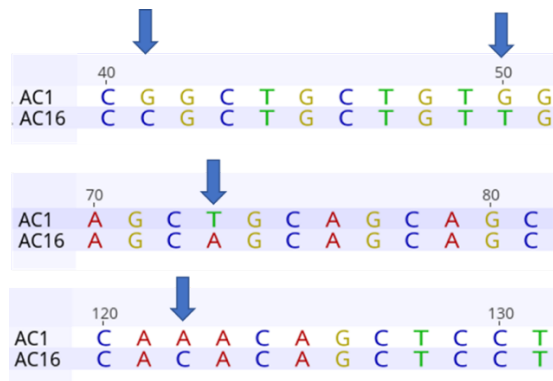
	Indivíduo	Parentesco	L15a	L15b	L16a	L16b	L20a	L20b	L23a	L23b	L29a	L29b
Família 1	AC1	pai	98- 7cf295	100- 6ab47b	0	0	0	0	0	0	148- flcb17	148- flcb17
	AC2	mãe	98- 9bdf23	100- 6ab47b	91- 531363	81- 7f9743	0	0	0	0	0	0
	AC3	filhote	100- 6ab47b	100- 6ab47b	81- 7f9743	89- 8f1ddb	0	0	0	0	148- flcb17	148- flcb17
	AC4	filhote	98- 9bdf23	0	81- 7f9743	89- 8f1ddb	0	0	0	0	0	0
	AC5	filhote	100- 6ab47b	100- 6ab47b	81- 7f9743	89- 3e2797	0	0	0	0	166- 69dda5	0
Família 2	AC6	pai	96- c1b5e9	100- 6ab47b	89- 8f1ddb	99- dab55e	0	0	0	0	0	0
	AC7	mãe	0	0	89- 3e2797	89- 8f1ddb	0	0	130- 0e8266	0	0	0
	AC8	filhote	96- c1b5e9	96- c1b5e9	89- 8f1ddb	89- 3e2797	109- e7f483	109- e7f483	0	0	0	0
Família 3	AC9	pai	96- d785bd	100- 6ab47b	89- bc55ed	91- 531363	0	0	133- cbd1da	0	148- 3af74e	0
	AC10	mãe	102- 4530c3	110- dba4c0	81- 7f9743	89- 3e2797	0	0	0	0	0	0
	AC11	filhote	102- 4530c3	96- d785bd	91- 531363	81- 7f9743	0	0	0	0	0	0
	AC12	filhote	100- 6ab47b	0	89- 3e2797	91- 531363	0	0	0	0	148- 3af74e	148- 3af74e

	Indivíduo	Parentesco	L15a	L15b	L16a	L16b	L20a	L20b	L23a	L23b	L29a	L29b
Família 4	AC9	pai	96- d785bd	100- 6ab47b	89- bc55ed	91- 531363	0	0	133- cbd1da	0	148- 3af74e	0
	AC13	mãe	96- d785bd	102- 4530c3	0	0	0	0	133- d75386	139- 77c452	0	0
	AC14	filhote	98- 7cf295	98- 7cf295	89- bc55ed	89- bc55ed	0	0	0	0	145- d6ec0d	145- d6ec0d
Família 5	AC1	pai	98- 7cf295	100- 6ab47b	0	0	0	0	0	0	148- f1cb17	148- f1cb17
	AC15	mãe	98- 7cf295	100- 6ab47b	93- b528e9	93- b528e9	0	0	142- 1e8dec	0	0	0
	AC16	filhote	100- 6ab47b	100- 6ab47b	0	0	0	0	0	0	148- 39865f	148- 39865f
Família 6	AC17	pai	98- 7cf295	0	93- c566e2	89- 3e2797	0	0	0	0	148- 39865f	148- 39865f
	AC18	mãe	100- b6f94c	98- 7cf295	89- 3e2797	89- 3e2797	0	0	0	0	148- 39865f	148- 39865f
	AC19	filhote	98- 7cf295	112- 4b731c	93- c566e2	0	0	0	0	0	0	0

	Indivíduo	parentesco	L36a	L36b	L37a	L37b	L38a	L38b	L39a	L39b
<b>Família 1</b>	AC1	pai	118- 0fa61a	118- 0fa61a	119- bf4bd5		134- fa9466	104- ab9f82	97- ea259f	103-dbd76f
	AC2	mãe	139- f22238	127- f672bf		0	116- 3ebf23	116- 3ebf23	97- 20e048	121-0ae92d
	AC3	filhote	118- 0fa61a		0	0	116- 3ebf23	134- fa9466	121- 0ae92d	97-ea259f
	AC4	filhote		0	0	0	134- fa9466	116- 3ebf23	121- 0ae92d	103-dbd76f
	AC5	filhote		0	125- 8b7bb6		116- 3ebf23	134- fa9466	103- dbd76f	121-0ae92d
<b>Família 2</b>	AC6	pai	136- ecf5c8		122- 9dbb1d	125- 8b7bb6	134- a20c61	122- ea2fef	103- dbd76f	103-c065bf
	AC7	mãe	127- bf524b		122- acac76	122- acac76	134- 8d78c9	0	103- f29a4d	103-f29a4d
	AC8	filhote	145- 56d212		122- acac76	125- 8b7bb6	134- a20c61	134- a20c61	103- f29a4d	103-dbd76f
<b>Família 3</b>	AC9	pai	118- 0fa61a	118- 0fa61a	125- 8b7bb6	125- 8b7bb6	116- 27a623	131- 94c6f7	103- f29a4d	97-ea259f
	AC10	mãe	148- 141eaf	148- 141eaf	125- 8b7bb6	125- 8b7bb6	104- 6a9d12	137- 43304c	97- 20e048	97-20e048
	AC11	filhote		0	125- 8b7bb6	125- 8b7bb6	116- 27a623	104- 6a9d12	97- 20e048	103-f29a4d
	AC12	filhote		0	125- 8b7bb6	125- 8b7bb6	116- 27a623	137- 43304c	103- f29a4d	97-20e048

	Indivíduo	parentesco	L36a	L36b	L37a	L37b	L38a	L38b	L39a	L39b
Família 4	AC9	pai	118-0fa61a	118-0fa61a	125-8b7bb6	125-8b7bb6	116-27a623	131-94c6f7	103-f29a4d	97-ea259f
	AC13	mãe	0	0	0	0	baddda	43304c	d8eeeb	cb204f
	AC14	filhote	0	0	122-acac76	125-8b7bb6	119-5fcfc4	104-ab9f82	97-20e048	103-dbd76f
Família 5	AC1	pai	118-0fa61a	118-0fa61a	119-bf4bd5	0	134-fa9466	104-ab9f82	97-ea259f	103-dbd76f
	AC15	mãe	133-fld225	133-fld225	122-f02b57	122-acac76	122-e0f4b3	125-66a356	97-20e048	97-20e048
	AC16	filhote	0	0	119-bf4bd5	122-f02b57	104-ab9f82	122-e0f4b3	103-dbd76f	97-20e048
Família 6	AC17	pai	0	0	0	0	104-ab9f82	134-8d78c9	0	0
	AC18	mãe	118-0fa61a	118-0fa61a	122-21c65d	128-b6d990	104-ab9f82	122-f3da73	103-dbd76f	121-0ae92d
	AC19	filhote	0	0	128-b6d990	122-acac76	104-ab9f82	122-f3da73	121-0ae92d	97-20e048

	<b>Indivíduo</b>	<b>parentesco</b>	<b>L49a</b>	<b>L49b</b>	<b>L57a</b>	<b>L57b</b>
<b>Família 1</b>	AC1	pai	105- 9383de	113- d5c755	0	0
	AC2	mãe	105- 9383de	113- d5c755	0	0
	AC3	filhote	113- d5c755	105- 9383de	0	0
	AC4	filhote	105- 9383de	105- 9383de	0	0
	AC5	filhote	105- 9383de	113- d5c755	0	0
<b>Família 2</b>	AC6	pai	105- 9383de	105- 9383de	135- 2e9ec9	150- c828e8
	AC7	mãe	105- 9383de	105- 9383de	0	0
	AC8	filhote	105- 9383de	105- 9383de	0	0
<b>Família 3</b>	AC9	pai	109- 43f73a	109- 43f73a	0	0
	AC10	mãe	105- 9383de	113- d5c755	0	0
	AC11	filhote	109- 43f73a	113- d5c755	125- 805a7f	125- 805a7f
	AC12	filhote	105- 9383de	109- 43f73a	0	0
<b>Família 4</b>	AC9	pai	109- 43f73a	109- 43f73a	0	0
	AC13	mãe	105- 9383de	105- 9383de	0	0
	AC14	filhote	109- 43f73a	109- 43f73a	145- f33326	145- f33326
<b>Família 5</b>	AC1	pai	105- 9383de	113- d5c755	0	0
	AC15	mãe	105- 9383de	113- d5c755	145- bfd640	145- f33326
	AC16	filhote	105- 9383de	113- d5c755	0	0
<b>Família 6</b>	AC17	pai	109- 43f73a	109- 43f73a	0	0
	AC18	mãe	105- 9383de	105- 9383de	150- f07225	0
	AC19	filhote	105- 9383de	109- 43f73a	0	0

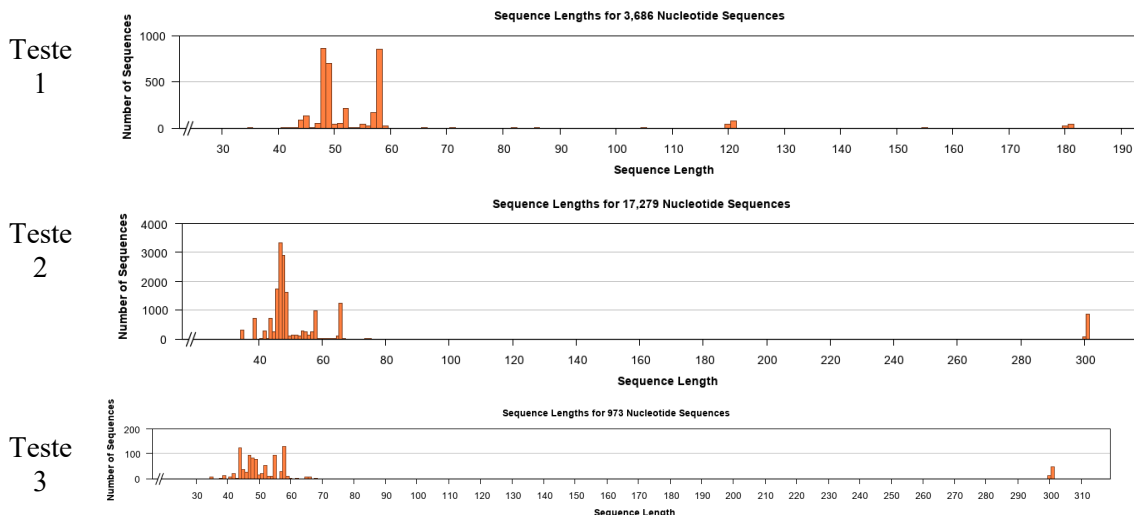


**Figura 43- Mismatches nas sequências de pai (AC1) e filhote (AC16) no locus L29.** A análise de paternidade do programa Cervus identificou um *mismatch* entre o possível pai (AC1) e o filhote (AC16). Ambos possuíam alelos diferentes, porém com mesmo tamanho. As diferenças observadas entre os dois alelos foram quatro SNVs fora da região do microsatélites e são indicados na imagem pela seta azul. Imagem gerada no programa Geneious.

#### 1.5.1.8 Controles negativos

Os controles negativos de todos os testes também foram sequenciados e as suas sequências analisadas. A ausência de resultados de genotipagem para o controle negativo indica que estas sequências não possuem informações suficientes para as análises. Conforme mostrado na figura 44, abaixo, a grande maioria das *reads* dos controles negativos eram de sequências curtas (de até cerca de 60pb), sendo basicamente dímeros de *primers*. Já nos gráficos de tamanho das *reads* das amostras sequenciadas, a maioria possuía tamanho entre 100 e 180pb, abrangendo as sequências alvo (anexo I). Em alguns casos, havia também algumas sequências que não correspondiam à sequência referência nem a dímeros de *primers*, podendo ser resultado de sequências de *index*, adaptadores ou de combinação de *primers* de diferentes *loci*, como, por exemplo, mostrado na figura 45, abaixo, do mapeamento das *reads* do controle negativo do teste 2 contra a referência de L37. Vale ressaltar que as sequências demonstradas nas imagens são de *reads* brutas, e que quando utilizados filtros de trimagem das *reads* de acordo com o tamanho, boa parte das sequências alinhadas nos controles negativos são descartadas.





**Figura 44- Tamanho das *reads* dos controles negativos.** As imagens mostram os gráficos de tamanho das *reads* brutas dos controles negativos dos três testes realizados no MiSeq, e, para todos, a grande maioria das *reads* tem tamanho de até cerca de 60 pares de bases, tamanho compatível com os dímeros de *primers*. Os gráficos foram gerados pelo programa Geneious 11.1.15 (<https://www.geneious.com>).

Figura a disposição sob consulta com o Prof. Dr. Evanguedes Kalapothakis.

**Figura 45- *Reads* do controle negativo do teste 2 mapeadas contra a referência do L37.** A- Parte das *reads* continham a sequência do *primer forward* e sequência complementar do *primer reverse*, resultante da formação de dímeros de *primers*. B- A sequência complementar do *primer reverse* está destacada em vermelho, seguido por sequências posteriores que não continham sequências do *amplicon* em análise. O mapeamento foi feito com o programa Geneious 11.1.15 (<https://www.geneious.com>).

Apesar de não influenciarem nos resultados das análises dos genótipos, a existência destas sequências pode prejudicar o rendimento do sequenciamento, uma vez que utiliza reagentes que poderiam ser utilizados para sequências de interesse. As sequências geradas nos controles negativos poderiam ser possivelmente evitadas se as PCRs fossem realizadas com reagentes do tipo *Hot start*, que ao reduzir a amplificação não específica e formação de dímeros de *primers*, otimizam o rendimento das amplificações desejadas (Green&Sambrook 2018). Uma outra possível alternativa poderia ser a substituição do uso da Taq polimerase (enzima encontrada em *Thermus aquaticus*) pela Pfu polimerase (enzima encontrada em *Pyrococcus furiosus*), que previne o anelamento dos *primers* antes que a temperatura de anelamento seja atingida e também possui menor taxa de erro (Goldbio 2018).

1.5.2 DNA mitochondrial de *Saltator similis*

MITOCHONDRIAL DNA PART B  
2019, VOL. 4, NO. 1, 2058–2060  
<https://doi.org/10.1080/23802359.2019.1618210>



MITOGENOME ANNOUNCEMENT

OPEN ACCESS

## Complete mitochondrial genome of *Saltator similis* (Aves, Passeriformes)

Ana Luiza Lemos Queiroz , Adriana Heloísa Pereira , Nazaré Lucio , Yan Kalapothakis ,  
Kelton Gonçalves Miranda and Evanguedes Kalapothakis

Laboratory of Biotechnology and Molecular Markers, Biological Sciences Institute, Federal University of Minas Gerais, Belo Horizonte, MG, Brazil

### ABSTRACT

*Saltator similis*, popularly known as Green-winged Saltator, is one of the most trafficked species in Brazil. DNA from a muscle tissue sample was sequenced on MiSeq (Illumina) sequencer. The reads were assembled to reference using Geneious. The mtDNA consisted of 16,750 base pairs containing 2 ribosomal RNA, 22 transporter RNA, 13 protein-coding genes, and 1 control region. Most of the tRNA and PCGs were encoded on the heavy strand.

### ARTICLE HISTORY

Received 24 April 2019  
Accepted 3 May 2019

### KEYWORDS

Wildlife trafficking; green-winged saltator; next-generation sequencing; mtDNA

*Saltator similis* (d'Orbigny & Lafresnaye, 1837 in Guala & Döring 2019) popularly known as Green-winged Saltator, is a member of order Passeriformes. It inhabits in parts of Brazil, Bolivia, Uruguay, Paraguay and Argentina (BirdLife International 2018). The current decrease in population is due to the loss of habitats and the illegal capture (UFV 2009, Manhães & Loures-Ribeiro 2011). *Saltator similis* is one of the most seized species in actions to combat the illegal trade of wild animals in Brazil (Borges et al. 2006; SEMA/PMA 2006; Gogliath et al. 2010; Destro 2012; Nunes et al. 2012; Franco et al. 2012; Felker et al. 2013; Zocche and Vianna 2013; Silva 2015; Freitas et al. 2015). In Brazilian legislation, captive breeding is permitted. To be a legal bird, it must be born in captivity and have an official inviolable band.

A population of female individual from *Saltator similis* species deceased as a consequence of illegal trafficking. It was identified and donated by IBAMA (Brazilian Environmental Agency) with proper licensing from the competent authorities (SISBIO 56471-1, IEF 024/2016 and CEUA-UFMG 37/2017) and mitochondrial DNA was extracted following the protocol from Françoso et al. (2015). The specimen was taxidermized and is in the Center of Taxonomic Collections of Federal University of Minas Gerais (voucher DZ7289).

DNA library was constructed and sequenced using a single-read 300-bp strategy on a MiSeq system (Illumina, San Diego, CA) using MiSeq Reagent Kit V2-300. Reads were trimmed with Phred 30 and assembled to reference (mtDNA of *Geospiza magnirostris* – GenBank MG682351.1) using

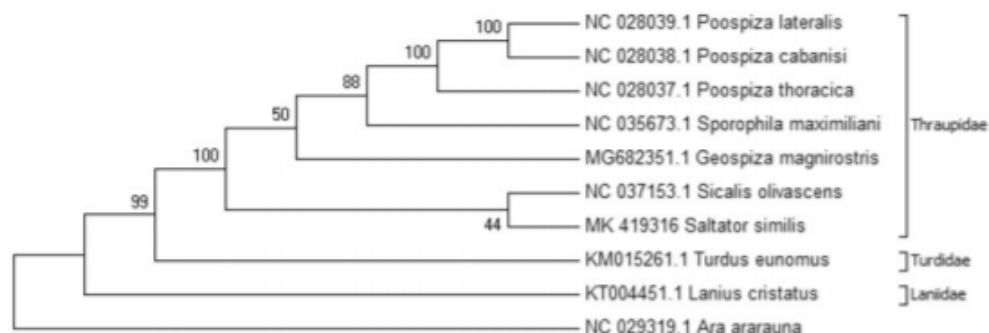
Geneious 11.1.5 (<https://www.geneious.com>). The complete mitochondrial DNA of *S. similis* (GenBank accession number MK419316) was found to be circular in shape with 16,750 bp length, with average coverage of 148 reads. mtDNA was submitted to nucleotide BLAST (Altschul et al. 1990), which resulted in high similarity with the mitogenome of other Passeriformes species. The mtDNA displayed a GC content of 46.93% and base frequencies were 30% A, 14.36% G, 23% T, and 32.57% C.

The mitogenome was annotated with MITOS (Bernt et al. 2013) and verified with ExPASy (Gasteiger et al. 2003). Genes in *S. similis* were arranged similarly to other avian mitogenomes (Raposo do Amaral et al. 2015; Ludwig et al. 2017; Liang et al. 2018), with a different organization of the genes near the control region when compared to typical vertebrate mitogenome (Quinn and Wilson 1993). The mtDNA structure contained 2 rRNA, 22 tRNA, 13 protein-coding genes (PCGs), and 1 control region (D-loop). Protein-coding genes commonly had ATG as start codon (11 PCGs). Nine of the 13 PCGs contained a TAA stop codon. Eight of the 22 tRNAs and one PCG (Nd6) were encoded on the light (L) strand. The remaining genes were encoded on the heavy (H) strand. We found 10 overlapping regions.

Phylogenetic analyses were conducted using MEGA version 6.06 (MEGA Inc., Englewood, NJ) (Tamura et al. 2013) using the neighbor-joining method (Saitou and Nei 1987). D-loop region was excluded from phylogenetic analysis due to its hypervariability (Gonder et al. 2007) (Figure 1).

**CONTACT** Evanguedes Kalapothakis [kalapothakis@gmail.com](mailto:kalapothakis@gmail.com) Laboratory of Biotechnology and Molecular Markers, Department of General Biology, Institute of Biological Sciences, Federal University of Minas Gerais, Av. Antônio Carlos, 6627, Pampulha CEP:31270-901, Belo Horizonte, Brazil

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.  
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Figure 1.** Consensus neighbor-joining tree of 10 avian species. Analyses were conducted in MEGA6 (Tamura et al. 2013). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches (Felsenstein 1985). The evolutionary distances were computed using the Kimura 2-parameter method (Kimura 1980) and are in the units of the number of base substitutions per site. The analysis involved 10 nucleotide sequences. Codon positions included were 1st + 2nd + 3rd. Thraupidae Family grouped together with a high bootstrap value including *Saltator similis*. Turdidae and Laniidae Family were grouped in separate clades. Tree was rooted using *Ara ararauna* as the outgroup.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Acknowledgments

We thank IBAMA, IEF and colleagues of the Laboratory of Biotechnology and Molecular Markers for their partnership. We thank Jean Oliveira and team of Taxonomy Collection Center for their collaborations.

## Funding

This study was supported by CAPES [Edital Ciências Forenses n<sup>o</sup> 25/2014, Processo 23038.006839/2014-33, 3405/2014] and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) [307550/2015-0].

## ORCID

Ana Luiza Lemos Queiroz  <http://orcid.org/0000-0002-1362-7713>  
 Adriana Heloisa Pereira  <http://orcid.org/0000-0003-1805-1028>  
 Nazaré Lucio  <http://orcid.org/0000-0002-8766-9916>  
 Yan Kalapothakis  <http://orcid.org/0000-0001-8914-0023>  
 Kelton Gonçalves Miranda  <http://orcid.org/0000-0002-2899-4620>  
 Evanguedes Kalapothakis  <http://orcid.org/0000-0002-8326-249X>

## References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol.* 215:403–410.
- Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsche G, Pütz J, Middendorf M, Stadler PF. 2013. MITOS: improved de novo metazoan mitochondrial genome annotation. *Mol Phylogenet Evol.* 69:313–319.
- BirdLife International. 2018. Species factsheet: *Saltator similis*; [accessed 2018 Oct 25]. <http://www.birdlife.org>.
- Borges RC, Oliveira A, Bernardo N, Costa RMMC. 2006. Diagnóstico da fauna silvestre apreendida e recolhida pela Polícia Militar de Meio Ambiente de Juiz de Fora, MG (1998 e 1999). *Revista Brasileira de Zoológicas.* 8:23–33.
- Destro GFG. 2012. Efforts to combat wild animals trafficking in Brazil. In: Lammed GA, editor. *Biodiversity enrichment in a diverse world*. Sao Paulo: InTech; p. 421–436.
- Felker RM, Dörr AC, Rovedder AP, Piazza EM, Dick G. 2013. Levantamento parcial da avifauna apreendida pelo Escritório Regional do Ibama de Santa Maria-RS. *Revista Eletrônica em Gestão, Educação e Tecnologia Ambiental*, Santa Maria, v 11, n. 11, p. 2506–2510; [accessed 2016 Aug 18]. <http://cascavel.ufsm.br/revistas/ojs-2.2.2/index.php/reget/article/download/8734/pdf>
- Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution.* 39:783–791.
- Franco MR, Câmara FM, Rocha DCC, Souza RM, Oliveira NJF. 2012. Animais silvestre apreendidos no período de 2002 a 2007 na macro-região de Montes Claros. Minas Gerais. *Enciclopédia Biosfera, Centro Científico Conhecer.* 8:1007–1018.
- Françoso E, Gomes F, Arias MC. 2015. A protocol for isolating insect mitochondrial genomes: a case study of NUMT in *Melipona flavolineata* (Hymenoptera: Apidae). *Mitochondrial DNA Part A.* 27:1–4.
- Freitas ACP, Oviedo-Pastrana ME, Vilela DAR, Pereira PLL, Loureiro LOC, Haddad JPA, Martins NRS, Soares DFM. 2015. Diagnóstico de animais ilegais recebidos no centro de triagem de animais silvestres de Belo Horizonte, Estado de Minas Gerais, no ano de 2011. *Cienc Rural.* 45: 163–170.
- Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. 2003. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* 31:3784–3788.
- Gogliath M, Bisaggio EL, Ribeiro LB, Resgalla AE, Borges RC. 2010. Avifauna apreendida e entregue voluntariamente ao Centro de Triagem de Animais Silvestres (Cetas) do Ibama de Juiz de Fora, Minas Gerais. *Atualidades Ornitológicas on-Line;* 154:55–59 [accessed 2016 Aug 18]. [http://www.ao.com.br/download/ao154\\_55.pdf](http://www.ao.com.br/download/ao154_55.pdf).
- Gonder MK, Mortensen HM, Reed FA, de Sousa A, Tishkoff SA. 2007. Whole-mtDNA genome sequence analysis of ancient African Lineages. *Mol Biol Evol.* 24:757–768.
- Guala G, Döring M. 2019. Integrated Taxonomic Information System (ITIS). National Museum of Natural History, Smithsonian Institution. [accessed via GBIF.org on 2019-05-17] Checklist dataset <https://doi.org/10.15468/rjarmt>
- Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol.* 16:111–120.
- Liang W, Yunlin Z, Zhenggang X, Tian H, Libo Z, Shiquan L. 2018. The complete mitochondrial genome and phylogeny of *Geospiza magnirostris* (Passeriformes: Thraupidae). *Conserv Gen Res.* 2: 1–3.
- Ludwig S, Martins APV, Queiroz ALL, Carmo AO, Oliveira-Mendes BBR, Kalapothakis E. 2017. Complete mitochondrial genome of *Sporophila maximiliani* (Ave, Passeriformes). *Mitochondrial DNA Part B.* 2:2: 417–418.
- Manhães MA, Loures-Ribeiro A. 2011. The avifauna of the Poço D'Anta Municipal Biological Reserve, Juiz de Fora, MG. *Biota Neotrop.* 11: 275–286.



- Nunes PB, Barreto AS, Franco EZ. 2012. Subsídios à ação fiscalizatória no combate ao tráfico de aves silvestres e exóticas em Santa Catarina. *Ornithologia. Matinhos*, n. 5, n. 1, p. 28–33; [accessed 2016 Aug 2]. <http://www.cemave.net/publicacoes/index.php/ornithologia/article/view/75/78>.
- Quinn TW, Wilson AC. 1993. Sequence evolution in and around the mitochondrial control region in birds. *J Mol Evol.* 37:417–425.
- Raposo do Amaral F, Neves LG, Resende MF Jr, Mobili F, Miyaki CY, Pellegrino KC, Biondo C. 2015. Ultraconserved elements sequencing as a low-cost source of complete mitochondrial genomes and microsatellite markers in non-model amniotes. *PLoS One.* 10: e0138446.
- Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol.* 4:406–425.
- SEMA/PMA. Polícia Militar Ambiental De São Paulo (PMA-SP). 2006. Tráfico de Animais Silvestres da Fauna Nacional – Dados estatísticos e estratégias operacionais 2001–2005; [accessed 2016 Aug 3]. [http://www.pea.org.br/educativo/relatorio\\_policia\\_ambiental.pdf](http://www.pea.org.br/educativo/relatorio_policia_ambiental.pdf)
- Silva NS. 2015. Espécimes recebidos no centro de triagem de animais silvestres de salvador/BA durante os anos de 2012 a 2014. Trabalho de Conclusão de Curso. Salvador: Universidade Federal da.
- Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol.* 30: 2725–2729.
- Universidade Federal De Viçosa (UFV). 2009. Museu de Zoologia João Moojen. Bicho da vez n° 4. Trinca-ferro-verdadeiro (*Saltator similis*); [accessed 2018 Dec 15]. <http://www.museudezoologia.ufv.br/bichoda-vez/edicao04.pdf>
- Zocche JJ, Viana IR. 2013. Avifauna apreendida no extremo sul catarinense: apreensões feitas durante oito anos de fiscalização e combate à captura de aves silvestres. *Revista Brasileira de Biociências. Porto Alegre.* 11:395–404.

### 1.6. Conclusão

O uso da tecnologia de NGS permitiu um enorme avanço na área de microssatélites, uma vez que tornou a descoberta destas regiões repetitivas muito mais eficiente e rápida quando comparada à metodologia tradicional. Além disto, ao analisar as sequências e não somente o tamanho do fragmento na genotipagem, é possível discriminar alelos de mesmo comprimento e também detectar SNVs (Butler 2004, Zavodna 2014). Este estudo foi pioneiro em desenvolver marcadores moleculares para *S. similis* baseado em NGS tanto para o desenvolvimento quanto para análise. Dos 59 marcadores microssatélites que foram selecionados, apenas dez *loci* chegaram na fase final (L15, L16, L20, L23, L29, L37, L38, L39, L49, L57), mostrando que apesar da facilidade que o NGS oferece no desenvolvimento de novos marcadores deste tipo, o desenvolvimento ainda requer várias etapas para padronização das PCR, avaliação da qualidade do sequenciamento (como pouca formação de dímeros e ausência de anelamento inespecífico em outras partes do DNA que não na sequência alvo) e descoberta de alvos que possuam graus de polimorfismo satisfatório e baixa frequência de alelos nulos.

Na padronização, as PCR foram testadas com diversas modificações, sendo que os melhores resultados foram obtidos quando utilizado o pré-mix IVB (Phoneutria), temperatura de anelamento de 60°C e 30 ciclos. A escolha de manter as mesmas condições para todas as reações permite que as reações que aqui foram feitas como pseudomultiplex sejam possivelmente realizadas como multiplex, reduzindo ainda mais os custos e tempo de manipulação. Além disto, a genotipagem poderá possivelmente ser realizada no mesmo laboratório que a PCR de sexagem, visando facilitar para o responsável pelo criatório que poderá enviar uma única vez a amostra em papel FTA.

A genotipagem dos *loci* aqui selecionados permitiu a correta identificação de possíveis progenitores mesmo que a análise de paternidade tenha sido realizada com um número inferior a dez *loci*. Isto demonstra que estes *loci* são extremamente úteis para este tipo de análise, e que, quando combinados com um número maior de *loci*, resultarão em ainda mais confiança dos resultados de filiação, e, portanto, estão aptos a compor o painel de genotipagem mencionada na resolução do CONAMA n° 487, de 15 de maio de 2018. Os *loci* que foram selecionados e verificados no teste de paternidade piloto foram dois dinucleotídeos (L15 e L16), quatro trinucleotídeos (L37, L38, L39 e L49) e um pentanucleotídeo (L57). Além disto, *primers* de regiões microssatélite também podem

funcionar em estudos com DNA de espécies proximamente relacionadas e não somente para aquela espécie para o qual foram desenhados, permitindo estudos multiespécies (Hodel 2016). Assim, os *primers* desenvolvidos neste trabalho para as regiões microssatélites de *Saltator similis* possuem potencial de funcionarem também para espécies próximas, como, por exemplo, para *Saltator maximus*, espécie também presente nos criatórios brasileiros.

Apesar do foco do trabalho ter sido desenvolver marcadores moleculares baseados em NGS para genotipagem e teste de filiação de *S. similis* visando a possível identificação de animais criados ilegalmente, estes mesmos marcadores poderão ser utilizados futuramente também em projetos de conservação, manejo em cativeiro e em estudos de genética de populações. Caso seja detectada estruturação genética entre os diferentes locais de ocorrência da espécie, a utilização destes marcadores também poderá auxiliar na descoberta de rotas de tráfico.

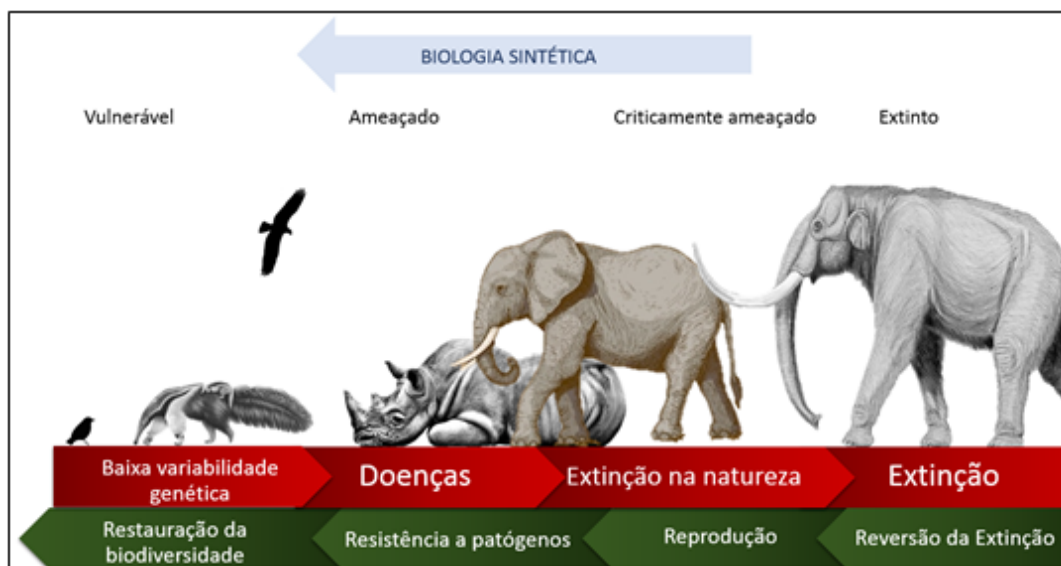
Este trabalho também foi pioneiro no sequenciamento, montagem e anotação do mtDNA completo de *S. similis*. Este poderá ser utilizado em estudos evolutivos e também possibilitará a identificação da espécie quando o animal não apresentar características morfológicas para a correta identificação, se tratar de ovos ou então apenas vestígios estiverem presentes e as informações apenas do COI não forem suficientes para distinguir espécies proximamente relacionadas.

## 2. Capítulo 2 - Biologia sintética: uma nova abordagem para a conservação de espécies ameaçadas de extinção

### 2.1 Introdução

A biologia sintética abrange o desenho e a construção de novas partes biológicas e sistemas, assim como a capacidade de alterar sistemas biológicos naturais com determinados fins (Nature portfolio 2021). A combinação de técnicas e tecnologias permite ler, interpretar, modificar, projetar e fabricar DNA para interferir em formas e funções celulares e de organismos, podendo atingir a escala de interferir em toda uma espécie ou ecossistema. Enquanto esta área avança, muito ainda é necessário aprender sobre quais genes influenciam quais características e como eles interagem com outros genes e também com fatores ambientais (Redford 2019).

Uma das abordagens da biologia sintética é em relação à conservação de espécies, que possui grande potencial já que projetos de conservação frequentemente se iniciam quando a espécie já se encontra em um nível crítico de ameaça e/ou possui poucos indivíduos, dificultando a chance de sucesso. A diversidade genética é um importante fator na sobrevivência da população (Bouzat 2010), sendo que a sua perda resulta em redução na qualidade de espermas, redução no número de filhotes por ninhada (Hedrick 2010), aumento da mortalidade de indivíduos jovens (Ralls 1988), aumento da susceptibilidade a doenças e infecções (Coltman 1999) e redução da habilidade de se adaptar a mudanças ambientais (Hoffmann 1997). Populações ameaçadas ou fragmentadas geralmente sofrem com endogamia, fluxo gênico restrito, deriva genética e pequeno tamanho populacional, contribuindo para a redução da diversidade genética e aumentando seu risco de extinção (Saccheri 1998, Madsen 1999, Butchart 2010, Market 2010, Rands 2010). Assim, o uso da biologia sintética, através principalmente da edição gênica, pode ser a chave para a eficiência dos projetos de conservação, uma vez que sejam tomadas ações para resolver problemas relacionados à patógenos, baixa diversidade genética, espécies invasoras, conversão de habitats, superexploração, poluição é até mesmo reversão da extinção (Piaggio 2016, Redford 2019) (figura 46).



**Figura 46- Biologia sintética na restauração da biodiversidade.** As pressões exercidas sobre os animais podem resultar na extinção de espécies (setas vermelhas). A utilização da biológica sintética nos projetos de conservação auxilia a manutenção das espécies, reduzindo o risco de extinção (setas verdes). Figura baseada em: <http://reviverestore.org/what-we-do/extinction-continuum/>.

### 2.1.1 De-extinction

Em 2013, surgiu um novo termo no campo da conservação de espécies: *de-extinction* (Revive & Restore 2013). Este termo ainda não possui tradução para o português, mas se trata da reversão da extinção de uma espécie. Em 2016, a IUCN (*International Union for Conservation of Nature*) definiu que *de-extinction* é a geração de uma espécie próxima à espécie extinta e funcionalmente equivalente, tendo como objetivo a restauração das funções ecológicas vitais que sustentam a dinâmica dos ecossistemas e aumentam a biodiversidade (IUCN 2016). Dentre as formas de se realizar a *de-extinction*, a hibridização precisa via edição gênica é a única aplicável a maioria das espécies alvo. Para isso, é necessário obter a sequência de DNA da espécie extinta e uma espécie modelo que poderá ter seu genoma editado para produzir a nova espécie-próxima com características da espécie extinta (Novak 2018).

Atualmente, já se tem a sequência de DNA de diversos animais extintos que poderiam ser candidatos aos programas de reversão de extinção. Porém, estes projetos são extensos e caros, exigindo investimento financeiro e equipes multidisciplinares. Assim, é preciso selecionar as espécies levando em consideração critérios técnicos e considerações ecológicas e sociais, tendo o foco principal em espécies que possuem alto impacto em seus ecossistemas (Novak 2018).



Em 2017, havia sete projetos de reversão de extinção de espécies com diferentes abordagens. Um destes projetos é a *de-extinction* do mamute-lanoso (*Mammuthus primigenius*), que utiliza o genoma de elefante-asiático (*Elephas maximus*) como organismo modelo (Novak 2018). *Mammuthus primigenius* é uma das espécies de mamute que existiu no Pleistoceno, cujos últimos registros encontrados datam de 12-10.000 anos atrás (The Academy of Natural Sciences 2007). Análises dos genótipos das duas espécies mostram que elas possuem cerca de um milhão de diferenças, mas que a maioria é apenas variação natural acumulada ao longo dos anos. Assim, para se produzir uma espécie próxima do mamute-lanoso, a edição gênica é focada nos genes que conferem características específicas de sobrevivência do mamute, como as adaptações para sobrevivência no frio (Palkopoulou 2015, Palkopoulou 2018, Lynch 2015). Ao se trabalhar em um projeto de reversão de extinção, comumente a espécie usada como modelo também recebe abordagem de conservação, que, como no caso do elefante-asiático, possui um projeto envolvendo o EEHV (herpesvírus endoteliotrófico de elefante).

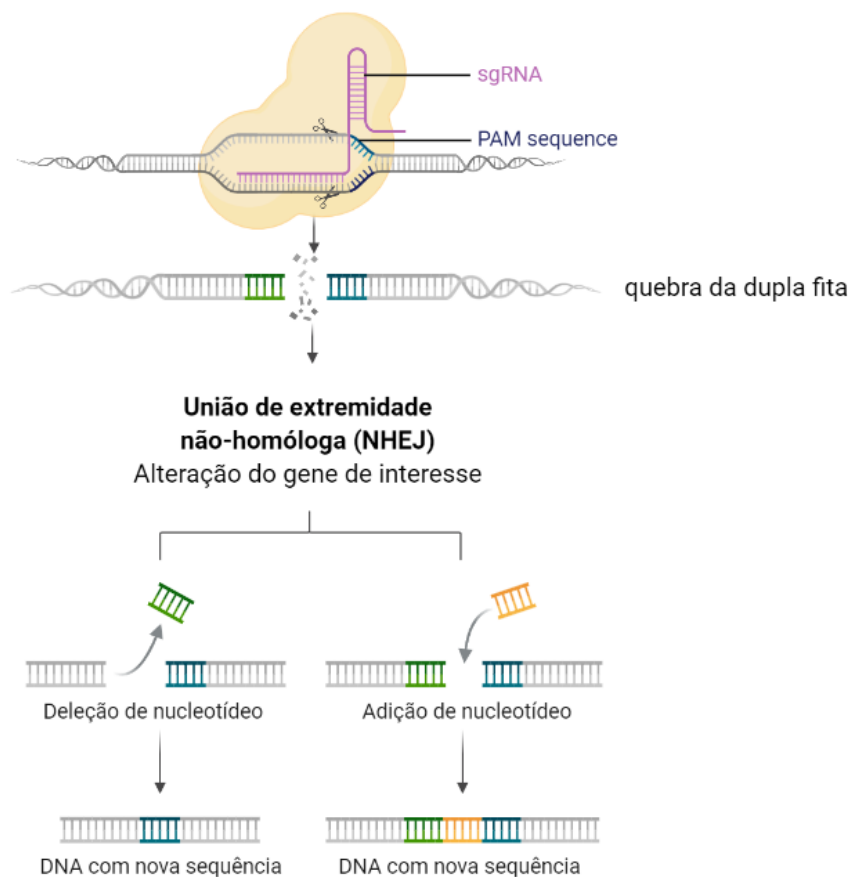
O mamute-lanoso foi uma das espécies escolhidas para reversão da extinção por ser uma espécie-chave no seu ecossistema, uma vez que é capaz de moldar seu habitat de modo a beneficiar outras espécies e disponibilidade de recursos. O caminhar da megafauna foi vantajoso para as gramíneas, e sua ausência resultou no crescimento de tundra e taiga (Putshkov 1999), sendo que as gramíneas sequestram e armazenam carbono de maneira mais eficaz (Amthor 1998), tendo importância nas mudanças climáticas. Além disto, a pastagem do mamute faz com que o ar frio do inverno chegue até o solo e o mantenha congelado durante o verão. Isso, juntamente com o albedo, previne o derretimento do permafrost, que é uma das maiores fontes de liberação de hidrocarbonetos na atmosfera. A restauração do ecossistema para pastagem não é uma solução rápida para se evitar e controlar as mudanças climáticas, mas oferece uma proteção a longo prazo (Amthor 1998). Para se criar um animal a partir de edições gênicas no genoma de elefante capaz de sobreviver ao frio, precisam ser encontradas no genoma do mamute os genes que permitiam sua sobrevivência neste ambiente, como crescimento de pelos, gordura subcutânea para isolamento e jejum e mecanismo de oxigenação a baixas temperaturas (Revive & Restore).

### 2.1.2 CRISPR-Cas9

Apesar de ter dado maior visibilidade à edição gênica, o CRISPR (repetições palindrômicas curtas agrupadas intercaladas regularmente, do inglês “*clustered regularly interspaced short palindromic repeats*”) não foi a primeira metodologia desenvolvida para este fim. Outras metodologias como TALEN (*transcription activator-like effector nucleases*) e *zinc-finger nucleases* já eram utilizadas há alguns anos, porém, apresentam certas dificuldades em relação à complexidade e possibilidade de realizar múltiplas edições, além do alto custo. Todos esses aspectos foram contornados com o CRISPR, permitindo uma nova era na edição gênica (Synthego, Ran 2013, Ledford 2015).

Descoberto inicialmente em uma fábrica de iogurtes e queijos, o sistema CRISPR funciona como uma espécie de sistema imunológico adaptativo em bactérias, que permite que elas lutem contra a repetidas infecções de certos bacteriófagos. Pouco tempo depois desta descoberta, grupos de cientistas foram capazes de utilizar essa metodologia para deletar, adicionar, ativar ou silenciar genes alvos em diversos animais de laboratório, demonstrando a ampla utilidade da técnica (Pennisi 2013), sendo uma ferramenta importante também para o estudo das funções gênicas (Nelson 2018).

Para alterar a sequência de nucleotídeos, é necessário um RNA-guia (sgRNA) que possui uma região complementar à sequência alvo de 20 à 25 pb seguida por uma sequência PAM (*protospacer adjacent motif*), que será variável de acordo com a espécie da qual o sistema CRISPR-Cas9 é derivado, sendo que no sistema com SpCas9 (*Streptococcus pyogenes* Cas9) o DNA alvo deve preceder um 5'-NGG PAM (Jinek 2012). Juntamente com a nuclease Cas9, esse complexo formado dentro da célula cliva separadamente as duas fitas de DNA. O sítio de clivagem situa-se três nucleotídeos acima ou abaixo da sequência PAM, dependendo da fita. Posteriormente ocorre a reparação pela união de extremidade não-homóloga (NHEJ, do inglês “non homologous end joining”), que pode resultar em mutações aleatórias de inserção ou deleção de nucleotídeos de diversos tamanhos, gerando mudança da janela de leitura e resultando em *knockout* do gene alvo (Vasconcelos 2015). O processo é exemplificado na figura 47, abaixo.



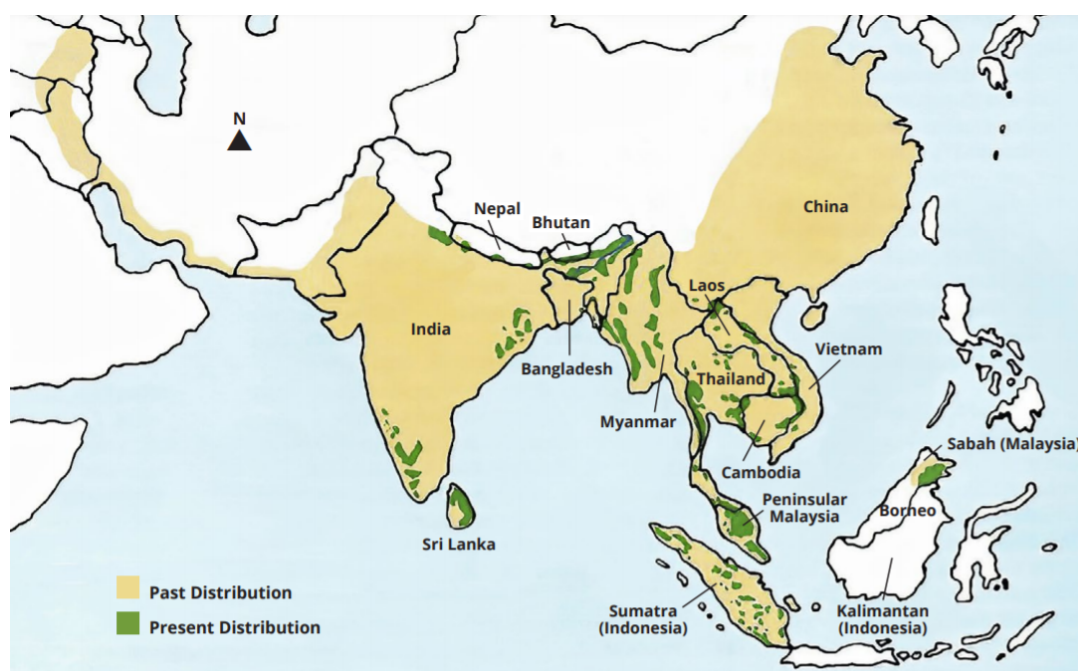
**Figura 47- Mecanismo de edição gênica com sistema CRISPR-Cas9.** Após a formação do complexo CRISPR-Cas9 com alinhamento do sgRNA à sequência alvo e reconhecimento do PAM, há a quebra da dupla fita de DNA que é então reparada por união de extremidade não homóloga, que pode ser feita com deleção ou adição de nucleotídeos, resultando em uma fita de DNA final com uma sequência diferente de nucleotídeos da fita de DNA original. Figura criada com BioRender.

### 2.1.3 *Elephas maximus*

Conhecido como elefante-asiático ou elefante-inidiano, o *Elephas maximus* é a única espécie ainda existente do gênero *Elephas* e é classificado como ameaçado de extinção na lista vermelha de espécies ameaçadas da IUCN (Williams 2020), enquanto que uma das suas três subespécies, o *Elephas maximus sumatranus*, é classificada como criticamente ameaçada pela mesma lista (Gopala 2011). Seu genoma contém 3.1 bilhões de pares de bases e é organizado em 56 cromossomos (Houck 2001). Atualmente, estima-se que há cerca de 50 mil indivíduos em vida livre e cerca de 15 mil indivíduos em cativeiro, sendo que seu tamanho populacional vem diminuindo (The IUCN SSC Asian Elephant Specialist Group). A espécie tem sua população severamente fragmentada e ocorre em diversos países: Bangladesh, Butão, Camboja, China, Índia, Indonésia, República Popular do Laos, Malásia, Mianmar, Nepal, Sri Lanka, Tailândia e Vietnã

(Williams 2020), sendo que já desapareceu de 95% da sua área de ocorrência histórica (Sukumar 2006) (figura 48).

Dentre as ameaças para sua conservação, está a perda de habitats e fragmentação, conflito com humanos, caça e tráfico (Sukumar 2006, Sukumar 2016, Menon 2017). A caça com interesse no marfim resulta em um desequilíbrio dos sexos, uma vez que atinge apenas machos, podendo resultar em aumento do número de acasalamentos consanguíneos (Sukumar 1998, Sukumar 2003). A caça não ocorre apenas por interesse no marfim, mas também por outras partes como carne e pele, afetando, neste caso, ambos os sexos (Kemf 2000, Menon 2002). Outro problema para a conservação da espécie, tanto para animais de vida livre como de cativeiro são as infecções causadas pelo Herpesvírus endoteliotrópico de elefante (EEHV).



**Figura 48- Área de ocorrência de *Elephas maximus*.** A região demarcada em amarelo representa as áreas de ocorrência do elefante-asiático no passado. As regiões em verde são áreas de ocorrência atuais (The IUCN SSC Asian Elephant Specialist Group).

#### 2.1.4 Herpesvírus endoteliotrópico de elefante (EEHV)

Há 19 cepas conhecidas do herpesvírus endoteliotrópico de elefante (EEHV), vírus de dupla fita DNA e pertencente à ordem Herpervirales, responsável por causar hemorragia aguda em elefantes asiáticos (*Elephas maximus*) filhotes e jovens, tanto de vida livre como de cativeiro, sendo a cepa EEHV1 a principal causadora da doença

(EEHV Advisory Group 2019). A doença causada por EEHV é um das as principais ameaças à sobrevivência ao elefante-asiático, já ameaçado de extinção, afetando cerca de 24% dos filhotes nascidos na América do Norte e na Europa nos últimos 20 anos e causando a morte de cerca de 60% dos elefantes-asiáticos monitorados com idades entre oito meses e 15 anos (Hayward 2012). Apesar da relevância do EEHV na conservação do elefante-asiático, ainda não se tem certeza sobre quais os tipos celulares são alvos do vírus (Guntawang 2020) ou a dinâmica da infecção (Fuery 2018, Srivorakul 2019).

Para poder conhecer melhor sobre o mecanismo de patogenicidade do vírus e resposta celular à infecção e também estudar sobre possíveis tratamentos, é necessário um modelo de cultura celular, que também serve para crescimento e propagação do vírus (Pringproa 2015, Meister 2019). Porém, há inúmeras tentativas fracassadas de se cultivar EEHV em diversas linhagens celulares como Vero (células oriundas de rim de macaco), MDCK (tipo celular oriundo de rim de cachorro), BHK-21 (células provenientes de rim de filhote de hamster) e HrT-18G (células de tumor retal humano), assim como em culturas primárias de diversos tipos celulares (Ossent 1990, Richman 1999, Seilern-Moy 2016, Pavulraj 2019). Essa dificuldade em se isolar e cultivar o vírus resulta na ausência de vacina, de tratamento específico e de protocolos de prevenção para a doença (Photichai 2020).

Em 2002, foi publicado pela primeira vez a síntese do material genético de um vírus sem a necessidade de um organismo molde: o vírus da pólio (Cello 2002), quebrando a necessidade de se precisar de um espécime viral isolado da natureza para se estudar aspectos relacionados aos vírus, iniciando um novo ramo da biologia sintética: a virologia sintética. Mesmo ainda estando nos primeiros anos da técnica e com apenas poucos vírus construídos com essa metodologia, essa é uma ferramenta importante para reconstruir vírus de difícil propagação e assim conseguir estudar mecanismos de patogenicidade e entendimento sobre a ligação patógeno-hospedeiro (Wimmer 2009, Wimmer 2011, Weber 2012, tenOever 2020). Apesar do avanço científico, a nova metodologia também levantou questionamentos éticos e de segurança, uma vez que tornou possível produzir qualquer vírus cuja sequência gênica esteja disponível em banco de dados ou até mesmo produzir novos vírus (Wimmer 2011).

## 2.2 Justificativa e relevância do projeto

Com a extinção de espécies ocorrendo de forma cada vez mais acelerada na sexta extinção em massa da história, novas abordagens são necessárias para proteção e conservação de espécies ameaçadas. A biologia sintética tem potencial inovador nesta área, podendo até mesmo criar espécies com sequências gênicas de mais de uma espécie e que mantenham a função ecológica de uma espécie extinta para manutenção do ecossistema no qual estavam inseridas. Apesar de focar em uma espécie, este é um projeto piloto que auxiliará no desenvolvimento das metodologias e futuramente poderá ser utilizado também para a conservação de outras espécies em risco de extinção.

## 2.3 Objetivo

Padronizar protocolo para estudo da estrutura nuclear de mamute-lanoso e realizar algumas etapas da edição gênica do projeto de *de-extinction* de mamute-lanoso e da construção sintética de EEHV.

### 2.3.1 Objetivos específicos

- Amplificar, extrair e purificar fragmentos de EEHV1 para futura montagem com virologia sintética;
- Desenhar sgRNA para *knockout* de genes de elefante;
- Realizar *knockout* de genes específicos com metodologia de CRISPR-Cas9 em cultura de células;
- Desenhar conjuntos de *primers* para avaliação da edição gênica;
- Padronizar protocolo de FISH em célula e tecido de elefante-asiático para posterior utilização em tecido de mamute-lanoso para melhor compreensão da organização nuclear da espécie e de como o DNA é preservado em tecido congelado antigo.

## 2.4 Metodologia

### 2.4.1 Fragmentos de EEHV

A sequência genética do EEHV1 sequenciada por colaboradores foi dividida em 20 fragmentos, sendo 18 fragmentos com ~10kb de tamanho e dois com ~5kb (fragmentos 1 e 20), mantendo uma sobreposição de nucleotídeos entre as sequências de 40pb que permitirá, no próximo passo, a união dos fragmentos com a utilização do GenBuilder™ DNA Assembly (GenScript). As sequências foram sintetizadas pela empresa GenScript em vetor pUC57. Para cada fragmento, foi desenhado um conjunto de *primers forward* e *reverse* para amplificação da sequência alvo. Foi realizado um gradiente de temperatura de anelamento para a amplificação, sendo as temperaturas utilizadas de: 56.9°C, 57.1°C, 57.8°C, 58.8°C, 60.1°C, 61.5°C, 63°C e 64.4°C. Também foram testados diferentes tempos de extensão (6 a 10 minutos) e diferentes números de ciclos (32 a 37 ciclos). Os reagentes utilizados na montagem da PCR são mostrados na tabela 18 e a programação do termociclador é mostrada na tabela 19, abaixo.

**Tabela 18- Reagentes utilizados para montagem das PCR.**

Reagente	Quantidade por reação (µl)
Água ultrapura	23
KAPA HiFi HotStart <i>Ready Mix</i> (2x) (KAPA biosystems)	25
Mix de <i>primers forward</i> e <i>reverse</i> (10pmol/µl)	1
DNA	1
<b>Volume total</b>	<b>50</b>

**Tabela 19-Programação do termociclador.** Os números destacados em negrito foram alterados em algumas reações durante a padronização

Temperatura (°C)	Tempo	
95	5 min	Desnaturação
98	30 seg	Desnaturação
<b>56.9-64.4</b>	30 seg	Anelamento
72	<b>10 min</b>	Extensão
		<b>37 ciclos</b>
72	10 min	Extensão final
4	∞	



Para visualização das amostras foi aplicado 20µl de produto de PCR no E-Gel 2% (Invitrogen) com corrente por 15 minutos com padrão de tamanho 1kb Plus DNA ladder (New England BioLabs) em equipamento E-Gel Power Snap System (Invitrogen). Os resultados dos géis foram visualizados com o mesmo equipamento. A banda alvo foi extraída do gel com o kit de purificação Zymoclean Large Fragment DNA Recovery Kit (Zymo Research) de acordo com o protocolo do fabricante e a amostra extraída foi quantificada em nanodrop.

#### 2.4.2 Edição gênica

##### 2.4.2.1 Desenho de sgRNA

Os sgRNA foram desenhados com auxílio da plataforma Synthego Knockout Design Guide (<https://design.synthego.com/#/>) para utilização com a enzima SpCas9 (*Streptococcus pyogenes* Cas9), que especificamente reconhece a sequência 5'-NGG-3' PAM (Guo 2019). Foi utilizado como referência o genoma de *Loxodonta africana* (RefSeq Loxafr3.0 (GCF\_000001905.1) rel101) disponível na plataforma. Posteriormente foi confirmado se não havia diferenças na sequência alvo no genoma de *Elephas maximus*. Os sgRNA desenhados foram sintetizados pela Synthego, que também fornece a enzima SpCas9.

##### 2.4.2.2 Desenho de primers

Os primers foram desenhados manualmente em regiões flangeadoras ao local da edição gênica utilizando as sequências dos genes disponíveis no banco de dados do National Center for Biotechnology Information. Os primers foram sintetizados pela empresa IDT Integrated DNA Technologies.

##### 2.4.2.3 Imortalização de células

Células mesenquimais de elefante-asiático foram cultivadas em placa de cultura de 6 poços com meio DMEM (Dulbecco's Modified Eagle Medium - Gibco) suplementado com 15% de FBS (soro fetal bovino). Para a imortalização, as células receberam 2 ml do mesmo meio de cultura suplementado com 2 µl de sulfato de protamina 8mg/ml (diluído em água e filtrado), e foram incubadas com 10 µl de EF1α-hSTEMCCA(OKSM) Lentivirus (Millipore). Após dois dias da incubação com o lentivírus, o meio de cultura foi removido e as células lavadas vagarosamente com PBS e depois adicionado novo meio de cultura suplementado com 15% FBS e Gibco MEM

aminoácidos não-essenciais, e foram mantidas em estufa a 37°C em atmosfera de 5% de CO<sub>2</sub>.

#### 2.4.2.4 CRISPR-Cas9

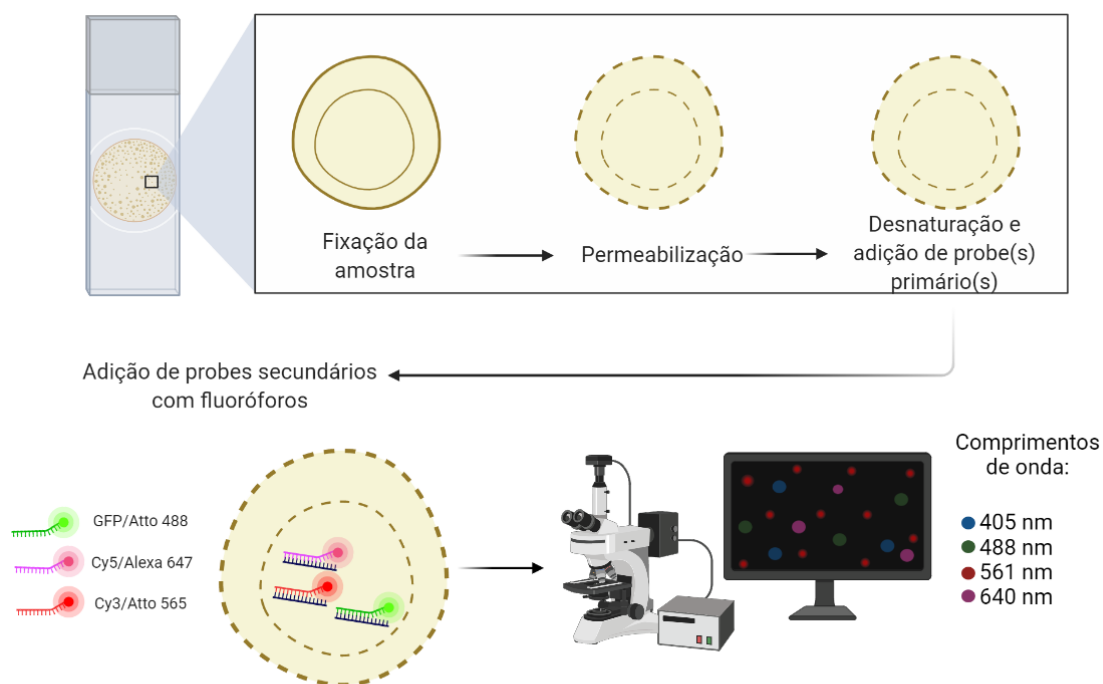
Células mesenquimais de elefante asiático imortalizadas no laboratório foram cultivadas em meio de cultura DMEM (Dulbecco's Modified Eagle Medium - Gibco), com 15% de soro fetal bovino (Gibco), antibiótico-antimicótico (Gibco), e Gibco MEM aminoácidos não-essenciais (para aumentar o crescimento celular e viabilidade) e mantidas em estufa a 37°C em atmosfera de 5% de CO<sub>2</sub>. Dez dias antes do experimento, foi adicionado TSA (*Trichostatin A*) (STEMCELL™ Technologies) às células de acordo com as especificações do fabricante. Um dia antes da transfecção, as células receberam novo meio de cultura DMEM (Gibco) com 15% de soro fetal bovino, sem antibióticos e com TSA. A transfecção foi realizada com Amaxa™ P3 Primary Cell 4D-Nucleofactor™ X kit (Lonza) de acordo com o protocolo do fornecedor, utilizando seleção de tipo celular hES cell- H9 e pulsação CB 150. As células de uma garrafa de cultura T75 (75 cm<sup>2</sup> de área de crescimento) com cerca de 75% de confluência foi utilizada com 40 µl de mix dos sgRNA, 2 µl de SpCas9 (Synthego) e 4 µl de GFP (Green Fluorescent Protein – fornecido pelo kit Lonza), utilizado para avaliar o grau de transfecção das células. Após o procedimento, as células foram mantidas em meio DMEM (Gibco) com adição de TSA e de fator de crescimento bFGF (Fibroblast Growth Factor-basic – Millipore) de acordo com especificação do fabricante.

#### 2.4.3 Fluorescence in situ hybridization (FISH)

*Fluorescence in situ hybridization* (FISH) é uma técnica utilizada para detecção e localização de uma sequência específica de DNA no cromossomo, sendo uma poderosa ferramenta para estudar a estrutura e organização nuclear. A técnica de FISH utilizada neste trabalho foi adaptada do protocolo de Beliveau (2015) (figura 49), com diferenças no tempo de permeabilização da amostra (1 hora) e tentativas com diferentes períodos (3, 5, 10 e 20 minutos) e temperaturas de desnaturação (80°C e 95°C). Com o intuito de se reduzir o *background* das amostras de tecido, também foi adicionada a etapa de adição de TrueBlack (Biotium) no final do protocolo e antes da adição de DAPI. A solução estoque de TrueBlack foi diluída 50X em etanol 80% no momento da utilização, sendo posteriormente adicionado 100 µl da solução de uso à cada lâmina por 30-60 segundos, seguido de 10 lavagens de 10 minutos com PBS ou PBT (*Phosphate Buffered Solution/0.1% Tween*). Como o PBT pode resultar em *background* do tecido, apenas a

segunda lavagem foi realizada com PBT. Outra etapa testada para reduzir o *background* da amostra foi a utilização de Scale U2 *Buffer* (0,1M Triton, 4M uréia e 30% glicerol) por diferentes períodos em temperatura ambiente. Todas as soluções utilizadas foram preparadas no dia dos experimentos.

Na montagem das lâminas foi utilizado o DAPI SlowFade™ Gold Antifade Mountant (ThermoFisher), e, em alguns casos, também o YOYO-3 Iodide (ThermoFisher) e TO-PRO-3 Iodide (ThermoFisher).



**Figura 49- Metodologia de FISH.** Desenho esquemático da metodologia de FISH mostrando de forma resumida os principais pontos do protocolo, que são a fixação e permeabilização da amostra, seguido por desnaturação e adição dos *probes* primários, que são hibridizados com o DNA da amostra. Depois é adicionado à lâmina os *probes* secundários marcados com fluoróforos e então são visualizados com microscopia confocal utilizando diversos comprimentos de onda. A figura foi criada com BioRender.

#### 2.4.3.1 Preparação das lâminas

Lâminas de tecido foram montadas pela equipe do Beth Israel Deaconess Medical Center após fixação em solução de 4% de paraformaldeído. Os cortes do tecido em parafina foram realizados a cada 5µm. Para a preparação de lâminas de células, lâminas novas foram esterilizadas com álcool 70%, às quais foi adicionado 200 µl de células em meio de cultura (retirado de um total de 5ml de células em suspensão de uma garrafa de cultura T75 com 70-80% de confluência). As lâminas foram colocadas em placas de petri estéreis com algodão embebido em água estéril em incubadora a 37°C em atmosfera de

5% de CO<sub>2</sub> por 3 horas. Em seguida, foram retiradas da estufa e imersas em solução de PBS (*phosphate buffered saline*) com 4% de paraformaldeído por 30 minutos seguido de lavagem com PBS. As lâminas de células foram armazenadas a 4°C em solução de PBS. Foram utilizadas células mesenquimais de elefante-asiático e também células humanas (PGP1 iPSCs) para controle negativo.

Previamente ao protocolo de FISH, tanto lâminas de células como de tecidos foram fixadas novamente por 30 minutos em solução de PBS com 4% de paraformaldeído em temperatura ambiente e depois lavadas com PBS 1x.

#### 2.4.3.2 Probes

Os *probes* foram desenhados por colaboradores e sintetizados pela empresa IDT technologies e amplificados de acordo com o protocolo “*probe amplification using T7*”, disponível em <https://nirlab.org/2015/09/30/t7-reaction/>. Estes *probes* foram desenhados no genoma de elefante-asiático tendo como alvo a região dos centrômeros, cromossomo 11 e cromossomo 24. Os *probes* secundários marcados com fluorescência foram sintetizados pela mesma empresa e possuíam as seguintes sequências: /5Alex647N/CACCGACGTCGCATAGAACGGAAGAGCGTGTG/3AlexF647N/ (secundário 1), /5Atto488N/TAGCGCAGGAGGTCCACGACGTGCAAGGGTGT /3Atto488N/ (secundário 5), e /5Atto565N/CACACGCTCTCCGTCTTGGCCGTGGTCGATCA/3Atto565N/ (secundário 6). Essas sequências determinam que o *probe* secundário 1 será visível quando utilizado o laser com comprimento de onda 640 nm, que possui cor rosa, enquanto que o *probe* secundário 5 será visualizado com comprimento de onda 488 nm, que possui cor verde, e o *probe* secundário 6 no comprimento de onda 561 nm, com cor vermelha. Também foi utilizado *probes* secundários com Cy5, que também pode ser visualizado com o comprimento de onda de 640 nm.

#### 2.4.3.3 Imunofluorescência

A imunofluorescência foi realizada de acordo com o protocolo “IF protocol following FISH” disponível em <https://nirlab.org/2019/06/06/if-protocol-following-fish/>, utilizando apenas o anticorpo primário rabbit anti-lamin B1 (abcam) e secundário anti-rabbit Cy5 (rosa) ou anti-rabbit pacific blue (invitrogen), de cor azulada.

#### 2.4.3.4 Obtenção de imagens

O microscópio foi calibrado com Tetraspeck Microspheres (ThermoFisher), preparado de acordo com o protocolo “Protocol for Tetraspeck Calibration Slide

Preparation used for generating Biplane PSF” disponível em <https://nirlab.org/2017/01/25/protocol-for-tetraspeck-calibration-slide-preparation-used-for-generating-biplane-psf/>. As imagens foram obtidas utilizando o microscópio confocal Nikon Eclipse Ti do Wyss Institute com aumento de 60x com lente objetiva com imersão em água e uso do *software* Metamorph 7.10.2.240. As imagens foram tratadas com o *software* Fiji (Schindelin 2012), NIS-Elements *software* (Nikon instruments Inc.) e IMARIS *software* (Oxford Instruments). Os comprimentos de onda utilizados para captura das imagens foram 405 (DAPI - azul), 488 (GFP - verde), 561 (dsRed /Cy3 – vermelho) e 640 nm (Cy5 – rosa).

## 2.5 Resultados e discussão

### 2.5.1 Edição gênica

Os genes para *knockout* foram selecionados por colaboradores após análise comparativa dos genomas de mamute-lanoso e elefante asiático. Foram desenhados 31 sgRNA de 20pb para *knockout* de genes em cultura de células de elefante utilizando a metodologia de CRISPR-Cas9 (alguns genes encontram-se exemplificados na tabela 20), levando em consideração a menor chance de *off-target* e estar no começo da região codificante. A grande maioria foi desenhada para que o corte da dupla fita ocorra do exon 2 (18 sgRNA), seguido por sete sgRNA desenhados no exon 1, quatro no exon3, um no exon 5 e um no exon 10.

Apenas um experimento utilizando todos os sgRNA sintetizados como *multiplex* foi realizado, porém, após dois dias, a grande maioria das células estava morta (incluindo grupo controle sem edição gênica), possivelmente devido à presença de TSA nas células por muito tempo, uma vez que também é indutor de apoptose (Schnichels 2012). O experimento será repetido por colaboradores da mesma maneira, porém, as células serão expostas ao TSA somente dois dias antes da edição gênica. O TSA inibe seletivamente algumas classes de histonas desacetilases, resultando em alteração na expressão gênica, e a metodologia de CRISPR funciona melhor em sítios ativos, e, portanto, é importante a tentativa de se manter as células com TSA.

**Tabela 20- Complexo CRISPR-Cas9 para *knockout* dos genes selecionados.** À esquerda da tabela encontra-se a abreviação e o nome do gene alvo, seguida da informação do exon na qual a sequência alvo está inserida. A sequência do RNA guia está representada em verde, em negrito a sequência alvo, em azul a região PAM e as setas amarelas indicam os locais onde as fitas de DNA serão cortadas. Imagens produzidas com Synthego Knockout Design Guide (<https://design.synthego.com/#/>).

Gene alvo	Exon	Complexo CRISPR-Cas9
Nome do gene alvo disponível sob consulta	2	Sequência do sgRNA desenhado disponível sob consulta
Nome do gene alvo disponível sob consulta	2	Sequência do sgRNA desenhado disponível sob consulta
Nome do gene alvo disponível sob consulta	2	Sequência do sgRNA desenhado disponível sob consulta
Nome do gene alvo disponível sob consulta	2	Sequência do sgRNA desenhado disponível sob consulta

Nome do gene alvo disponível sob consulta	2	Sequência do sgRNA desenhado disponível sob consulta
Nome do gene alvo disponível sob consulta	2	Sequência do sgRNA desenhado disponível sob consulta
Nome do gene alvo disponível sob consulta	3	Sequência do sgRNA desenhado disponível sob consulta
Nome do gene alvo disponível sob consulta	3	Sequência do sgRNA desenhado disponível sob consulta

Para cada um dos sgRNA, também foram desenhados manualmente conjuntos de *primers* (exemplificados na tabela 21) para que, após a edição gênica, o DNA seja extraído para ser utilizado como molde para posterior PCR e sequenciamento para verificação da edição gênica. Os *primers* desenhados possuem de 21 à 30pb de tamanho e os tamanhos do *amplicons* esperados são de 121 à 310pb.

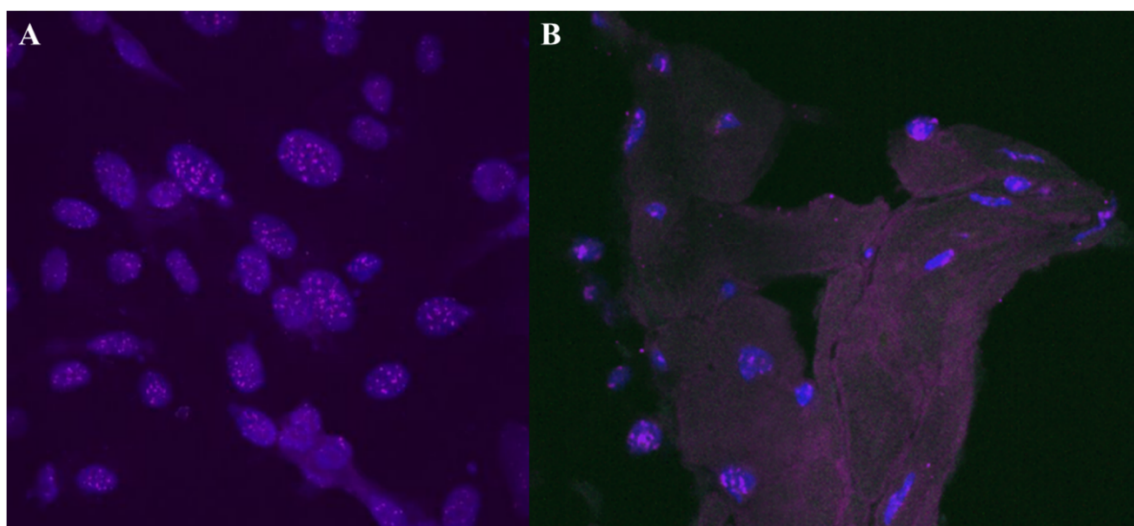
**Tabela 21- Conjuntos de primers desenhados para amplificação após *knockout* para sequenciamento.**

<i>primer F 5'→3'</i>	Tamanho <i>primer F</i> (pb)	Tamanho <i>primer R</i> (pb)	<i>primer R 5'→3'</i>	Tamanho do <i>amplicon</i> esperado (pb)
Sequência disponível sob consulta	28	24	Sequência disponível sob consulta	199
Sequência disponível sob consulta	23	26	Sequência disponível sob consulta	171
Sequência disponível sob consulta	27	29	Sequência disponível sob consulta	288
Sequência disponível sob consulta	30	24	Sequência disponível sob consulta	249
Sequência disponível sob consulta	24	29	Sequência disponível sob consulta	161
Sequência disponível sob consulta	25	26	Sequência disponível sob consulta	153
Sequência disponível sob consulta	26	27	Sequência disponível sob consulta	217
Sequência disponível sob consulta	23	22	Sequência disponível sob consulta	200

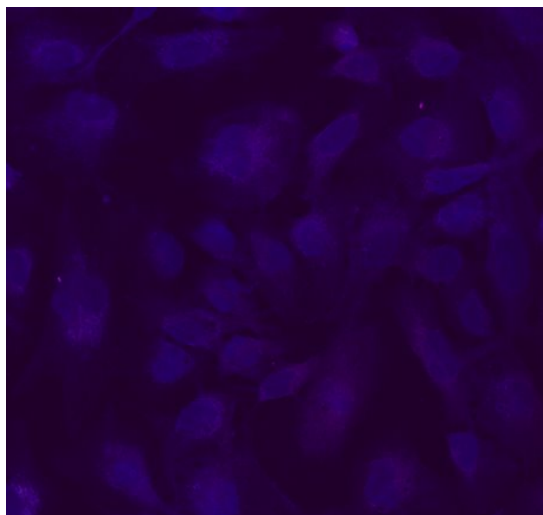


### 2.5.2 Padronização do protocolo de FISH

Com o intuito de se reduzir o uso de lâminas com tecido de mamute-lanoso devido à sua quantidade limitada, todos os protocolos foram primeiro testados em lâminas com células mesenquimais ou fígado de elefante-asiático. Primeiramente foi realizado o protocolo padrão de FISH em células mesenquimais de elefante-asiático para testagem do protocolo e funcionamento dos *probes* desenhados para os centrômeros, utilizando marcação rosa (Cy5) para os centrômeros e finalização com DAPI, que teve resultado satisfatório, conforme mostrado na figura 50-A. Na imagem, é possível observar os núcleos corados em azul e dentro deles diversos pontos rosados, que representam os centrômeros do elefante-asiático, que possui 56 cromossomos. O mesmo protocolo foi realizado em lâmina com fígado de elefante-asiático e apresentou resultado similar, conforme mostrado na figura 50-B, abaixo. No tecido, é esperado observar menos pontos rosados, uma vez que o processo de corte do tecido em lâminas pode resultar que nem todos os centrômeros estejam no mesmo plano. Como controle negativo, os mesmos procedimentos foram realizados em uma lâmina contendo células humanas, na qual não se observa a coloração de nenhum centrômero, apenas núcleos corados em azul (figura 51).



**Figura 50- Centrômeros de elefante-asiático.** Após a metodologia de FISH utilizando Cy5 (rosa) para os centrômeros de elefante-asiático e DAPI (azul) para o núcleo, as imagens foram obtidas utilizando o microscópio confocal Nikon Eclipse Ti com aumento de 60x com lente objetiva com imersão em água e uso do *software* Metamorph 7.10.2.240. As imagens foram tratadas com o *software* Fiji (Schindelin 2012) e são sobreposições dos lasers de comprimento de onda de 405, 488, 561 e 640 nm. A - células mesenquimais. B - fígado.

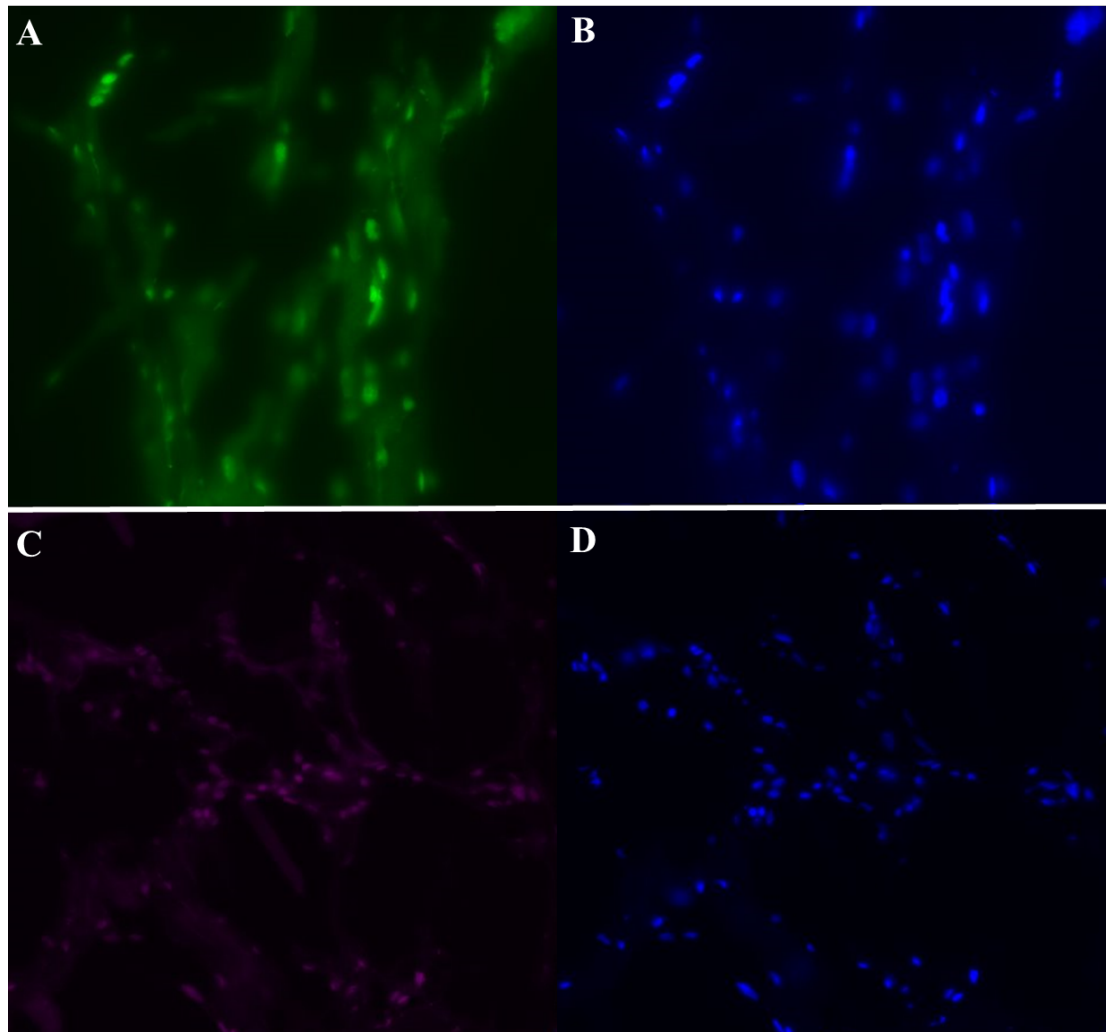


**Figura 51- Controle negativo (células humanas).** Resultado de FISH em células humanas (PGP1 iPSCs) utilizando Cy5 (rosa) para os centrômeros de elefante-asiático e DAPI (azul) para o núcleo. A imagem foi obtida utilizando o microscópio confocal Nikon Eclipse Ti com aumento de 60x com lente objetiva com imersão em água e uso do *software* Metamorph 7.10.2.240. A imagem foi tratada com o *software* Fiji (Schindelin 2012). Imagem com sobreposição dos lasers de comprimento de onda de 405, 488, 561 e 640 nm.

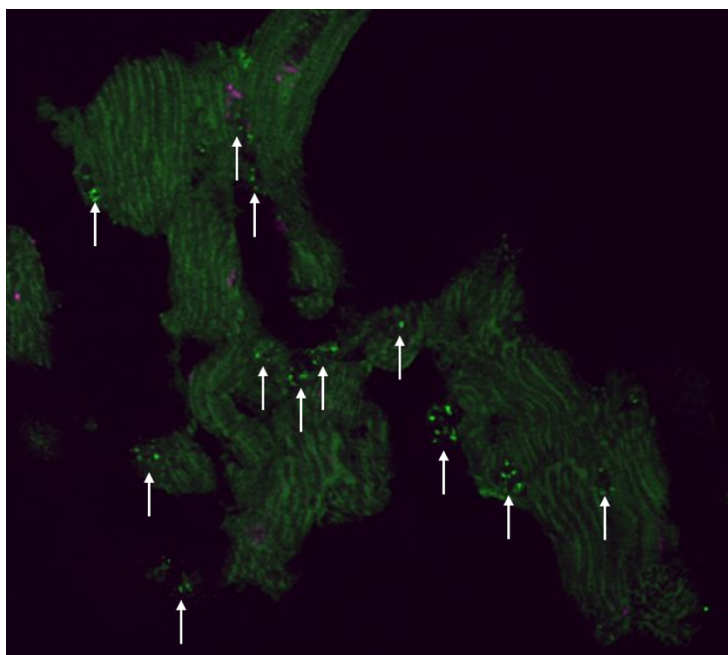
Como o sinal de alguns *probes* com alguns fluoróforos é menos intenso, é interessante a possibilidade de se fazer um “jogo de cores”, visando selecionar a melhor combinação de *probes* a serem utilizados e suas respectivas cores. Dentre os experimentos realizados por nós, a coloração em azul sempre apresentou bom resultado quanto à intensidade, então seria interessante que este pudesse ser utilizado em um *probe* em vez de utilizá-lo na coloração do núcleo com o DAPI. Assim, foram testados também TO-PRO e YO-YO, que colorem o núcleo de rosa e verde, respectivamente. O protocolo de FISH foi então realizado sem a utilização de *probes*, visando apenas a etapa final de montagem da lâmina, na qual TO-PRO ou YO-YO foram diluídos em DAPI. Ambos foram capazes de colorir os núcleos também, com ligeiramente menos núcleos corados do que DAPI e com mais *background* (figura 52).

Em seguida, foi realizado FISH em fígado de elefante-asiático com *probes* de centrômero, porém, desta vez marcados com fluoróforo verde e TO-PRO diluído em água, sem DAPI. Neste, foi possível observar regiões que apresentavam aspecto de núcleo por apresentarem marcação dos centrômeros (pontos verdes), porém, estes não apresentavam coloração quando visualizados com o comprimento de onda de 640 nm (imagem 53). Após visualização no microscópio, esta mesma lâmina teve sua lamínula removida, foi lavada uma vez com PBS e então foi adicionado TO-PRO novamente, porém, desta vez diluído em DAPI. A figura 54 mostra que então foi possível observar os núcleos corados

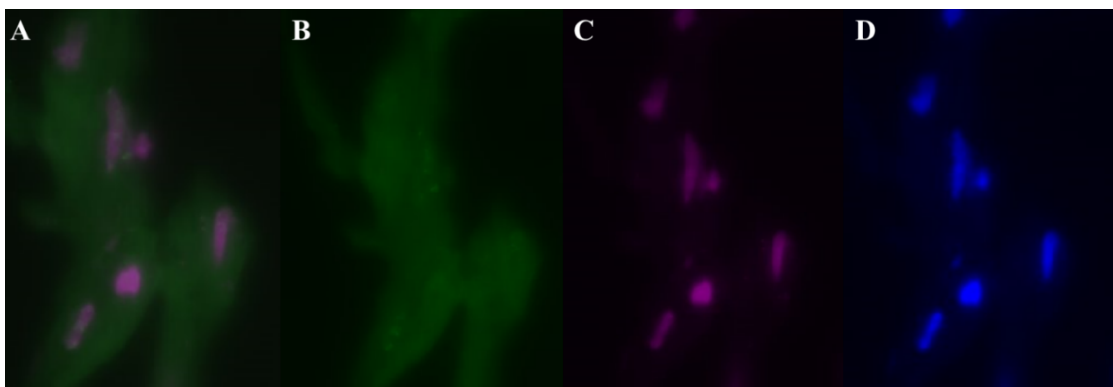
tanto em rosa quanto azul, além dos centrômeros em verde, como anteriormente. O mesmo fato de o TO-PRO só corar o núcleo quando diluído em DAPI e não corar quando diluído em água também ocorreu quando realizado o FISH em células (figura 55). Resultado similar também foi encontrado quando utilizado o YO-YO, que também só corou os núcleos tanto em células como em tecido quando diluído em DAPI e não quando diluído em água. Assim, as duas outras alternativas ao DAPI aqui experimentadas não se mostraram viáveis.



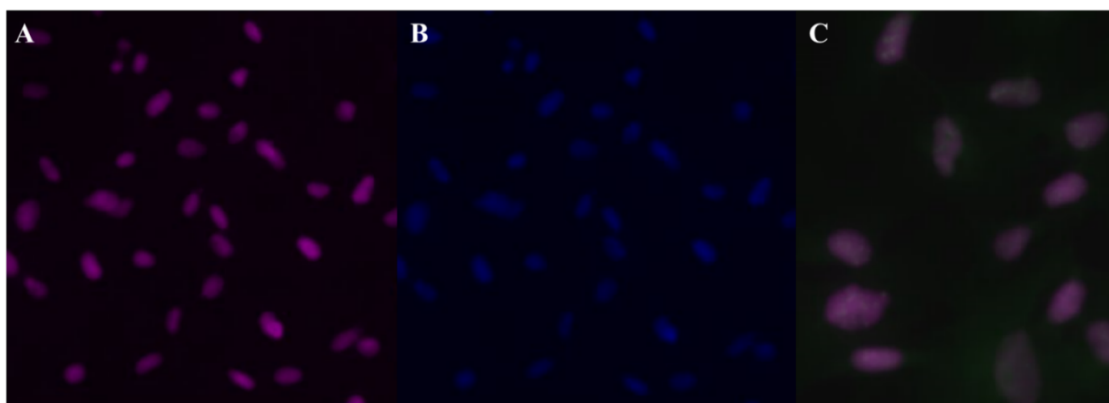
**Figura 52- Marcação de núcleo em tecido com diferentes cores.** Apenas os marcadores de núcleo foram adicionados nas lâminas contendo fígado de elefante-asiático após o protocolo de FISH sem utilização de *probes*. As imagens foram obtidas utilizando o microscópio confocal Nikon Eclipse Ti com aumento de 60x com lente objetiva com imersão em água e uso do *software* Metamorph 7.10.2.240. As imagens foram tratadas com o *software* Fiji (Schindelin 2012). A e B – lâmina de fígado com YO-YO diluído em DAPI, sendo que A é a imagem gerada somente com o comprimento de onda de 488 nm e B com 405 nm. Nota-se que em B há alguns núcleos a mais marcados do que em A e menos *background*. C e D – lâmina de fígado com TO-PRO diluído em DAPI, sendo que C é a imagem gerada somente com comprimento de onda de 640nm e em D com 405 nm. Na imagem D há alguns núcleos a mais marcados do que na imagem C e menos *background*.



**Figura 53- Centrômeros e TO-PRO diluído em água em fígado de elefante-asiático.** Fígado de elefante-asiático com FISH de centrômeros marcados com a cor verde (atto 488) e lâmina finalizada com TO-PRO (rosa) diluído em água. As setas brancas indicam regiões com aparência de núcleo, porém nenhuma delas foi corada com TO-PRO. A imagem foi obtida utilizando o microscópio confocal Nikon Eclipse Ti com aumento de 60x com lente objetiva com imersão em água e uso do *software* Metamorph 7.10.2.240. A imagem foi tratada com o *software* Fiji (Schindelin 2012) e tem sobreposição dos lasers de comprimento de onda de 488 e 640 nm.

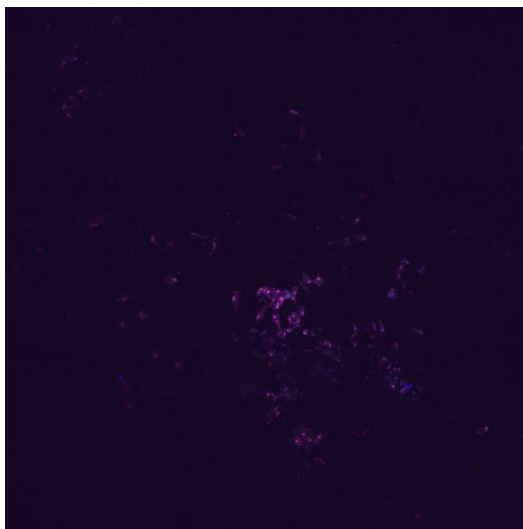


**Figura 54- Centrômeros e TO-PRO diluído em DAPI em fígado de elefante-asiático.** Após capturar a imagem anterior, a mesma lâmina de fígado de elefante-asiático com FISH com *probes* para centrômeros (atto 488) teve sua lamínula removida, foi lavada com PBS 1x e então foi adicionado TO-PRO novamente, mas desta vez diluído em DAPI. As imagens foram obtidas utilizando o microscópio confocal Nikon Eclipse Ti com aumento de 60x com lente objetiva com imersão em água e uso do *software* Metamorph 7.10.2.240 e tratadas com o *software* Fiji (Schindelin 2012). A – sobreposição das imagens com comprimentos de onda 488 e 640nm, sendo possível observar a marcação dos núcleos com TO-PRO (rosa) e também centrômeros em verde. B - imagem utilizando apenas o comprimento de onda 488 nm, sendo possível observar os centrômeros. C - imagem utilizando apenas o comprimento de onda 640 nm com os núcleos corados com TO-PRO. D - imagem utilizando apenas o comprimento de onda 405 nm com os mesmos núcleos da figura C corados também com DAPI.



**Figura 55- Centrômeros e TO-PRO diluído em DAPI em células de elefante-asiático.** Células mesenquimais de elefante-asiático com FISH de centrômeros marcados com a cor verde (atto 488) e lâmina finalizada com TO-PRO (rosa) diluído em DAPI. Os mesmos núcleos marcados na imagem A com o comprimento de onda de 640 nm aparecem marcados na imagem B, com comprimento de onda de 405 nm. A imagem C tem sobreposição dos lasers de comprimento de onda de 488 e 640 nm, sendo possível visualizar o núcleo (em rosa) e os centrômeros (verde).

Um dos fatores que dificulta a realização de FISH em amostras de mamute-lanoso é o estado de preservação da amostra. Em uma tentativa de se simular um tecido mais degradado, foi adicionado pepsina na lâmina com fígado de elefante-asiático por 10, 20 ou 30 minutos a 37°C antes do início do protocolo de FISH, porém, este passo resultou na destruição do tecido com todos os períodos testados (figura 56).



**Figura 56-FISH em fígado de elefante-asiático após pepsina.** Em uma tentativa de se reduzir o *background* do tecido, foi aplicado pepsina no tecido por 30 minutos a 37°C antes do início do protocolo de FISH, porém, este passo resultou na destruição do tecido. A imagem foi obtida utilizando o microscópio confocal Nikon Eclipse Ti com aumento de 60x com lente objetiva com imersão em água e uso do *software* Metamorph 7.10.2.240. A imagem foi tratada com o *software* Fiji (Schindelin 2012) e é uma sobreposição das imagens com lasers de comprimento de onda de 405, 488, 561 e 640 nm.

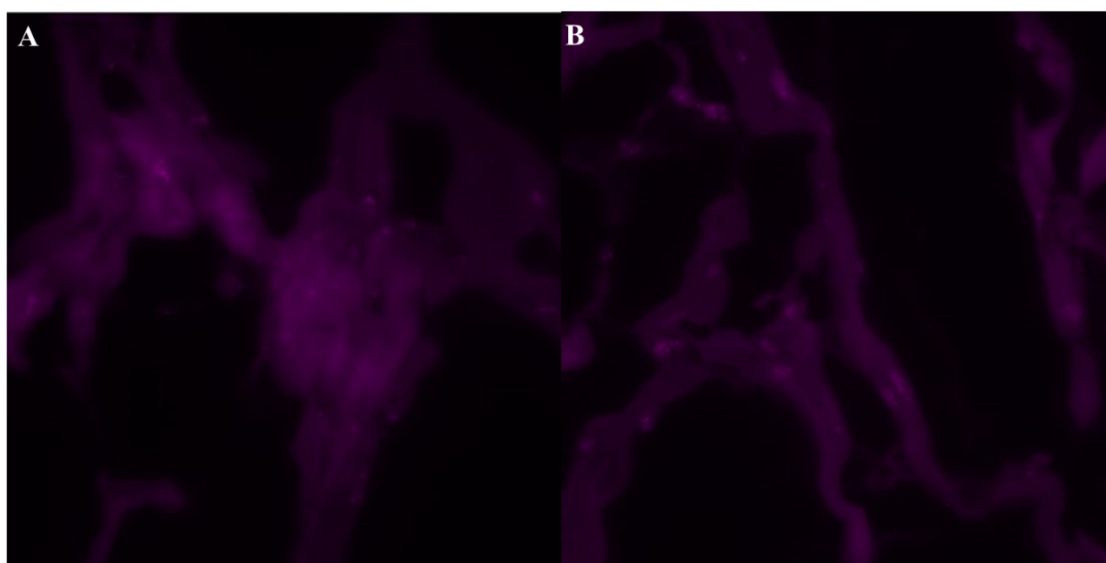
Outro problema encontrado quando da realização de FISH em lâminas de tecido e de osso de mamute-lanoso é o excesso de *background*, assim, foram experimentadas algumas etapas para se reduzir o sinal do *background*. Uma das formas foi adicionar a etapa de TrueBlack (Biotium) diluído em álcool 80% por 30 segundos a um minuto ao final do protocolo de FISH. O TrueBlack é utilizado para reduzir a fluorescência da lipofuscina, presente em tecidos mais envelhecidos. Após o uso desta solução, é preciso realizar diversas lavagens para remover o produto. Quando a lavagem era realizada com PBT (Phosphate *Buffered* Solution/0.1% Tween), a remoção do produto era mais eficiente, porém, o uso de Tween pode resultar em mais *background*. Assim, ficou padronizado a adição de TrueBlack por 45 segundos seguido de uma lavagem em PBS de 10 minutos, uma lavagem em PBT de 10 minutos e mais 8 lavagens de 10 minutos cada em PBS. Conforme pode ser visto na figura 57, este procedimento se mostrou eficaz para reduzir o *background* do tecido.

Outras tentativas de se reduzir o *background* do tecido foi adicionar uma etapa de 5 minutos com Tris 100mM a temperatura ambiente ou então adicionar glicina 20mM por 30 minutos a temperatura ambiente no início do protocolo de FISH, logo após a etapa de fixação do tecido, porém, ambos resultaram em formação de sinal até mesmo fora da região do tecido (figura 58). Uma etapa já realizada no protocolo de FISH para redução do *background* é a imersão da lâmina por 30 minutos em temperatura ambiente em *Scale*

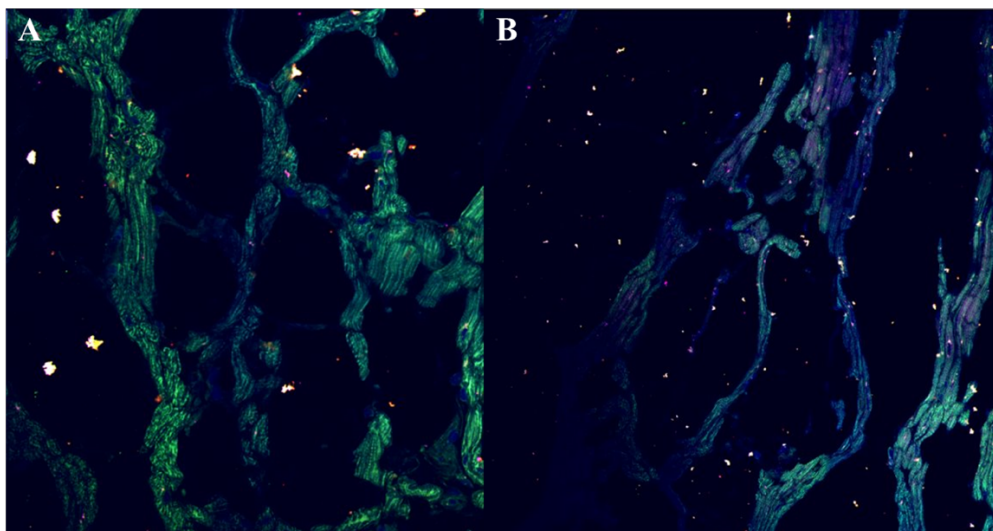


U2 *Buffer* logo após a adição do *probe* secundário. Foram realizados experimentos com tempos maiores de incubação (1 hora, 2 horas, 1 dia e 2 dias), mas não houve diferença significativa, optando-se, então, por manter essa etapa por 30 minutos.

Outras alterações testadas no protocolo de FISH foram mudanças na temperatura de hibridização (95°C) conforme sugerido por Beliveau (2014). Também foi experimentado mais tempo na etapa de desnaturação que o protocolo original, mas nestes casos o protocolo de FISH não funcionou.



**Figura 57-Centrômeros em fígado de elefante-asiático com e sem TrueBlack.** Ao final do protocolo de FISH, foi adicionado a etapa do TrueBlack seguido por 10 lavagens anteriormente à montagem da lâmina com DAPI. O TrueBlack (Biotium) é utilizado para reduzir a fluorescência da lipofuscina, presente em tecidos mais envelhecidos, e nota-se que ele foi eficaz em reduzir o *background*. A - sem TrueBlack. B – com TrueBlack. As imagens foram obtidas utilizando o microscópio confocal Nikon Eclipse Ti com aumento de 60x com lente objetiva com imersão em água e uso do *software* Metamorph 7.10.2.240 e tratadas com o *software* Fiji (Schindelin 2012). As imagens mostradas foram capturadas com o comprimento de onda de 640 nm, e é possível observar várias regiões com sinal mais forte composto por vários pontos referente aos *probes* de centrômeros marcados em rosa.

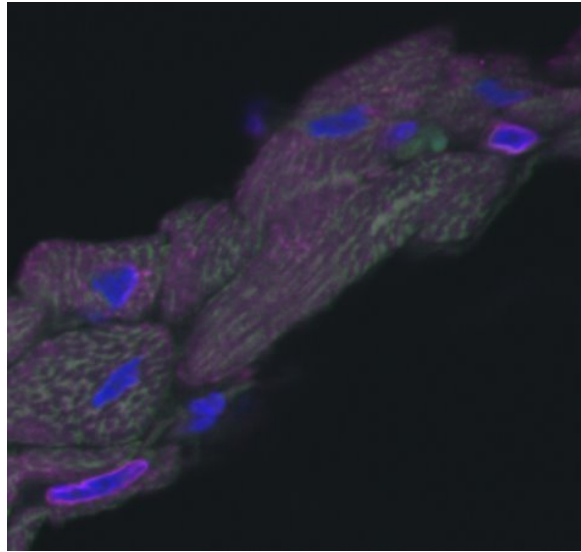


**Figura 58- FISH em fígado de elefante-asiático após glicina e tris.** Em uma tentativa de se reduzir o *background* do tecido, foi adicionado, após a etapa de fixação do tecido, Tris 100mM por 5 minutos a temperatura ambiente (imagem A) ou Glicina 20mM por 30 minutos a temperatura ambiente (imagem B), seguido pelo protocolo de FISH utilizando *probes* para centrômeros (atto 488 e alex 640). Nos dois casos, a adição dessa etapa inicial resultou em formação de sinal inclusive fora da região do tecido. As imagens foram obtidas utilizando o microscópio confocal Nikon Eclipse Ti com aumento de 60x com lente objetiva com imersão em água e uso do *software* Metamorph 7.10.2.240. As imagens foram tratadas com o *software* Fiji (Schindelin 2012) e mostram sobreposições dos lasers de comprimento de onda de 405, 488, 561 e 640 nm.

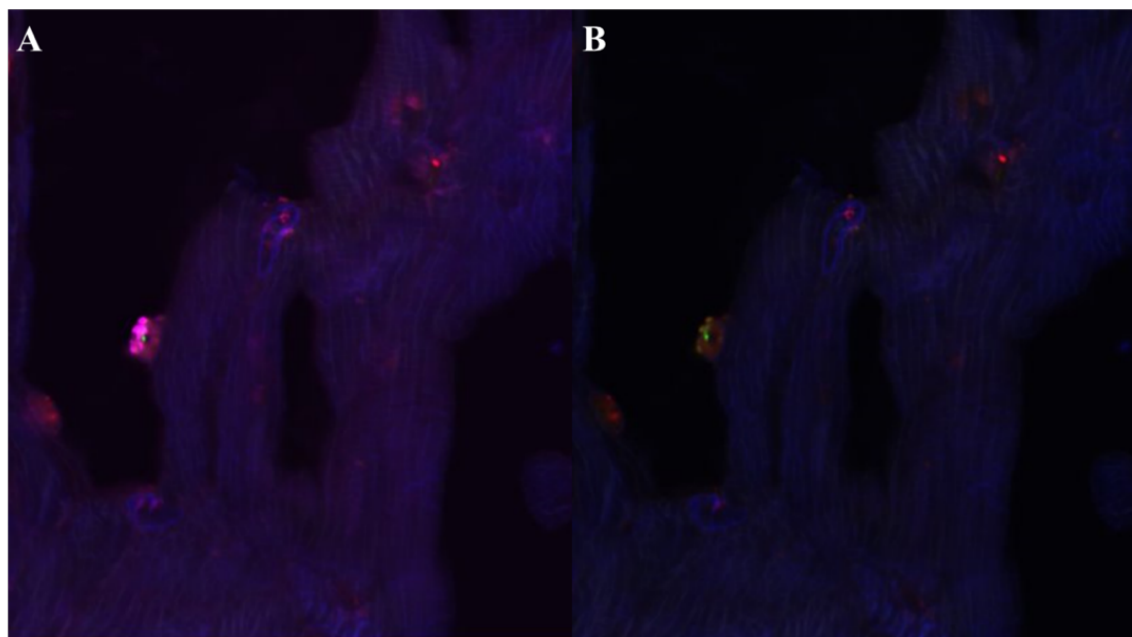
Quando o DAPI é utilizado nas amostras de mamute-lanoso, dificilmente é possível encontrar a imagem de um núcleo e todo o tecido possui coloração azul. Isto pode ocorrer possivelmente por contaminação por microrganismos ou porque os núcleos não se encontram mais intactos e têm seu material genético espalhado pelo tecido. Com isso, foi realizada uma nova abordagem com o intuito de se identificar estruturas nucleares com imunofluorescência utilizando anticorpo primário lamin B1 (lamínula nuclear). A lamínula nuclear é formada por proteínas que constituem o envelope nuclear. Na figura 59, abaixo, é possível ver o resultado do protocolo de imunofluorescência em fígado de elefante-asiático quando realizado apenas a imunofluorescência com marcação da lamínula nuclear em rosa (Cy5). Como os fluoróforos do espectro rosa foram os que apresentaram as melhores intensidades, seria interessante que fosse usado em um dos *probes*, portanto, testamos também a imunofluorescência para lamin B1 com anticorpo secundário marcados com pacific blue, que colore a lamínula nuclear com azul, e o resultado também foi satisfatório. Então, em um próximo experimento foi realizado o protocolo de FISH com os *probes* para cromossomo 11 (parte 1 e 2) e cromossomo 24, seguido de imunofluorescência de lamin B1 com anticorpo secundário pacific blue. O



resultado é mostrado na figura 60, e nota-se que a intensidade da imunofluorescência é menor que o de FISH.

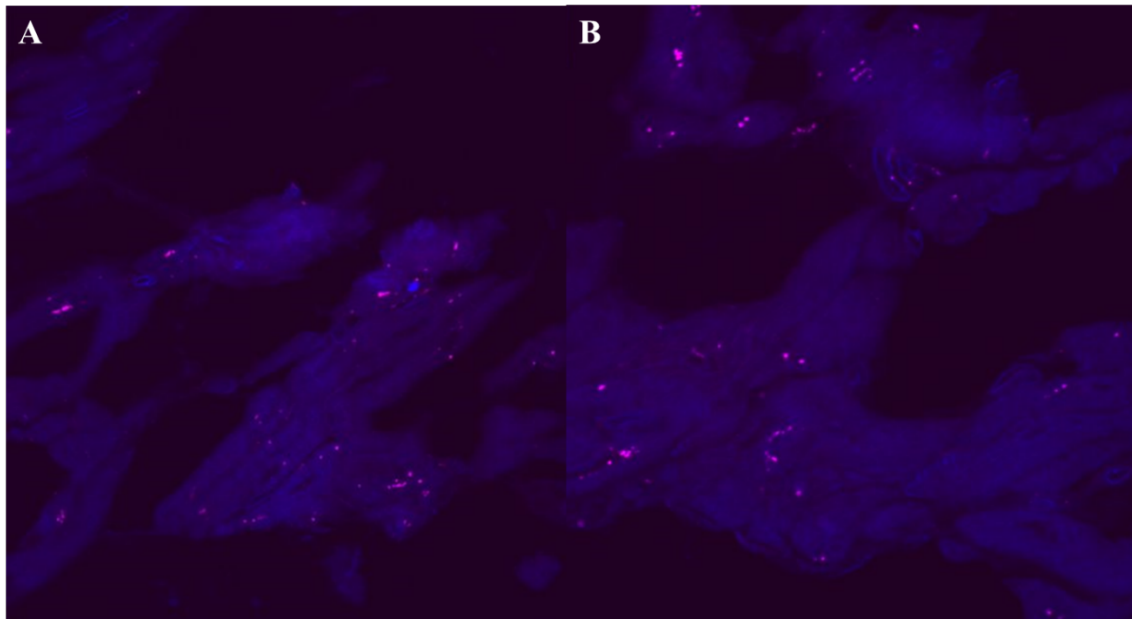


**Figura 59- imunofluorescência com lamin B1 (Cy5) em fígado de elefante-asiático.** Imagem de fígado de elefante-asiático após imunofluorescência com lamin B1 (contorno rosa no núcleo das células) e DAPI (azul). Lâmina apenas com imunofluorescência, sem FISH. Para que fosse possível capturar a imagem de lamin, foi necessário um maior tempo de exposição, resultando em excesso de *background*. A imagem foi obtida utilizando o microscópio confocal Nikon Eclipse Ti com aumento de 60x com lente objetiva com imersão em água e uso do *software* Metamorph 7.10.2.240. A imagem foi tratada com o *software* Fiji (Schindelin 2012). Imagem com sobreposição dos lasers de comprimento de onda de 405, 488, 561 e 640 nm.



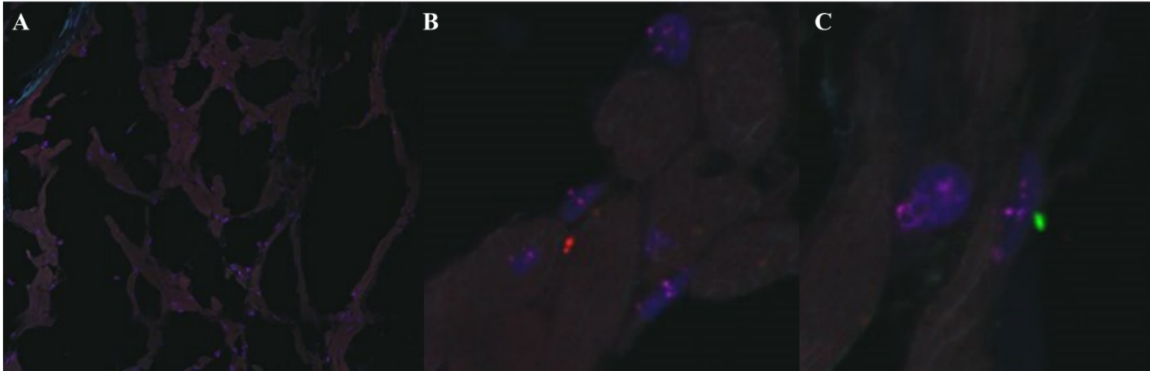
**Figura 60- FISH após imunofluorescência com lamin B1 (pacific blue) em fígado de elefante-asiático.** Primeiro foi feito o protocolo de FISH com os *probes* para cromossomo 11 parte 1 e 2 (visíveis em comprimentos de onda 640 e 488 nm, respectivamente) e cromossomo 24 (visível em comprimento de onda 561 nm) sem adição de DAPI, seguido pelo protocolo de imunofluorescência utilizando o anticorpo de lamin B1 com secundário pacific blue, que marca o alvo (contorno dos núcleos) na cor azul. O sinal da imunofluorescência é menos intenso que o de FISH. A - Imagem com sobreposição dos comprimentos de onda 405, 488, 561 e 640 nm. B – Imagem com sobreposição dos comprimentos de onda 405, 488 e 561 nm do mesmo fragmento de A, ficando um pouco mais evidente a o contorno nuclear. as imagens foram obtidas utilizando o microscópio confocal Nikon Eclipse Ti com aumento de 60x com lente objetiva com imersão em água e uso do *software* Metamorph 7.10.2.240. As imagens foram tratadas com o *software* Fiji (Schindelin 2012).

Como os tecidos de mamute-lanoso não estão no estado de conservação ideal, as lâminas com as amostras, apesar de terem sido preparadas com cortes de 5  $\mu\text{m}$ , geralmente possuem espessuras superiores, pois o estado do tecido não permite cortes tão finos. Isso faz com que a obtenção de imagens no microscópio seja dificultada, sendo necessário captar imagem de diferentes camadas (Zstacks). As sucessivas lavagens as quais as lâminas são submetidas no protocolo de FISH e de imunofluorescência faz com que haja desprendimento do tecido da lâmina, resultando em perda considerável de amostra. Assim, foi testado a adição de mais uma etapa de fixação do tecido em solução de PBS com 4% de paraformaldeído por 30 minutos ao final do protocolo de imunofluorescência e antes do início do protocolo de FISH, porém, isso resultou no surgimento de *background* nos comprimentos de onda 561 e 640 nm, conforme pode ser visto na figura 61, abaixo.

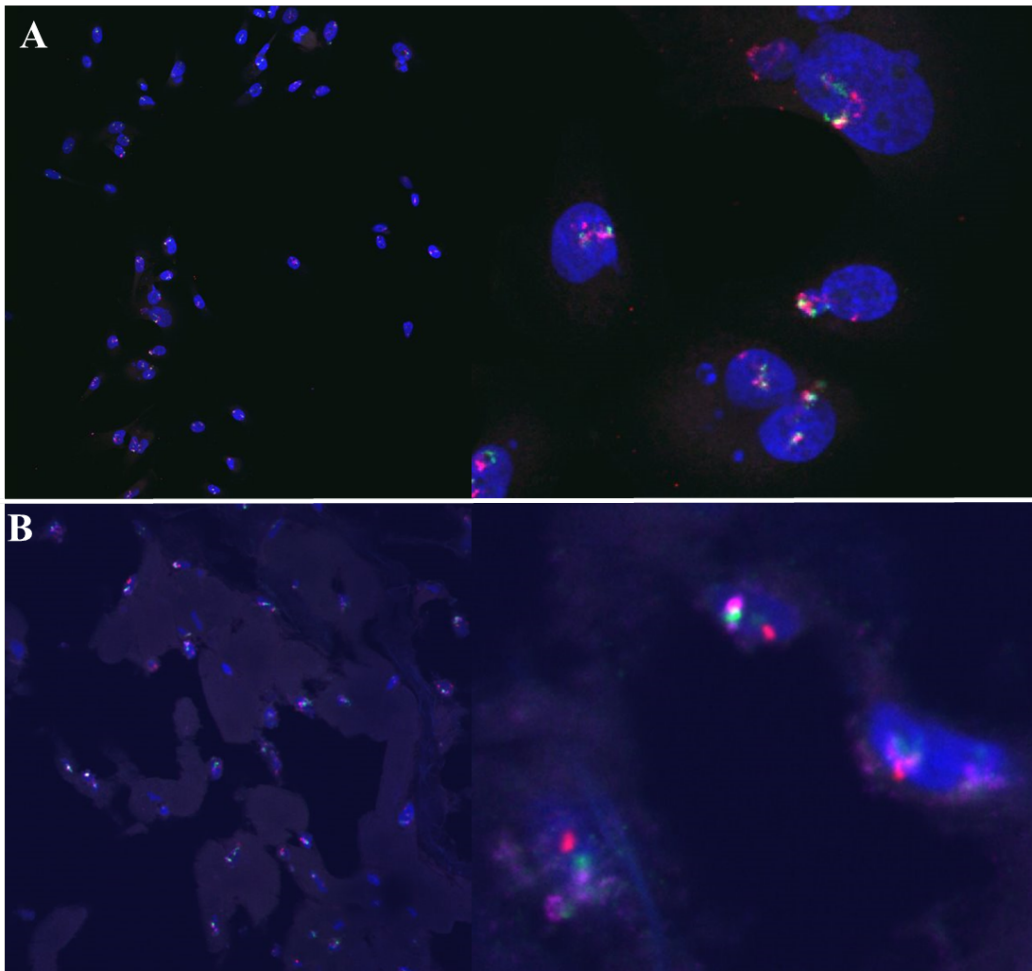


**Figura 61- Imunofluorescência com lamin B1 (pacific blue) com fixação extra ao final do protocolo em fígado de elefante-asiático.** Nessa lâmina foi realizada apenas a imunofluorescência com uma etapa de fixação a mais ao final do protocolo em uma tentativa de que o tecido ficasse mais íntegro para um FISH posterior. Porém, isto resultou em surgimento de *background* fora do núcleo nos comprimentos de onda 561 e 640 nm. As imagens foram obtidas utilizando o microscópio confocal Nikon Eclipse Ti com aumento de 60x com lente objetiva com imersão em água e uso do *software* Metamorph 7.10.2.240. As imagens foram tratadas com o *software* Fiji (Schindelin 2012) e mostram sobreposições dos lasers de comprimento de onda de 405, 488, 561 e 640 nm. A - zoom-out. B - zoom-in.

Outras combinações de cores e *probes* foram testadas utilizando o protocolo de FISH estabelecido com uso de TrueBlack, e todos apresentaram bons resultados. O uso de *probes* de centrômero (Cy5), cromossomo 11 parte 2 (atto 488) e cromossomo 24 (visível em comprimento de onda 561 nm) pode ser visto na figura 62, realizado em fígado de elefante-asiático, cuja imagem possui baixo *background* e bom sinal dos alvos. Resultado similar foi encontrado também na combinação de *probes* de cromossomo 11 parte 1 e 2 (visíveis nos comprimentos de onda 640 e 488 nm, respectivamente) e cromossomo 24 (visível em comprimento de onda 561 nm), que pode ser visto tanto em células como em tecido de elefante-asiático na figura 63.



**Figura 62- Centrômero, Cromossomo 11 e cromossomo 24 de elefante-asiático.** Resultado de FISH de fígado de elefante utilizando os *probes* para centrômero (visível em comprimento de onda 640 nm), cromossomo 11 parte 2 (visível em comprimento de onda 488 nm) e cromossomo 24 (visível em comprimento de onda 561 nm) e núcleo corado com DAPI. As imagens foram obtidas utilizando o microscópio confocal Nikon Eclipse Ti com aumento de 60x com lente objetiva com imersão em água e uso do *software* Metamorph 7.10.2.240. As imagens foram tratadas com o *software* Fiji (Schindelin 2012) e mostram sobreposições dos lasers de comprimento de onda de 405, 488, 561 e 640 nm. A - zoom-out. B e C - zoom-in.

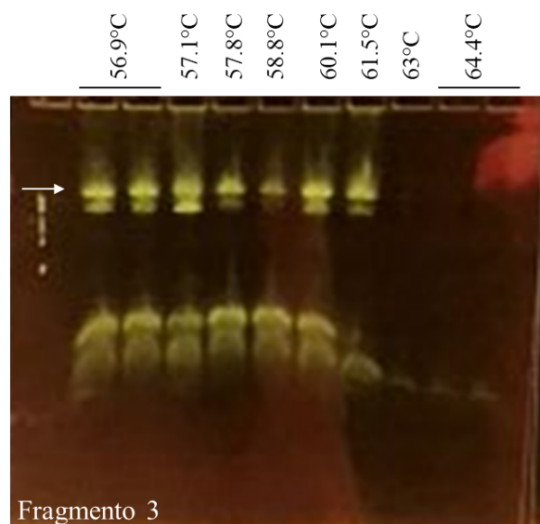


**Figura 63- Marcação do cromossomo 11 e cromossomo 24 de elefante-asiático.** Resultado de FISH utilizando os *probes* Cromossomo 11 parte 1 e 2 (visíveis em comprimentos de onda de 640 e 488 nm, respectivamente) e Cromossomo 24 (visível em comprimento de onda 561 nm) e núcleo corado com DAPI. As imagens foram obtidas utilizando o microscópio confocal Nikon Eclipse Ti com aumento de 60x com lente objetiva com imersão em água e uso do *software* Metamorph 7.10.2.240. As imagens foram tratadas com o *software* Fiji (Schindelin 2012) e mostram sobreposições dos lasers de comprimento de onda de 405, 488, 561 e 640 nm. A - Células mesenquimais de elefante-asiático (zoom-out à esquerda e zoom-in à direita). B – Fígado de elefante-asiático (zoom-out à esquerda e zoom-in à direita).

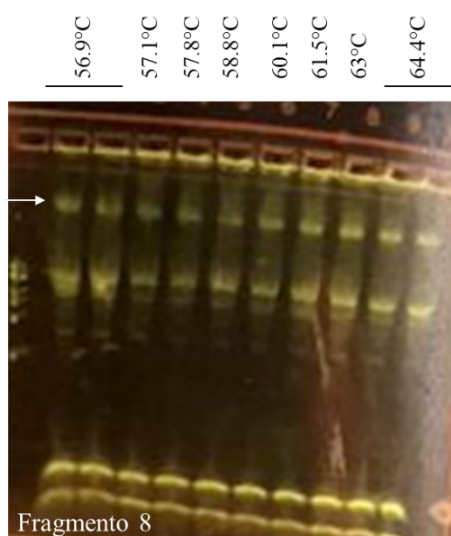
### 2.5.3 Herpervírus endoteliotrópico de elefante

Com exceção do fragmento 1, todos os demais foram amplificados nas PCR testadas. Todas as reações de PCR tiveram melhores resultados quando utilizado 37 ciclos e 10 minutos de extensão para os fragmentos com ~10kb de tamanho e 6 minutos de extensão para o fragmento com ~5kb. Alguns fragmentos (fragmentos 2, 3, 9, 16 e 19) amplificaram apenas com determinadas temperaturas de anelamento, como, por exemplo, o fragmento 3 (figura 64), que amplificou somente até 61.5°C e não amplificou em temperaturas maiores. Para estes, mais de um gel foi utilizado com o restante da PCR das temperaturas que apresentaram amplificação para obtenção de maior volume de fragmento alvo extraído e purificado. Outros fragmentos (fragmentos 6, 7, 8, 18 e 20) amplificaram em todas as temperaturas do gradiente, porém, com bandas de diferentes intensidades conforme exemplificado na figura 65 o gel da amplificação do fragmento 8. Já para outros fragmentos (fragmentos 4, 5, 10, 11, 12, 13, 14, 15, e 17) todas as temperaturas de anelamento testadas resultaram em amplificação com boa intensidade, conforme exemplificado na figura 66 as amplificações para os fragmentos 12 e 15.

A formação de bandas inespecíficas que podem ser vistas nos géis não interferem na qualidade do fragmento alvo, uma vez que os fragmentos desejados foram posteriormente extraídos. Para todos as reações cuja amplificação foi bem-sucedida, foi possível obter ao menos 30µl de produto de PCR purificado da banda de tamanho correspondente ao desejado e purificado, com concentração mínima de 29.4µg/µl. Independente da intensidade da banda, foi extraído o produto de PCR do gel de todas as bandas que apresentaram amplificação no tamanho alvo. Estes produtos de PCR purificados serão enviados para sequenciamento na empresa Genewiz para confirmação.

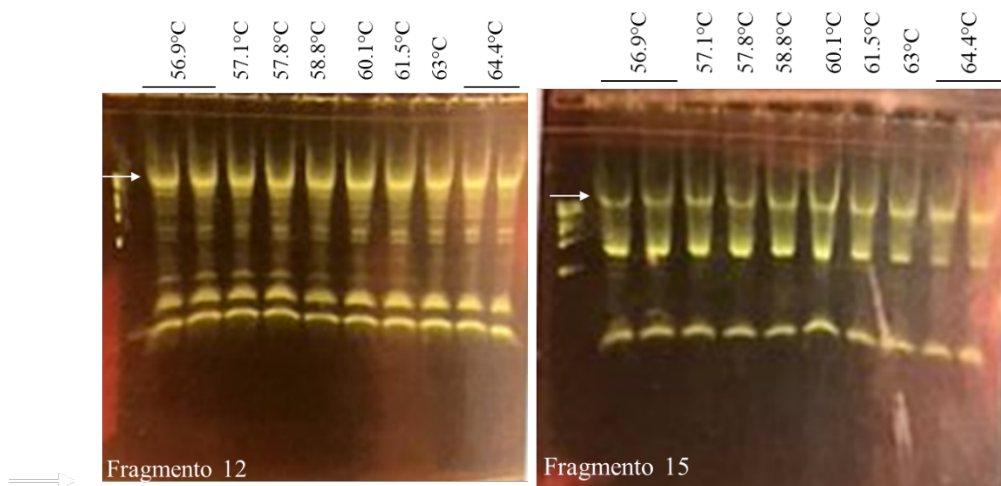


**Figura 64- Gradiente de temperatura de anelamento para o fragmento 3 de EEHV.** Foram aplicados 20 $\mu$ l de produto de PCR em cada poço do E-Gel 2% (invitrogen), com corrente por 15 minutos com padrão de peso molecular 1kb Plus DNA ladder (New England BioLabs) na primeira canaleta à esquerda em equipamento E-Gel Power Snap System (invitrogen) e visualização com o mesmo equipamento. A seta branca da imagem indica a banda do fragmento alvo que foi extraída. A amplificação ocorreu somente com temperaturas de anelamento de 56.9°C a 61.5°C.



**Figura 65- Gradiente de temperatura de anelamento para o fragmento 8 de EEHV.** Foram aplicados 20 $\mu$ l de produto de PCR em cada poço do E-Gel 2% (invitrogen), com corrente por 15 minutos com padrão de peso molecular 1kb Plus DNA ladder (New England BioLabs) na primeira canaleta à esquerda em equipamento E-Gel Power Snap System (invitrogen) e visualização com o mesmo equipamento. A seta branca da imagem indica a banda do fragmento alvo que foi extraída. A amplificação ocorreu com todas as temperaturas de anelamento testadas (56.9°C, 57.1°C, 57.8°C, 58.8°C, 60.1°C, 61.5°C, 63°C e 64.4°C), sendo que em algumas a intensidade da banda foi maior do que em outras temperaturas.





**Figura 66- Gradiente de temperatura de anelamento para os fragmentos 12 e 15 de EEHV.**

Para os fragmentos 12 e 15, à esquerda e direita, respectivamente, ocorreu amplificação satisfatória com todas as temperaturas de anelamento testadas (56.9°C, 57.1°C, 57.8°C, 58.8°C, 60.1°C, 61.5°C, 63°C e 64.4°C). Foram aplicados 20µl de produto de PCR em cada poço do E-Gel 2% (invitrogen), com corrente por 15 minutos com padrão de peso molecular 1kb Plus DNA ladder (New England BioLabs) na primeira canaleta à esquerda em equipamento E-Gel Power Snap System (invitrogen) e visualização com o mesmo equipamento. A seta branca da imagem indica a banda do fragmento alvo que foi extraída.



## 2.6 Conclusão

### 2.6.1 Conclusão FISH

Foi possível estabelecer um protocolo de FISH e de imunofluorescência combinada com FISH para uso em amostras de tecido de mamute-lanoso com redução de *background* utilizando TrueBlack e utilização de *probes* para centrômeros, cromossomo 11 e cromossomo 24 utilizando amostras de células e de fígado de elefante-asiático como controle positivo. A utilização destes protocolos ajudará na compreensão da preservação do DNA em amostras antigas congeladas, assim como sobre a estrutura nuclear da espécie.

### 2.6.2 Conclusão edição gênica

Foi possível desenhar sgRNA para o *knockout* de todos os genes pretendidos utilizando a enzima SpCas9, sendo a maioria com região de corte da dupla fita no exon 2. Todos foram selecionados visando a redução de *off-targets* e idealmente no início da região codificante. A edição gênica em *multiplex* ainda não teve seu protocolo padronizado com sucesso, porém, já tem seus *primers* desenhados para a realização de PCR para verificação da edição.

### 2.6.3 Conclusão EEHV

Todos os fragmentos de EEHV, com exceção do fragmento 1, foram amplificados, extraídos e purificados, e, após a confirmação pelo sequenciamento, estarão prontos para serem utilizados na construção do genoma viral. O fragmento 1 segue em fase experimental com o desenho de novos *primers* para amplificação.

## Considerações finais

O trabalho desenvolvido conseguiu alcançar o principal objetivo de desenvolver ferramentas genéticas para identificação de tráfico de *Saltator similis*, sendo uma importante abordagem para a sua conservação, uma vez que esta é uma das principais ameaças à espécie. Além disto, os marcadores microssatélites desenvolvidos poderão ser utilizados futuramente também com amostras de indivíduos oriundos da natureza de diferentes regiões geográficas, e, assim, obter informações sobre a ocorrência de estruturação genética. Caso haja, essa informação será útil também para as investigações de rota de tráfico, uma vez que poderá ser possível inferir a origem geográfica de um animal capturado ilegalmente na natureza. Adicionalmente, a metodologia descrita nesta tese poderá ser utilizada no desenvolvimento de outros marcadores, e, assim, contribuir também com a conservação, manejo e combate ao tráfico também de outras espécies. Ainda no contexto de conservação, o uso de biologia sintética tem se mostrado uma importante ferramenta nesta área, mas ainda se encontra em etapa inicial de desenvolvimento e requer mais estudos até a sua aplicabilidade. Também é válido destacar que, apesar de o projeto piloto possuir a espécie alvo em questão definida, a metodologia desenvolvida poderá, futuramente, ser utilizada na conservação de diversas espécies sob diferentes níveis de risco de extinção.

## Referências

### Referências capítulo 1

Abdelkrim, J., Robertson, B.C., Stanton, J.A.L. and Gemmell, N.J., 2009. Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. *BioTechniques*, 46(3), pp.185-192.

Aliabadian, M., Kaboli, M., Nijman, V. and Vences, M., 2009. Molecular identification of birds: performance of distance-based DNA barcoding in three genes to delimit parapatric species. *PLoS One*, 4(1), p.e4119.

Almeida, J.G. de, 2015. Análises comparativas das probabilidades de paternidade obtidas a partir de diferentes bancos de dados populacionais do Brasil. 81 f. Dissertação (programa de pós-graduação stricto sensu em genética) - Pontifícia Universidade Católica de Goiás, Goiânia - GO.

Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F. and Schreier, P.H., 1981. Sequence and organization of the human mitochondrial genome. *Nature*, 290(5806), pp.457-465.

Arif, I.A. and Khan, H.A., 2009. Molecular markers for biodiversity analysis of wildlife animals: a brief review. *Animal biodiversity and conservation*, 32(1), pp.9-17.

Avise, J.C., 1986. Mitochondrial DNA and the evolutionary genetics of higher animals. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 312(1154), pp.325-342.

B10K Genomes Project. 2016. China National GeneBank. Disponível em: <https://b10k.genomics.cn/index.html>. Acesso em 26.11.2020.

Beier, S., Thiel, T., Münch, T., Scholz, U. and Mascher, M., 2017. MISA-web: a web server for microsatellite prediction. *Bioinformatics*, 33(16), pp.2583-2585.

Bell, D., Robertson, S. and Hunter, P.R., 2004. Animal origins of SARS coronavirus: possible links with the international trade in small carnivores. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 359(1447), pp.1107-1114.

Bender, K., Schneider, P.M. e Rittner, C., 2000. Application of mtDNA sequence analysis in forensic casework for the identification of human remains. *Forensic science international*, 113(1-3), pp.103-107.

Benson, G., 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic acids research*, 27(2), pp.573-580.

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritsch, G., Pütz, J., Middendorf, M. and Stadler, P.F., 2013. MITOS: improved de novo metazoan mitochondrial genome annotation. *Molecular phylogenetics and evolution*, 69(2), pp.313-319.

Besser, J., Carleton, H.A., Gerner-Smidt, P., Lindsey, R.L. and Trees, E., 2018. Next-generation sequencing technologies and their application to the study and control of bacterial infections. *Clinical microbiology and infection*, 24(4), pp.335-341.

BirdLife International 2018. *Saltator similis*. *The IUCN Red List of Threatened Species* 2018:e.T22723882A132169900. <http://dx.doi.org/10.2305/IUCN.UK.2018-2.RLTS.T22723882A132169900.en>. Download em 04.12.2018.

Blower, D.C., Corley, S., Hereward, J.P., Riginos, C.R. and Ovenden, J.R., 2015. Characterisation and cross-amplification of 21 novel microsatellite *loci* for the dusky shark, *Carcharhinus obscurus*. *Conservation genetics resources*, 7(4), pp.909-912.

BOLD Systems. 2020. Identification Engine. Disponível em: [https://www.boldsystems.org/index.php/IDS\\_OpenIdEngine](https://www.boldsystems.org/index.php/IDS_OpenIdEngine). Acesso em 25.11.2020.

Borges, R.C., de Oliveira, A., Bernardo, N. and da Costa, R.M.M.C., 2006. Diagnóstico da fauna silvestre apreendida e recolhida pela Polícia Militar de Meio Ambiente de Juiz de Fora, MG (1998 e 1999). *Revista Brasileira de Zootecias*, 8(1).

Botstein, D., White, R.L., Skolnick, M. and Davis, R.W., 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American journal of human genetics*, 32(3), p.314.

Brasil. 1998. Lei nº 9.605, de 12 de fevereiro de 1998. Disponível em: <[http://www.planalto.gov.br/ccivil\\_03/leis/L9605.htm](http://www.planalto.gov.br/ccivil_03/leis/L9605.htm)>. Acesso em: 11.05.2020.

Broad, S., 2020. Wildlife trade, COVID-19 and zoonotic disease risk. TRAFFIC-the wildlife monitoring network. Disponível em: <https://www.traffic.org/site/assets/files/12764/covid-19-briefing-vfinal.pdf>. Acesso em 05.05.2020.

Brookes, C., Bright, J.A., Harbison, S. and Buckleton, J., 2012. Characterising *stutter* in forensic STR *multiplexes*. *Forensic Science International: Genetics*, 6(1), pp.58-63.

Brown, W.M., George, M. and Wilson, A.C., 1979. Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences*, 76(4), pp.1967-1971.

Budowle, B., Allard, M.W., Wilson, M.R. and Chakraborty, R., 2003. Forensics and mitochondrial DNA: applications, debates, and foundations. *Annual review of genomics and human genetics*, 4(1), pp.119-141.

Butler, J.M., 2005. Constructing STR *multiplex* assays. In *Forensic DNA Typing Protocols* (pp. 53-65). Humana Press.

Butler, J.M., Buel, E., Crivellente, F. and McCord, B.R., 2004. Forensic DNA typing by capillary electrophoresis using the ABI Prism 310 and 3100 genetic analyzers for STR analysis. *Electrophoresis*, 25(10-11), pp.1397-1412.

Cao, M.D., Balasubramanian, S. and Bodén, M., 2015. Sequencing technologies and tools for short tandem *repeat* variation detection. *Briefings in bioinformatics*, 16(2), pp.193-204.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S.M., Betley, J., Fraser, L., Bauer, M. and Gormley, N., 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME journal*, 6(8), pp.1621-1624.

Carvalho, M.C.D.C.G.D. and Silva, D.C.G.D., 2010. Next generation DNA sequencing and its applications in plant genomics. *Ciência Rural*, 40(3), pp.735-744.

Castoe, T.A., Poole, A.W., de Koning, A.J., Jones, K.L., Tomback, D.F., Oyler-McCance, S.J., Fike, J.A., Lance, S.L., Streicher, J.W., Smith, E.N. and Pollock, D.D., 2012. Rapid microsatellite identification from Illumina *paired-end* genomic sequencing in two birds and a snake. *PloS one*, 7(2), p.e30953.

Castoe, T.A., Poole, A.W., de Koning, A.J., Jones, K.L., Tomback, D.F., Oyler-McCance, S.J., Fike, J.A., Lance, S.L., Streicher, J.W., Smith, E.N. and Pollock, D.D., 2012. Rapid microsatellite identification from Illumina *paired-end* genomic sequencing in two birds and a snake. *PloS one*, 7(2), p.e30953.

Chabot, C.L., 2012. Characterization of 11 microsatellite *loci* for the brown smooth-hound shark, *Mustelus henlei* (Triakidae), discovered with next-generation sequencing. *Conservation Genetics Resources*, 4(1), pp.23-25.

Chakraborty, R., Andrade, M.D., Daiger, S.P. and Budowle, B., 1992. Apparent heterozygote deficiencies observed in DNA typing data and their implications in forensic applications. *Annals of human genetics*, 56(1), pp.45-57.

Chapuis, M.P. and Estoup, A., 2007. Microsatellite null alleles and estimation of population differentiation. *Molecular biology and evolution*, 24(3), pp.621-631.

Chevreur, B., Wetter, T. and Suhai, S., 1999, October. Genome sequence assembly using trace signals and additional sequence information. In *German conference on bioinformatics* (Vol. 99, No. 1, pp. 45-56).

Cohen, J., 2020. Mining coronavirus genomes for clues to the outbreak's origins. Disponível em: [https://scim.ag/ COVID-19genomeclues](https://scim.ag/COVID-19genomeclues). Acesso em 05.05.2020.

CONAMA (Conselho Nacional do meio Ambiente). 2018. Resolução nº 487, de 15 de maio de 2018. Diário Oficial da União, Brasília-DF, 16 mai. 2018. Edição 93, Seção 1, p 117.

Contente, A., Dittmer, A., Koch, M.C., Roth, J. and Dobbstein, M., 2002. A polymorphic microsatellite that mediates induction of PIG3 by p53. *Nature genetics*, 30(3), pp.315-320.

Cook, R., Karesh, W., Osofsky, S., 2014. One World, One Health: Building Interdisciplinary Bridges to Health in a Globalized World. Wildlife Conservation Society, Bronx, New York, USA. Disponível em: [http://www.oneworldonehealth.org/sept2004/owoh\\_sept04.html](http://www.oneworldonehealth.org/sept2004/owoh_sept04.html) Acesso em: 05.05.2020.

Csencsics, D., Brodbeck, S. and Holderegger, R., 2010. Cost-effective, species-specific microsatellite development for the endangered dwarf bulrush (*Typha minima*) using next-generation sequencing technology. *Journal of Heredity*, 101(6), pp.789-793.

Cyranoski, D., 2020. Mystery deepens over animal source of coronavirus. *Nature*, 579(7797), pp.18-20.

Da Maia, L.C., Palmieri, D.A., De Souza, V.Q., Kopp, M.M., de Carvalho, F.I.F. and Costa de Oliveira, A., 2008. SSR locator: tool for simple sequence *repeat* discovery integrated with *primer* design and PCR simulation. *International journal of plant genomics*, 2008.

Dakin, E.E. and Avise, J.C., 2004. Microsatellite null alleles in parentage analysis. *Heredity*, 93(5), pp.504-509.

Dakin, E.E. and Avise, J.C., 2004. Microsatellite null alleles in parentage analysis. *Heredity*, 93(5), pp.504-509.

Darby, B.J., Erickson, S.F., Hervey, S.D. and Ellis-Felege, S.N., 2016. Digital fragment analysis of short tandem repeats by high-throughput amplicon sequencing. *Ecology and evolution*, 6(13), pp.4502-4512.

Dashnow, H., Lek, M., Phipson, B., Halman, A., Sadedin, S., Lonsdale, A., Davis, M., Lamont, P., Clayton, J.S., Laing, N.G. and MacArthur, D.G., 2018. STRetch: detecting and discovering pathogenic short tandem repeat expansions. *Genome biology*, 19(1), pp.1-13.

Dawnay, N., Ogden, R., McEwing, R., Carvalho, G.R. and Thorpe, R.S., 2007. Validation of the barcoding gene COI for use in forensic genetic species identification. *Forensic science international*, 173(1), pp.1-6.

Destro, G.F.G., Pimentel, T.L., Sabaini, R.M., Borges, R.C. and Barreto, R., 2012. Efforts to combat wild animals trafficking in Brazil. *Biodiversity enrichment in a diverse world*, 1, pp.421-436.

Dohm, J.C., Lottaz, C., Borodina, T. and Himmelbauer, H., 2008. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic acids research*, 36(16), p.e105.

Edwards, A., Civitello, A., Hammond, H.A. and Caskey, C.T., 1991. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *American journal of human genetics*, 49(4), p.746.

Ellegren, H., 2000. Heterogeneous mutation processes in human microsatellite DNA sequences. *Nature genetics*, 24(4), pp.400-402..

Ellegren, H., 2004. Microsatellites: simple sequences with complex evolution. *Nature reviews genetics*, 5(6), pp.435-445.

Elwick, K., Zeng, X., King, J., Budowle, B. and Hughes-Stamm, S., 2018. Comparative tolerance of two massively parallel sequencing systems to common PCR inhibitors. *International journal of legal medicine*, 132(4), pp.983-995.

Faircloth, B.C., 2008. MSATCOMMANDER: Detection of microsatellite repeat arrays and automated, locus-specific primer design. *Molecular ecology resources*, 8(1), pp.92-94.

Fazekas, A.J., Steeves, R. and Newmaster, S.G., 2010. Improving sequencing quality from PCR products containing long mononucleotide repeats. *Biotechniques*, 48(4), pp.277-285.

Felker, R.M., Dörr, A.C., Rovedder, A.P., Piazza, E.M. and Dick, G., 2013. Levantamento parcial da avifauna apreendida pelo escritório regional do IBAMA de Santa Maria-RS. *Revista Eletrônica em Gestão, Educação e Tecnologia Ambiental*, 11, pp.2506-2510. Disponível em: <http://cascavel.ufsm.br/revistas/ojs-2.2.2/index.php/reget/article/download/8734/pdf>. Acesso em: 18.08.2016.

Feng, S., Stiller, J., Deng, Y., Armstrong, J., Fang, Q., Reeve, A.H., Xie, D., Chen, G., Guo, C., Faircloth, B.C. and Petersen, B., 2020. Dense sampling of bird diversity increases power of comparative genomics. *Nature*, 587(7833), pp.252-257.

Fernandez-Silva, I., Whitney, J., Wainwright, B., Andrews, K.R., Ylitalo-Ward, H., Bowen, B.W., Toonen, R.J., Goetze, E. and Karl, S.A., 2013. Microsatellites for next-generation ecologists: a post-sequencing bioinformatics pipeline. *PLoS one*, 8(2), p.e55990.

Ferreira, J.M., 2021. Contribuição da genética de populações à investigação sobre o tráfico de fauna no Brasil: Desenvolvimento de microssatélites e análise da estrutura genética em *Paroaria dominicana* e *Saltator similis* (Aves: Passeriformes:Thraupidae). 156 f. Tese (departamento de genética e biologia evolutiva) – Instituto de Biociências da Universidade de São Paulo, Sao Paulo-SP.

Franco, M., Câmara, F., Rocha, D.C., Souza, R.M. and Oliveira, N., 2012. Animais silvestres apreendidos no período de 2002 a 2007 na macrorregião de Montes Claros, Minas Gerais. *Enciclopédia Biosfera*, 8(14).

Françoso, E., Gomes, F. and Arias, M.C., 2016. A protocol for isolating insect mitochondrial genomes: a case study of NUMT in *Melipona flavolineata* (Hymenoptera: Apidae). *Mitochondrial DNA Part A*, 27(4), pp.2401-2404.

Freitas, A.C.P.D., Oviedo-Pastrana, M.E., Vilela, D.A.D.R., Pereira, P.L.L., Loureiro, L.D.O.C., Haddad, J.P.A., Martins, N.R.D.S. and Soares, D.F.D.M., 2015. Diagnóstico de animais ilegais recebidos no centro de triagem de animais silvestres de Belo Horizonte, Estado de Minas Gerais, no ano de 2011. *Ciência Rural*, 45(1), pp.163-170. Disponível em: <http://www.scielo.br/pdf/cr/v45n1/0103-8478-cr-45-01-00163.pdf>. Acesso em: 03.08.2016.

Gagneux, P., Boesch, C. and Woodruff, D.S., 1997. Microsatellite scoring errors associated with noninvasive genotyping based on nuclear DNA amplified from shed hair. *Molecular ecology*, 6(9), pp.861-868.

Gardner, M.G., Fitch, A.J., Bertozzi, T. and Lowe, A.J., 2011. Rise of the machines—recommendations for ecologists when using next generation sequencing for microsatellite development. *Molecular Ecology Resources*, 11(6), pp.1093-1101.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D. and Bairoch, A., 2003. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic acids research*, 31(13), pp.3784-3788.

Gettings, K.B., Kiesler, K.M., Faith, S.A., Montano, E., Baker, C.H., Young, B.A., Guerrieri, R.A. and Vallone, P.M., 2016. Sequence variation of 22 autosomal STR loci detected by next generation sequencing. *Forensic Science International: Genetics*, 21, pp.15-21.

Gianaroli L., Magli M.C., Ferraretti A.P., 2013. Capítulo “Sex Chromosomes” em Brenner's Encyclopedia of Genetics, 2a edição. Editor: Stanley Maloy, Kelly Hughes. Academic Press, p. 397-400.

Gogliath, M., Bisaggio, E.L., Ribeiro, L.B., Resgalla, A.E. and Borges, R.C., 2010. Avifauna apreendida e entregue voluntariamente ao Centro de Triagem de Animais Silvestres (CETAS) do IBAMA de Juiz de Fora, Minas Gerais. *Leonardo*, 1(3). Disponível em: [http://www.ao.com.br/download/ao154\\_55.pdf](http://www.ao.com.br/download/ao154_55.pdf). Acesso em: 18.08.2016.

Goldbio. 2018. Polymerase Chain Reaction (PCR) - Utilizing *Hot start* Pfu DNA Polymerase. Disponível em: <https://www.goldbio.com/uploads/documents/9ea27b6fa6c5446b4407c83210b674f5.pdf>. Acesso em 26.05.2021.

Gouy, A. and Zieger, M., 2017. STRAF—a convenient online tool for STR data evaluation in forensic genetics. *Forensic Science International: Genetics*, 30, pp.148-151.

Green, M.R. and Sambrook, J., 2018. *Hot start* polymerase chain reaction (PCR). *Cold Spring Harbor Protocols*, 2018(5), pp.pdb-prot095125.

Gregory, T.R., 2002. A bird's-eye view of the C-value enigma: genome size, cell size, and metabolic rate in the class Aves. *Evolution*, 56(1), pp.121-130.

Gregory, T.R., Andrews, C.B., McGuire, J.A. and Witt, C.C., 2009. The smallest avian genomes are found in hummingbirds. *Proceedings of the Royal Society B: Biological Sciences*, 276(1674), pp.3753-3757.

Gregory, T.R., Nicol, J.A., Tamm, H., Kullman, B., Kullman, K., Leitch, I.J., Murray, B.G., Kapraun, D.F., Greilhuber, J. and Bennett, M.D., 2007. Eukaryotic genome size databases. *Nucleic acids research*, 35(suppl\_1), pp.D332-D338.

Guichoux, E., Lagache, L., Wagner, S., Chaumeil, P., Léger, P., Lepais, O., Lepoittevin, C., Malausa, T., Revardel, E., Salin, F. and Petit, R.J., 2011. Current trends in microsatellite genotyping. *Molecular ecology resources*, 11(4), pp.591-611.

Gyllensten, U., Wharton, D., Josefsson, A. and Wilson, A.C., 1991. Paternal inheritance of mitochondrial DNA in mice. *Nature*, 352(6332), pp.255-257.

Gymrek, M., 2017. A genomic view of short tandem repeats. *Current opinion in genetics & development*, 44, pp.9-16.

Hackett, S.J., Kimball, R.T., Reddy, S., Bowie, R.C., Braun, E.L., Braun, M.J., Chojnowski, J.L., Cox, W.A., Han, K.L., Harshman, J. and Huddleston, C.J., 2008. A phylogenomic study of birds reveals their evolutionary history. *science*, 320(5884), pp.1763-1768.

Hancock, J.M., 1996. Simple sequences in a 'minimal' genome. *Nature genetics*, 14(1), pp.14-15.

Highnam, G., Franck, C., Martin, A., Stephens, C., Puthige, A. and Mittelman, D., 2013. Accurate human microsatellite genotypes from high-throughput resequencing data using informed error profiles. *Nucleic acids research*, 41(1), pp.e32-e32.

Hill, C.R., Butler, J.M. and Vallone, P.M., 2009. A 26plex autosomal STR assay to aid human identity testing. *Journal of Forensic Sciences*, 54(5), pp.1008-1015.

Hodel, R.G., Segovia-Salcedo, M.C., Landis, J.B., Crowl, A.A., Sun, M., Liu, X., Gitzendanner, M.A., Douglas, N.A., Germain-Aubrey, C.C., Chen, S. and Soltis, D.E., 2016. The report of my death was an exaggeration: a review for researchers using microsatellites in the 21st century. *Applications in plant sciences*, 4(6), p.1600025.

Hoogenboom, J., van der Gaag, K.J., de Leeuw, R.H., Sijen, T., de Knijff, P. and Laros, J.F., 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 Science International: Genetics*, 27, pp.27-40.

Hughes, A.L. and Hughes, M.K., 1995. Small genomes for better flyers. *Nature*, 377(6548), pp.391-391.

IBAMA. 2011. Instrução normativa nº 10, de 19 de setembro de 2011. Diário Oficial da União, Brasília-DF, 20 set. 2011. Seção 1, p 109.

IBAMA. 2014. Instrução normativa nº 23, de 31 de dezembro de 2014. Diário Oficial da União, Brasília-DF, 2 jan. 2015. Seção 1, p 115.

IBAMA. 2016. Relatório Técnico CETAS 2002-2014. Disponível em: [http://www.consultaesic.cgu.gov.br/busca/dados/Lists/Pedido/Attachments/539109/RES POSTA\\_PEDIDO\\_an.pdf](http://www.consultaesic.cgu.gov.br/busca/dados/Lists/Pedido/Attachments/539109/RES POSTA_PEDIDO_an.pdf). Acesso em 07.05.2020.

IBAMA. 2019a. Resposta do pedido de informação feito por meio do Sistema Eletrônico do Serviço de Informação ao Cidadão (e-SIC) protocolo n. 23480028207201826, de 11/12/2018. Formulário de resposta datado de 04/01/2019, encaminhado Diretoria de Uso Sustentável da Biodiversidade e Florestas (DBFlo).

Ibama. 2019b. Resposta do pedido de informação feito por meio do Sistema Eletrônico do Serviço de Informação ao Cidadão (e-SIC) protocolo n. 02680000132201973, de 18/01/2019. Formulário de resposta datado de 04/02/2019, encaminhado Diretoria de Uso Sustentável da Biodiversidade e Florestas (DBFlo).

Ibama. 2019c. Resposta do pedido de informação feito por meio do Sistema Eletrônico do Serviço de Informação ao Cidadão (e-SIC) protocolo n. 02680000531201934, de 12/03/2019. Formulário de resposta datado de 11/04/2019, encaminhado Diretoria de Uso Sustentável da Biodiversidade e Florestas (DBFlo).

Ibama. 2019d. Unidades dos Centros de Triagem de Animais Silvestres (Cetas). Disponível em: <https://www.Ibama.gov.br/institucional/cetas#cetas-mg>. Acesso em 05.09.2020.

Ibama. 2020. Resposta do pedido de informação feito por meio do Sistema Eletrônico do Serviço de Informação ao Cidadão (e-SIC) protocolo n. 02680.001506/2020-10, de 06/06/2020. Formulário de resposta datado de 15/06/2020, encaminhado Diretoria de Uso Sustentável da Biodiversidade e Florestas (DBFlo).

Illumina. 2011. Technical Note: SequencingQuality Scores for Next-Generation SequencingDisponível em: [https://www.illumina.com/documents/products/technotes/technote\\_Q-Scores.pdf](https://www.illumina.com/documents/products/technotes/technote_Q-Scores.pdf). Acesso em 15.09.2020.

Illumina. 2019. Targeted next-generation sequencing versus qPCR and Sanger sequencing. Disponível em: <https://www.illumina.com/content/dam/illumina-marketing/documents/products/other/infographic-targeted-ngs-vs-sanger-qpcr.pdf>. Acesso em 14.11.2020.

Illumina. 2020a. Key differences between next-generation sequencing and Sanger sequencing. Disponível em: <https://www.illumina.com/science/technology/next-generation-sequencing/ngs-vs-sanger-sequencing.html>. Acesso em 14.11.2020.

Illumina. 2020b. Illumina sequencing platforms. Disponível em: <https://www.illumina.com/systems/sequencing-platforms.html>. Acesso em 12.05.2020.

IMA (Instituto de meio ambiente- Estado de Alagoas), 2016. IMA e Ibama realizam operação para regularizar a criação amadora de pássaros. Disponível em: <https://www.ima.al.gov.br/operacao-flagra-crimes-contr-a-fauna-silvestre-alagoana/>. Acesso em 11/05/2020.

International Human Genome Sequencing Consortium, 2001. Initial sequencing and analysis of the human genome. *Nature*, 409(6822), pp.860-921.

Jarne, P. and Lagoda, P.J., 1996. Microsatellites, from molecules to populations and back. *Trends in ecology & evolution*, 11(10), pp.424-429.

John, J.C.S., 2014. Mitochondrial DNA: Its Transmission from Gametes and Embryos. In *Principles of cloning* (pp. 429-439). Academic Press.

Johns, G.C. and Avise, J.C., 1998. A comparative summary of genetic distances in the vertebrates from the mitochondrial cytochrome b gene. *Molecular biology and evolution*, 15(11), pp.1481-1490.

Kalinowski, S.T., Taper, M.L. and Marshall, T.C., 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular ecology*, 16(5), pp.1099-1106.

Karesh, W.B., Cook, R.A., Bennett, E.L. and Newcomb, J., 2005. Wildlife trade and global disease emergence. *Emerging infectious diseases*, 11(7), p.1000.

Katti, M.V., Ranjekar, P.K. and Gupta, V.S., 2001. Differential distribution of simple sequence repeats in eukaryotic genome sequences. *Molecular biology and evolution*, 18(7), pp.1161-1167.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C. and Thierer, T., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12), pp.1647-1649.

Kim, T.S., Booth, J.G., Gauch, H.G., Sun, Q., Park, J., Lee, Y.H. and Lee, K., 2008. Simple sequence repeats in *Neurospora crassa*: distribution, polymorphism and evolutionary inference. *BMC genomics*, 9(1), pp.1-19.

Knapik, E.W., Goodman, A., Ekker, M., Chevrette, M., Delgado, J., Neuhauss, S., Shimoda, N., Driever, W., Fishman, M.C. and Jacob, H.J., 1998. A microsatellite genetic linkage map for zebrafish (*Danio rerio*). *Nature genetics*, 18(4), pp.338-343.

Krehenwinkel, H., Meese, S., Mayer, C., Ruch, J., Schneider, J., Bilde, T., Künzel, S., Henderson, J.B., Russack, J., Simison, W.B. and Gillespie, R., 2019. Cost effective microsatellite isolation and genotyping by high throughput sequencing. *The Journal of Arachnology*, 47(2), pp.190-201.

Kvist, L., Martens, J., Nazarenko, A.A. and Orell, M., 2003. Paternal leakage of mitochondrial DNA in the great tit (*Parus major*). *Molecular Biology and Evolution*, 20(2), pp.243-247.

Kwok, S., Kellogg, D.E., McKinney, N., Spasic, D., Goda, L., Levenson, C. and Sninsky, J.J., 1990. Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus type 1 model studies. *Nucleic acids research*, 18(4), pp.999-1005.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W. and Funke, R., 2001. Initial sequencing and analysis of the human genome. *Nature*, 409(6822), pp.860-921.

Langmead, B. and Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nat Meth. Nature Research*, 9, pp.357-9.

Leroy, E.M., Rouquet, P., Formenty, P., Souquière, S., Kilbourne, A., Froment, J.M., Bermejo, M., Smit, S., Karesh, W., Swanepoel, R. and Zaki, S.R., 2004. Multiple Ebola virus transmission events and rapid decline of central African wildlife. *Science*, 303(5656), pp.387-390.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J. and Homer, N., 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools *Bioinformatics* 2009; 25: 2078-9.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R., 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), pp.2078-2079.

Lima R.M.G, 1998. Polimorfismos de microssatélites em DNA de eqüinos e seu uso na determinação de parentesco em animais da raça mangalarga machador. Tese (Doutorado em Ciência Animal) -Escola de Veterinária, Universidade Federal de Minas Gerais, Belo Horizonte. 91 p.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L. and Law, M., 2012. Comparison of next-generation sequencing systems. *Journal of Biomedicine and Biotechnology*, 2012.

Luo, S., Valencia, C.A., Zhang, J., Lee, N.C., Slone, J., Gui, B., Wang, X., Li, Z., Dell, S., Brown, J. and Chen, S.M., 2018. Biparental inheritance of mitochondrial DNA in humans. *Proceedings of the National Academy of Sciences*, 115(51), pp.13039-13044.

Maduna, S.N., Rossouw, C., Da Silva, C., Soekoe, M. and Bester-van der Merwe, A.E., 2017. Species identification and comparative population genetics of four coastal houndsharks based on novel NGS-mined microsatellites. *Ecology and evolution*, 7(5), pp.1462-1486.

Malausa, T., Gilles, A., Megléc, E., Blanquart, H., Duthoy, S., Costedoat, C., Dubut, V., Pech, N., CASTAGNONE-SERENO, P.H.I.L.I.P.P.E., Delye, C. and Feau, N., 2011. High-throughput microsatellite isolation through 454 GS-FLX Titanium pyrosequencing of enriched DNA libraries. *Molecular Ecology Resources*, 11(4), pp.638-644.

Manhaes, M.A. and Loures-Ribeiro, A., 2011. The avifauna of the Poço D'Anta Municipal Biological Reserve, Juiz de Fora, MG. *Biota Neotropica*, 11(3), pp.275-286.

MAPA. 2017. Instrução normativa nº 45, de 15 de dezembro de 2017. Diário Oficial da União, Brasília-DF, 26 dez. 2017. Seção 1, p 4.

Mardis, E.R., 2017. DNA sequencing technologies: 2006–2016. *Nature protocols*, 12(2), p.213.

Marshall, T.C., Slate, J.B.K.E., Kruuk, L.E.B. and Pemberton, J.M., 1998. Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular ecology*, 7(5), pp.639-655.

Mayrink R.R., 2016. Exame pericial para detecção de fraudes em anilhas oficiais de passeriformes: uma ferramenta para o combate ao tráfico de animais silvestres. Dissertação de mestrado. Universidade Federal de Santa Catarina, Florianópolis. 195 p.

McCracken, G.R., Wilson, K.L., Paterson, I., Perry, R., Keefe, D. and Ruzzante, D.E., 2014. Development of 17 novel microsatellite markers for the longnose sucker (*Catostomus catostomus*) and successful cross-specific amplification of 14 previously developed markers from congeneric species. *Conservation Genetics Resources*, 6(2), pp.329-332.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. and DePristo, M.A., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*, 20(9), pp.1297-1303.

MERCK. Whatman FTA card technology. Disponível em: <[https://www.sigmaaldrich.com/catalog/product/sigma/whawb120205?lang=pt&region=BR&gclid=EAJaIQobChMI9rPrmceF3AIVxAeRCh2pmAHdEAAYASAAEgLGyvD\\_BwE](https://www.sigmaaldrich.com/catalog/product/sigma/whawb120205?lang=pt&region=BR&gclid=EAJaIQobChMI9rPrmceF3AIVxAeRCh2pmAHdEAAYASAAEgLGyvD_BwE)>. Acesso em: 16.04.2019.

Mikota, S.K. and Maslow, J.N., 2011. Tuberculosis at the human–animal interface: An emerging disease of elephants. *Tuberculosis*, 91(3), pp.208-211.

Milach S.C.K., 1998. Marcadores de DNA em plantas. UFRGS, Porto Alegre, 141 p.

Montali, R.J., Mikota, S.K. and Cheng, L.I., 2001. Mycobacterium tuberculosis in zoo and wildlife species. *Scientific and Technical Review, International Office of Epizootics*.

Moore, W.S., 1995. Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. *Evolution*, 49(4), pp.718-726.

Moreira H.L.M., 1999. Análise da estrutura de plantéis e diversidade genética de estoques de reprodutores de tilápia do Nilo (*Oreochromis niloticus*) estimadas por microsatélites. Tese (Doutorado em Ciências) - Universidade Federal do Rio Grande do Sul, Porto Alegre. 112 p.

Morgante, M., Hanafey, M. and Powell, W., 2002. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nature genetics*, 30(2), pp.194-200.

National Center for Biotechnology Information (NCBI). 1988. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; Disponível em: <https://www.ncbi.nlm.nih.gov/>. Acesso em 01.12.20.

Neff, B.D. and Gross, M.R., 2001. Microsatellite evolution in vertebrates: inference from AC dinucleotide repeats. *Evolution*, 55(9), pp.1717-1733.

Nellemann, C., Henriksen, R., Kreilhuber, A., Stewart, D., Kotsovou, M., Raxter, P., Mrema, E. and Barrat, S., 2016. *The rise of environmental crime: a growing threat to natural resources, peace, development and security*. United Nations Environment Programme (UNEP).

NIH- National Institute of Health. 2020. The Human Genome Project. Disponível em: <https://www.genome.gov/human-genome-project>. Acesso em 16.11.2020.

Nowak, C., Zuther, S., Leontyev, S.V. and Geismar, J., 2014. Rapid development of microsatellite markers for the critically endangered Saiga (*Saiga tatarica*) using Illumina® Miseq next generation sequencing technology. *Conservation Genetics Resources*, 6(1), pp.159-162.

Nunes, P.B., Barreto, A.S. and Franco, E.Z., 2012. Subsídios à ação fiscalizatória no combate ao tráfico de aves silvestres e exóticas em Santa Catarina. *Ornithologia*, 5(1), pp.26-33.

Olejniczak, M. and Krzyzosiak, W.J., 2006. Genotyping of simple sequence repeat factors implicated in shadow band generation revisited. *Electrophoresis*, 27(19), pp.3724-3734.

Oliveira, E.J., Pádua, J.G., Zucchi, M.I., Vencovsky, R. and Vieira, M.L.C., 2006. Origin, evolution and genome distribution of microsatellites. *Genetics and Molecular Biology*, 29(2), pp.294-307.

ONU (Organização das Nações Unidas). 2015. Tackling illicit trafficking in wildlife. 69th session of the UN General Assembly, v. Resolution n. A/RES/69/314 n. A/69/PV.100 - 30 Julho 2015 - GA/11666 - A/69/L.80 & Add.1

Organ, C.L. and Edwards, S.V., 2011. Major events in avian genome evolution in Living Dinosaurs: The Evolutionary History of Modern Birds, DGK G, Editor.

Pena, I.F., 2013. Desenvolvimento de teste de paternidade por análise de DNA de Amazona aestiva (papagaio verdadeiro) com aplicabilidade na ciência forense. 113 f. Tese (programa de pós-graduação em genética) - Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais, Belo Horizonte-MG.

Peona, V., Weissensteiner, M.H. and Suh, A., 2018. How complete are “complete” genome assemblies?—An avian perspective.

Petrossian, G.A., Pires, S.F. and van Uhm, D.P., 2016. An overview of seized illegal wildlife entering the United States. *Global Crime*, 17(2), pp.181-201.

Phillips, C., Fondevila, M., García-Magariños, M., Rodriguez, A., Salas, A., Carracedo, A. and Lareu, M.V., 2008. Resolving relationship tests that show ambiguous STR results using autosomal SNPs as supplementary markers. *Forensic Science International: Genetics*, 2(3), pp.198-204.

Pimentel, J.D.S.M., Ludwig, S., Resende, L.C., Brandão-Dias, P.F.P., Pereira, A.H., de Abreu, N.L., Rosse, I.C., Martins, A.P.V., Facchin, S., Lopes, J.D.M. and Santos, G.B., 2020. Genetic evaluation of migratory fish: Implications for conservation and stocking programs. *Ecology and evolution*, 10(19), pp.10314-10324.

Pimentel, J.S., Carmo, A.O., Rosse, I.C., Martins, A.P., Ludwig, S., Facchin, S., Pereira, A.H., Brandão-Dias, P.F., Abreu, N.L. and Kalapothakis, E., 2018. High-throughput sequencing strategy for microsatellite genotyping using neotropical fish as a model. *Frontiers in Genetics*, 9, p.73.

Pinto L.A.F., 2014. O que é a transação penal? Disponível em <https://luizantoniofp.jusbrasil.com.br/artigos/148612891/o-que-e-transacao-penal>. Acesso em 11.05.2020.

Pirog, A., Blaison, A., Jaquemet, S., Soria, M. and Magalon, H., 2015. Isolation and characterization of 20 microsatellite markers from *Carcharhinus leucas* (bull shark)

and cross-amplification in *Galeocerdo cuvier* (tiger shark), *Carcharhinus obscurus* (dusky shark) and *Carcharhinus plumbeus* (sandbar shark). *Conservation Genetics Resources*, 7(1), pp.121-124. Walsh PS, Fildes NJ, Reynolds R. Sequence analysis and characterization of *stutter* products at the tetranucleotide *repeat locus* vWA. *Nucleic acids research*. 1996 Jul 1;24(14):2807-12.

Primmer, C.R., Møller, A.P. and Ellegren, H., 1995. Resolving genetic relationships with microsatellite markers: a parentage testing system for the swallow *Hirundo rustica*. *Molecular Ecology*, 4(4), pp.493-498.

Primmer, C.R., Raudsepp, T., Chowdhary, B.P., Møller, A.P. and Ellegren, H., 1997. Low frequency of microsatellites in the avian genome. *Genome Research*, 7(5), pp.471-482.

Provan, J., Powell, W. and Hollingsworth, P.M., 2001. Chloroplast microsatellites: new tools for studies in plant ecology and evolution. *Trends in ecology & evolution*, 16(3), pp.142-147.

Qiagen Bioinformatics – Variant types. Disponível em: [http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/700/Variant\\_types.html](http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/700/Variant_types.html). Acesso em 16.06.21

Quinn, T.W. and Wilson, A.C., 1993. Sequence evolution in and around the mitochondrial control region in birds. *Journal of molecular evolution*, 37(4), pp.417-425.

Rafalski, J.A. and Tingey, S.V., 1993. Genetic diagnostics in plant breeding: RAPDs, microsatellites and machines. *Trends in Genetics*, 9(8), pp.275-280.

Reyes, A., Yang, M.Y., Bowmaker, M. and Holt, I.J., 2005. Bidirectional replication initiates at sites throughout the mitochondrial genome of birds. *Journal of Biological Chemistry*, 280(5), pp.3242-3250.

Rojas, M., González, I., Pavón, M.Á., Pegels, N., Hernández, P.E., García, T. and Martín, R., 2011. Development of a real-time PCR assay to control the illegal trade of meat from protected capercaillie species (*Tetrao urogallus*). *Forensic science international*, 210(1-3), pp.133-138.

Roques, S., Fox, C.J., Villasana, M.I. and Rico, C., 2006. The complete mitochondrial genome of the whiting, *Merlangius merlangus* and the haddock, *Melanogrammus aeglefinus*: a detailed genomic comparison among closely related species of the Gadidae family. *Gene*, 383, pp.12-23.

Rubinoff, D., 2006. Utility of mitochondrial DNA barcodes in species conservation. *Conservation Biology*, 20(4), pp.1026-1033.

Santana, Q.C., Coetzee, M.P., Steenkamp, E.T., Mlonyeni, O.X., Hammond, G.N., Wingfield, M.J. and Wingfield, B.D., 2009. Microsatellite discovery by deep sequencing of enriched genomic libraries. *Biotechniques*, 46(3), pp.217-223.

Schirmer, M., Ijaz, U.Z., D'Amore, R., Hall, N., Sloan, W.T. and Quince, C., 2015. Insight into biases and sequencing errors for *amplicon* sequencing with the Illumina MiSeq platform. *Nucleic acids research*, 43(6), pp.e37-e37.

Schlötterer, C., 2000. Evolutionary dynamics of microsatellite DNA. *Chromosoma*, 109(6), pp.365-371.

Schlötterer, C., 2004. The evolution of molecular markers—just a matter of fashion?. *Nature reviews genetics*, 5(1), pp.63-69.

Schmieder, R. and Edwards, R., 2011. Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, 27(6), pp.863-864.

Schubert, I. and Vu, G.T., 2016. Genome stability and evolution: attempting a holistic view. *Trends in plant science*, 21(9), pp.749-757.

Schuster, S.C., 2008. Next-generation sequencing transforms today's biology. *Nature methods*, 5(1), pp.16-18.

Selkoe, K.A. and Toonen, R.J., 2006. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology letters*, 9(5), pp.615-629.

SEMA/PMA. Polícia Militar Ambiental De São Paulo (PMA-SP), 2006. Tráfico de Animais Silvestres da Fauna Nacional – Dados estatísticos e estratégias operacionais 2001-2005. Disponível em: [http://www.pea.org.br/educativo/relatorio\\_policia\\_ambiental.pdf](http://www.pea.org.br/educativo/relatorio_policia_ambiental.pdf). Acesso em: 03.08.2016.

Sequencing Library Preparation. Disponível em: [https://support.illumina.com/documents/documentation/chemistry\\_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf](https://support.illumina.com/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf). Acesso em 24.01.2019.

Sharp, C.P., LeBreton, M., Kantola, K., Nana, A., Dikko, J.L.D., Djoko, C.F., Tamoufe, U., Kiyang, J.A., Babila, T.G., Ngole, E.M. and Pybus, O.G., 2010. Widespread infection with homologues of human parvoviruses B19, PARV4, and human bocavirus of chimpanzees and gorillas in the wild. *Journal of virology*, 84(19), pp.10289-10296.

Sheikh, P.A., O'Regan, K.C., 2020. Wildlife Trade, COVID-19, and Other Zoonotic Diseases. Congressional Research Service. Disponível em: <https://crsreports.congress.gov/product/pdf/IF/IF11494>. Acesso em 06.05.2020.

Shokralla, S., Porter, T.M., Gibson, J.F., Dobosz, R., Janzen, D.H., Hallwachs, W., Golding, G.B. and Hajibabaei, M., 2015. Massively parallel *multiplex* DNA sequencing for specimen identification using an Illumina MiSeq platform. *Scientific reports*, 5(1), pp.1-7.

Silva N.S. 2015. Espécimes recebidos no centro de triagem de animais silvestres de salvador/ba durante os anos de 2012 a 2014. Trabalho de Conclusão de Curso. Universidade Federal da Bahia, Salvador. 47 p.

Smith, K.M., Anthony, S.J., Switzer, W.M., Epstein, J.H., Seimon, T., Jia, H., Sanchez, M.D., Huynh, T.T., Galland, G.G., Shapiro, S.E. and Sleeman, J.M., 2020. Zoonotic viruses associated with illegally imported wildlife products. *Australasian Policing*, 12(2), pp.24-32.

St. John, J.C. and Schatten, G., 2004. Paternal mitochondrial DNA transmission during nonhuman primate nuclear transfer. *Genetics*, 167(2), pp.897-905.

STABvida – Your easy genetics laboratory. 2021. Sexagem de aves. Disponível em: <https://www.stabvida.com/pt/sexagem-de-aves>. Acesso em 16.06.2021.

Suez, M., Behdenna, A., Brouillet, S., Graca, P., Higuete, D. and Achaz, G., 2016. MicNeSs: Genotyping microsatellite *loci* from a collection of (NGS) *reads*. *Molecular Ecology Resources*, 16(2), pp.524-533.

Summers, K. e Amos, W., 1997. Behavioral, ecological, and molecular genetic analyses of reproductive strategies in the Amazonian dart-poison frog, *Dendrobates ventrimaculatus*. *Behavioral Ecology*, 8(3), pp.260-267.

Sun, X., Liu, Y., Lutterbaugh, J., Chen, W.D., Markowitz, S.D. and Guo, B., 2006. Detection of mononucleotide *repeat* sequence alterations in a large *background* of normal DNA for screening high-frequency microsatellite instability cancers. *Clinical cancer research*, 12(2), pp.454-459.

Sutovsky, P., Moreno, R.D., Ramalho-Santos, J., Dominko, T., Simerly, C. and Schatten, G., 1999. Ubiquitin tag for sperm mitochondria. *Nature*, 402(6760), pp.371-372.

Taanman, J.W., 1999. The mitochondrial genome: structure, transcription, translation and replication. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 1410(2), pp.103-123.

Tae, H., McMahon, K.W., Settlage, R.E., Bavarva, J.H. and Garner, H.R., 2013. ReviSTER: an automated *pipeline* to revise misaligned *reads* to simple tandem *repeats*. *Bioinformatics*, 29(14), pp.1734-1741.

Tamura, K., Stecher, G., Peterson, D., Filipowski, A. and Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular biology and evolution*, 30(12), pp.2725-2729.

Tan, G., Opitz, L., Schlapbach, R. and Rehrauer, H., 2019. Long fragments achieve lower base quality in Illumina *paired-end* sequencing. *Scientific reports*, 9(1), pp.1-7.

Tang, H. and Nzabarushimana, E., 2017. STRScan: targeted profiling of short tandem *repeats* in whole-genome sequencing data. *BMC bioinformatics*, 18(11), pp.31-36.

Thermo Fischer Scientific. 2021. Qubit Fluorometric Quantification. Disponível em: <https://www.thermofisher.com/uk/en/home/industrial/spectroscopy-elemental-isotope-analysis/molecular-spectroscopy/fluorometers/qubit.html>. Acesso em 16.06.2021

Thiel, T., Michalek, W., Varshney, R. and Graner, A., 2003. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theoretical and applied genetics*, 106(3), pp.411-422

Tibihika, P.D., Curto, M., Dornstauder-Schrammel, E., Winter, S., Alemayehu, E., Waidbacher, H. and Meimberg, H., 2019. Application of microsatellite genotyping by sequencing (SSR-GBS) to measure genetic diversity of the East African *Oreochromis niloticus*. *Conservation Genetics*, 20(2), pp.357-372.

Tóth, G., Gáspári, Z. and Jurka, J., 2000. Microsatellites in different eukaryotic genomes: survey and analysis. *Genome research*, 10(7), pp.967-981.



U.S. Department of State. 2020. New Visa Restrictions for Wildlife and Timber Traffickers. Disponível em: <https://2017-2021.state.gov/new-visa-restrictions-for-wildlife-and-timber-traffickers/index.html>. Acesso em 10.03.2021.

Umar, A. and Kunkel, T.A., 1996. DNA-replication fidelity, *mismatch* repair and genome instability in cancer cells. *EJB Reviews 1996*, pp.163-173.

Unigen – Biologia pelo DNA. 2021. Exames de sexagem para aves. Disponível em: [http://www.unigen.com.br/animais/tipos\\_exam.html](http://www.unigen.com.br/animais/tipos_exam.html). Acesso em 16.06.2021.

Universidade Federal De Viçosa (UFV), 2009. Museu de Zoologia João Moojen. Bicho da vez nº 4. Trinca-ferro-verdadeiro (*Saltator similis*).

US Department of State. 2020. New Visa Restrictions for Wildlife and Timber Traffickers. Disponível em: <https://www.state.gov/new-visa-restrictions-for-wildlife-and-timber-traffickers/>. Acesso em 20.11.20.

van der Gaag, K.J., de Leeuw, R.H., Hoogenboom, J., Patel, J., Storts, D.R., Laros, J.F. and de Knijff, P., 2016. Massively parallel sequencing of short tandem repeats—population data and mixture analysis results for the PowerSeq™ system. *Forensic Science International: Genetics*, 24, pp.86-96.

Van Dijk, E.L., Auger, H., Jaszczyszyn, Y. and Thermes, C., 2014. Ten years of next-generation sequencing technology. *Trends in genetics*, 30(9), pp.418-426.

Veritas. 2019. Next in the Genomics Revolution: The Era of the "Social Genome". Disponível em: <https://www.veritasgenetics.com/our-thinking/next-genomics-revolution-era-social-genome/>. Acesso em 18.11.2020.

Viana, I.R. and Zocche, J.J., 2013. Avifauna apreendida no extremo sul catarinense: apreensões feitas durante oito anos de fiscalização e combate à captura de aves silvestres. *Revista Brasileira de Biociências*, 11(4).

Vinogradov, A.E., 1997. Nucleotypic effect in homeotherms: body-mass independent resting metabolic rate of passerine birds is related to genome size. *Evolution*, 51(1), pp.220-225.

Wagner HW and Sefc KM (1999) IDENTITY 1.0. Centre for Applied Genetics, Vienna.

Wandeler, P., Hoeck, P.E. and Keller, L.F., 2007. Back to the future: museum specimens in population genetics. *Trends in Ecology & Evolution*, 22(12), pp.634-642.

Wang, X., Lu, P. and Luo, Z., 2013. GMATo: a novel tool for the identification and analysis of microsatellites in large genomes. *Bioinformatics*, 9(10), p.541.

Wattier, R., Engel, C.R., Saumitou-Laprade, P. and Valero, M., 1998. Short allele dominance as a source of heterozygote deficiency at microsatellite *loci*: experimental evidence at the dinucleotide *locus* Gv1CT in *Gracilaria gracilis* (Rhodophyta). *Molecular Ecology*, 7(11), pp.1569-1573.

Wei, N.A., Bemmels, J.B. and Dick, C.W., 2014. The effects of *read* length, quality and quantity on microsatellite discovery and *primer* development: from Illumina to PacBio. *Molecular ecology resources*, 14(5), pp.953-965.

Wilson, M.R., DiZinno, J.A., Polansky, D., Replogle, J. and Budowle, B., 1995. Validation of mitochondrial DNA sequencing for forensic casework

analysis. *International journal of legal medicine*, 108(2), pp.68-74. Apud Tocchetto 2014.

World Animal Protection, 2020. Fim da Operação Delivery do Ibama pode facilitar o tráfico de passarinhos. Disponível em: <https://www.worldanimalprotection.org.br/Ibama-fim-da-operacao-delivery-favorece-traffic>. Acesso em 11.05.2020.

Wright, N.A., Gregory, T.R. and Witt, C.C., 2014. Metabolic ‘engines’ of flight drive genome size reduction in birds. *Proceedings of the Royal Society B: Biological Sciences*, 281(1779), p.20132780.

Ye, Y., Luo, H., Liao, L., Zhang, J., Wei, W., Wang, Z. and Hou, Y., 2014. A case study of SNPSTR efficiency in paternity testing with *locus incompatibility*. *Forensic Science International: Genetics*, 9, pp.72-75.

Zane, L., Bargelloni, L. and Patarnello, T., 2002. Strategies for microsatellite isolation: a review. *Molecular ecology*, 11(1), pp.1-16.

Zavodna, M., Bagshaw, A., Brauning, R. and Gemmell, N.J., 2014. The accuracy, feasibility and challenges of sequencing short tandem *repeats* using next-generation sequencing platforms. *PLoS One*, 9(12), p.e113862.

Zavodna, M., Bagshaw, A., Brauning, R. and Gemmell, N.J., 2014. The accuracy, feasibility and challenges of sequencing short tandem *repeats* using next-generation sequencing platforms. *PLoS One*, 9(12), p.e113862.

Zhan, L., Paterson, I.G., Fraser, B.A., Watson, B., Bradbury, I.R., Nadukkalam Ravindran, P., Reznick, D., Beiko, R.G. and Bentzen, P., 2017. MEGASAT: automated inference of microsatellite genotypes from sequence data. *Molecular ecology resources*, 17(2), pp.247-256.

Zhang, G., 2015. Bird sequencing project takes off. *Nature*, 522(7554), pp.34-34.

Zhang, Q. and Edwards, S.V., 2012. The evolution of intron size in amniotes: a role for powered flight?. *Genome biology and evolution*, 4(10), pp.1033-1043.

Zhao, X., Li, N., Guo, W., Hu, X., Liu, Z., Gong, G., Wang, A., Feng, J. and Wu, C., 2004. Further evidence for paternal inheritance of mitochondrial DNA in the sheep (*Ovis aries*). *Heredity*, 93(4), pp.399-403.

## Referências capítulo 2

Amthor, J.S. and Huston, M.A., 1998. Terrestrial ecosystem responses to global change: a research strategy. *ORNL Technical Memorandum*, 27, p.37.

Beliveau, B.J., Apostolopoulos, N. and Wu, C.T., 2014. Visualizing genomes with Oligopaint FISH probes. *Current Protocols in Molecular Biology*, 105(1), pp.14-23.

Beliveau, B.J., Boettiger, A.N., Avendaño, M.S., Jungmann, R., McCole, R.B., Joyce, E.F., Kim-Kiselak, C., Bantignies, F., Fonseka, C.Y., Erceg, J. and Hannan, M.A., 2015. Single-molecule super-resolution imaging of chromosomes and in situ haplotype visualization using Oligopaint FISH probes. *Nature communications*, 6(1), pp.1-13.

Bouzat, J.L., 2010. Conservation genetics of population bottlenecks: the role of chance, selection, and history. *Conservation Genetics*, 11(2), pp.463-478.

Butchart, S.H., Walpole, M., Collen, B., Van Strien, A., Scharlemann, J.P., Almond, R.E., Baillie, J.E., Bomhard, B., Brown, C., Bruno, J. and Carpenter, K.E., 2010. Global biodiversity: indicators of recent declines. *Science*, 328(5982), pp.1164-1168.

Cello, J., Paul, A.V. and Wimmer, E., 2002. Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. *science*, 297(5583), pp.1016-1018.

Coltman, D.W., Pilkington, J.G., Smith, J.A. and Pemberton, J.M., 1999. Parasite-mediated selection against inbred Soay sheep in a free-living island population. *Evolution*, 53(4), pp.1259-1267.

EEHV Advisory Group, 2019. EEHV *Background* information. Disponível em: <https://eehvinfo.org/eehv-information/>. Acesso em 07.02.2021.

Fuery, A., Leen, A.M., Peng, R., Wong, M.C., Liu, H. and Ling, P.D., 2018. Asian elephant T cell responses to elephant endotheliotropic herpesvirus. *Journal of virology*, 92(6), pp.e01951-17.

Gopala, A., Hadian, O., Sunarto, Sitompul, A., Williams, A., Leimgruber, P., Chambliss, S.E. & Gunaryadi, D. 2011. *Elephas maximus* ssp. *sumatranus*. *The IUCN Red List of Threatened Species* 2011: e.T199856A9129626. <https://dx.doi.org/10.2305/IUCN.UK.2011-2.RLTS.T199856A9129626.en>. Acesso em 4.07.2021.

Guntawang, T., Sittisak, T., Srivorakul, S., Kochagul, V., Photichai, K., Thitaram, C., Sthitmatee, N., Hsu, W.L. and Pringproa, K., 2020. In vivo characterization of target cells for acute elephant endotheliotropic herpesvirus (EEHV) infection in Asian elephants (*Elephas maximus*). *Scientific reports*, 10(1), pp.1-13.

Guo, M., Ren, K., Zhu, Y., Tang, Z., Wang, Y., Zhang, B. and Huang, Z., 2019. Structural insights into a high fidelity variant of SpCas9. *Cell research*, 29(3), pp.183-192.

Hedrick, Philip W., and Richard Fredrickson. "Genetic rescue guidelines with examples from Mexican wolves and Florida panthers." *Conservation genetics* 11, no. 2 (2010): 615-626.

Hoffmann, A.A. and Parsons, P.A., 1997. *Extreme environmental change and evolution*. Cambridge University Press.

Houck, M.L., Kumamoto, A.T., Gallagher Jr, D.S. and Benirschke, K., 2001. Comparative cytogenetics of the African elephant (*Loxodonta africana*) and Asiatic elephant (*Elephas maximus*). *Cytogenetic and Genome Research*, 93(3-4), pp.249-252.

IUCN, S., 2016. IUCN SSC Guiding principles on Creating Proxies of Extinct Species.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A. and Charpentier, E., 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *science*, 337(6096), pp.816-821.

Kemf, E. and Santiapillai, C. (2000). Asian Elephants in the Wild. A WWF Species Status Report. WWFInternational, Gland, Switzerland.

Ledford, H., 2015. CRISPR, the disruptor. *Nature News*, 522(7554), p.20.

Lynch, V.J., Bedoya-Reina, O.C., Ratan, A., Sulak, M., Drautz-Moses, D.I., Perry, G.H., Miller, W. and Schuster, S.C., 2015. Elephantid genomes reveal the molecular bases of woolly mammoth adaptations to the Arctic. *Cell reports*, 12(2), pp.217-228.

Madsen, T., Shine, R., Olsson, M. and Wittzell, H., 1999. Restoration of an inbred adder population. *Nature*, 402(6757), pp.34-35.

Markert, J.A., Champlin, D.M., Gutjahr-Gobell, R., Grear, J.S., Kuhn, A., McGreevy, T.J., Roth, A., Bagley, M.J. and Nacci, D.E., 2010. Population genetic diversity and fitness in multiple environments. *BMC evolutionary biology*, 10(1), pp.1-13.

Meister, T.L., Bruening, J., Todt, D. and Steinmann, E., 2019. Cell culture systems for the study of hepatitis E virus. *Antiviral research*, 163, pp.34-49.

Menon, V. 2017. *Right of passage: elephant corridors of India*. Wildlife Trust of India.

Nature portfolio, 2021. Synthetic biology. Disponível em: <https://www.nature.com/subjects/synthetic-biology>. Acesso em 03.07.21.

Nelson, D.L. and Cox, M.M., 2018. *Princípios de Bioquímica de Lehninger-7*. Artmed Editora.

Novak, B.J., 2018. *De-extinction*. *Genes*, 9(11), p.548.

Ossent, P., Guscetti, F., Metzler, A.E., Lang, E.M., Rübél, A. and Hauser, B., 1990. Acute and fatal herpesvirus infection in a young Asian elephant (*Elephas maximus*). *Veterinary Pathology*, 27(2), pp.131-133.

Palkopoulou, E., Lipson, M., Mallick, S., Nielsen, S., Rohland, N., Baleka, S., Karpinski, E., Ivancevic, A.M., To, T.H., Kortschak, R.D. and Raison, J.M., 2018. A comprehensive genomic history of extinct and living elephants. *Proceedings of the National Academy of Sciences*, 115(11), pp.E2566-E2574.

Palkopoulou, E., Mallick, S., Skoglund, P., Enk, J., Rohland, N., Li, H., Omrak, A., Vartanyan, S., Poinar, H., Götherström, A. and Reich, D., 2015. Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth. *Current Biology*, 25(10), pp.1395-1400.

Pavulraj, S., Eschke, K., Prahl, A., Flügger, M., Trimpert, J., van den Doel, P.B., Andreotti, S., Kaessmeyer, S., Osterrieder, N. and Azab, W., 2019. Fatal elephant endotheliotropic herpesvirus infection of two young Asian elephants. *Microorganisms*, 7(10), p.396.

Pennisi, E., 2013. The CRISPR craze. *Science*, 341(6148) 833-836

Photichai, K., Guntawang, T., Sittisak, T., Kochagul, V., Chuammitri, P., Thitaram, C., Thananchai, H., Chewonarin, T., Sringarm, K. and Pringproa, K., 2020. Attempt to Isolate Elephant Endotheliotropic Herpesvirus (EEHV) Using a Continuous Cell Culture System. *Animals*, 10(12), p.2328.

Piaggio, A.J., Segelbacher, G., Seddon, P.J., Alphey, L., Bennett, E.L., Carlson, R.H., Friedman, R.M., Kanavy, D., Phelan, R., Redford, K.H. and Rosales, M., 2017. Is it time for synthetic biodiversity conservation?. *Trends in Ecology & Evolution*, 32(2), pp.97-107.

Pringproa, K., Rungsiwiwut, R., Tantilertcharoen, R., Praphet, R., Pruksananonda, K., Baumgärtner, W. and Thanawongnuwech, R., 2015. Tropism and induction of cytokines in human embryonic-stem cells-derived neural progenitors upon inoculation with highly-pathogenic avian H5N1 influenza virus. *PLoS One*, 10(8), p.e0135850.

Putshkov, P.V., 1999. The impact of the mammoths on their biome: Clash of two paradigms. Em: Reumer, J.W. and Mol, D. eds., 2003. *Advances in Mammoth Research: bproceedings of the Second International Mammoth Conference, Rotterdam, 16-20 May 1999*. Natuurmuseum Rotterdam.

Ralls, K., Ballou, J.D. and Templeton, A., 1988. Estimates of lethal equivalents and the cost of inbreeding in mammals. *Conservation biology*, 2(2), pp.185-193.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A. and Zhang, F., 2013. Genome engineering using the CRISPR-Cas9 system. *Nature protocols*, 8(11), pp.2281-2308.

Rands, M.R., Adams, W.M., Bennun, L., Butchart, S.H., Clements, A., Coomes, D., Entwistle, A., Hodge, I., Kapos, V., Scharlemann, J.P. and Sutherland, W.J., 2010. Biodiversity conservation: challenges beyond 2010. *science*, 329(5997), pp.1298-1303.

Redford, K.H., 2019. *Genetic frontiers for conservation* (No. BOOK\_B). Union internationale pour la conservation de la nature. Disponível em: <https://www.iucn.org/resources/issues-briefs/synthetic-biology-and-its-implications-biodiversity-conservation>. Acesso em 03.07.21.

Revive & Restore, 2013. TEDx DeExtinction. Disponível em: <http://reviverestore.org/events/tedxdeextinction/>. Acesso em 04.07.2021.

Revive & Restore. The woolly mammoth revival – progress to date. Disponível em: <https://reviverestore.org/projects/woolly-mammoth/progress-to-date/>. Acesso em 25.06.21.

Richman, L.K., Montali, R.J., Garber, R.L., Kennedy, M.A., Lehnhardt, J., Hildebrandt, T., Schmitt, D., Hardy, D., Alcendor, D.J. and Hayward, G.S., 1999. Novel endotheliotropic herpesviruses fatal for Asian and African elephants. *Science*, 283(5405), pp.1171-1176.

Saccheri, I., Kuussaari, M., Kankare, M., Vikman, P., Fortelius, W. and Hanski, I., 1998. Inbreeding and extinction in a butterfly metapopulation. *Nature*, 392(6675), pp.491-494.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B. and Tinevez, J.Y., 2012. Fiji: an open-source platform for biological-image analysis. *Nature methods*, 9(7), pp.676-682.

Schnichels, S., Schultheiß, M., Hofmann, J., Szurman, P., Bartz-Schmidt, K.U. and Spitzer, M.S., 2012. Trichostatin A induces cell death at the concentration recommended to differentiate the RGC-5 cell line. *Neurochemistry international*, 60(6), pp.581-591.

Seilern-Moy, K., Darpel, K., Steinbach, F. and Dastjerdi, A., 2016. Distribution and load of elephant endotheliotropic herpesviruses in tissues from associated fatalities of Asian elephants. *Virus research*, 220, pp.91-96.

Srivorakul, S., Guntawang, T., Kochagul, V., Photichai, K., Sittisak, T., Janyamethakul, T., Boonprasert, K., Khammesri, S., Langkaphin, W., Punyapornwithaya, V. and Chuammitri, P., 2019. Possible roles of monocytes/macrophages in response to elephant endotheliotropic herpesvirus (EEHV) infections in Asian elephants (*Elephas maximus*). *PLoS One*, 14(9), p.e0222158.

Sukumar, R., 2003. *The living elephants: evolutionary ecology, behaviour, and conservation*. Oxford University Press.

Sukumar, R., 2006. A brief review of the status, distribution and biology of wild Asian elephants *Elephas maximus*. *International Zoo Yearbook*, 40(1), pp.1-8.

Sukumar, R., Ramakrishnan, U. and Santosh, J.A., 1998, November. Impact of poaching on an Asian elephant population in Periyar, southern India: a model of demography and tusk harvest. In *Animal Conservation forum* (Vol. 1, No. 4, pp. 281-291). Cambridge University Press.

Sukumar, R., Varma, S., Tiwari, S.K., Menon, V., 2016. Sustainable landscapes and corridors to conserve Asian elephants in India. Success, challenges, and prospects. *Tropical Conservation: Perspectives on Local and Global Priorities*.

Synthego. CRISPR 101 – Your guide to understanding CRISPR. Disponível em: <https://www.synthego.com/resources/crispr-101-ebook>. Acesso em 05.09.2019.

tenOever, B.R., 2020. Synthetic virology: Building viruses to better understand them. *Cold Spring Harbor Perspectives in Medicine*, 10(11), p.a038703.

The Academy of Natural Sciences. 2007. Woolly Mammoth (*Mammuthus primigenius*). The Academy of Natural Sciences. Disponível em: <http://www.ansp.org/museum/jefferson/otherFossils/mammuthus.php>. Acesso em 26.07.2018.

The IUCN SSC Asian Elephant Specialist Group. Disponível em <https://www.asesg.org/images/AESG%20brochure%20final.pdf>. Acesso em 05.07.2021.

Vasconcelos, M.J.V., Figueiredo, J.E.F., 2015. Tecnologia CRISPR-Cas para edição genômica. Sete Lagoas – Embrapa Milho e Sorgo. 37 p. Disponível em: <https://ainfo.cnptia.embrapa.br/digital/bitstream/item/149707/1/doc-197.pdf>. Acesso em 15.06.21.

Weber, W. and Fussenegger, M., 2012. Emerging biomedical applications of synthetic biology. *Nature Reviews Genetics*, 13(1), pp.21-35.

Williams, C., Tiwari, S.K., Goswami, V.R., de Silva, S., Kumar, A., Baskaran, N., Yoganand, K. & Menon, V. 2020. *Elephas maximus*. *The IUCN Red List of Threatened Species* 2020: e.T7140A45818198. <https://dx.doi.org/10.2305/IUCN.UK.2020-3.RLTS.T7140A45818198.en>. Acesso em 20.06.2021.

Wimmer, E. and Paul, A.V., 2011. Synthetic poliovirus and other designer viruses: what have we learned from them?. *Annual review of microbiology*, 65, pp.583-609.

Wimmer, E., Mueller, S., Tumpey, T.M. and Taubenberger, J.K., 2009. Synthetic viruses: a new opportunity to understand and prevent viral disease. *Nature biotechnology*, 27(12), pp.1163-1172.

**ANEXO I – Tamanho das *reads* brutas geradas nos sequenciamentos. Gráficos gerados pelo programa Geneious 11.1.15 (<https://www.geneious.com>).**

